technical note

Evaluation of En Route Automation's Trajectory Generation and Strategic Alert Processing

Analysis of ERAM Performance

Dr. Hollis F. Ryan, General Dynamics Information Technology

George Chandler, Simulation & Analysis Group,
Federal Aviation Administration

Confesor Santiago, Simulation & Analysis Group,
Federal Aviation Administration

Mike M. Paglione, Simulation & Analysis Group,
Federal Aviation Administration

Shurong Liu, General Dynamics Information Technology

November 2008

DOT/FAA/TC-TN08/10

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

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.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN08/10		

4. Title and Subtitle	5. Report Date
Evaluation of En Route Automation's Trajectory Generation and Strategic Alert Processing Analysis of ERAM Performance	November, 2008
	6. Performing Organization Code Simulation & Analysis Team
7. Author(s) Dr. Hollis F. Ryan, George Chandler, Confesor Santiago, Mike M. Paglione, Shurong Liu	8. Performing Organization Report No. DOT/FAA/CT-TN08/10
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 ERAM Program Office Conflict Probe Assessment Team Washington, D. C. 20590	13. Type of Report and Period Covered Technical Note
	14. Sponsoring Agency Code ATO-E

15. Supplementary Notes

16. Abstract

This technical note is the fifth in a series of metric reports for the En Route Automation Modernization Program (ERAM). This study examines the performance of the Flight Data Processing (FDP) subsystem to generate trajectories and the Conflict Probe Tool (CPT) subsystem to generate strategic conflict alerts. The metrics described in this report have been applied to quantify trajectory accuracy and the missed and nuisance detections of the strategic conflict probe. The report documents the results from the formal ERAM testing Run-For-Record (RFR) and the eight accuracy requirements tested, using reliable internal FAA tools. Next it compares results using these internal FAA tools to the official tools used by the ERAM development contractor during the RFR. Finally, the report presents new metrics and results that better capture the performance intended by the original ERAM requirements.

17. Key Words	18. Distribution Statement
Air traffic control, air traffic simulation, En Route Automation Modernization, ERAM, strategic conflict alerts, missed alerts, false alerts, warning time, aircraft flight paths, aircraft trajectory, HCS, Host Computer System, User Request Evaluation Tool, URET, accuracy	This report has restricted distribution. It is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405. To obtain a copy contact the authors to get FAA approval.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	114	

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

Acknowledgements

The authors of this technical note would like to first acknowledge the William J. Hughes Technical Center's Patrick Kelly and Thomas Choyce of the ERAM Test Group who supported this work from its inception.

Secondly, we would like to acknowledge the staff at the Lockheed Martin Transportation and Security Systems who ran our test scenarios through their ERAM and URET systems and re-formatted the output data to match our parsing tools. Most notably, we must acknowledge Andy Appel, who was largely responsible for making the runs and providing the URET and ERAM data output data essential for this work, and Charles Lapihuska, of Lockheed Martin's System Engineering Group, who provided many detailed analyses on ERAM's performance investigated during the study.

We would also like to thank the FAA's ERAM Program Office for funding this study via the ERAM Test Group at the William J. Hughes Technical Center.

Executive Summary

The Federal Aviation Administration (FAA) is developing a new Air Traffic Control (ATC) system to replace the existing Host Computer System (HCS) in the en route domain. This system is called ERAM, for En Route Automation Modernization, and is being developed by the Lockheed Martin Corporation at their Transportation and Security Solutions Division (LM) in Rockville, Maryland. The Factory Acceptance Test (FAT) was successfully performed by LM in September 2007. Although the FAA accepted the system, there were a number of deficiencies to be addressed. These deficiencies are currently being corrected by reworking parts of the computer software.

Two key functions of ERAM are the prediction of the future flight paths of the aircraft and the prediction of future conflicts between two aircraft or between an aircraft and a Special Use Airspace (SUA). A conflict occurs when the required minimum separation is not maintained. Metrics have been defined to measure the ability of ERAM to perform these functions. The metrics have been effectively used in the past, particularly for the User Request Evaluation Tool (URET) and have been extensively documented. The accuracy of ERAM's performance is determined by applying the metrics. A study has been conducted to measure the performance of ERAM, as exhibited at the FAT, to perform these two functions. The data analyzed in this report is from Runs 4 and 5 of the FAT Run-For-Record (RFR). All of the results are based on a test scenario supplied by the FAA and run on both URET and ERAM by LM.

The FAA requires the subsystems FDP and CPT of the new system, ERAM, perform at least as well as the old system being replaced, URET. At the FAT, ERAM failed most of the FDP and CPT accuracy tests. This study reports on the details of ERAM performance and establishes a baseline to monitor its future improvements. The two functions under study are implemented in ERAM by the Flight Data Processing (FDP) and the Conflict Probe Tool (CPT) subsystems. The principal metrics are altitude prediction, strategic missed conflict alert rate, and strategic false conflict alert rate. A strategic alert is one in which the predicted conflict is far enough into the future (up to 20 minutes) that immediate action is not required and there is time to adjust the traffic flows to avoid the conflict.

The study was performed by the Conflict Probe Assessment Team (CPAT) in the Systems and Analysis Group (AJP-661) at the William J. Hughes Technical Center (WJHTC) with the assistance of LM engineering staff. An extensive set of computer analysis tools, developed over several years by CPAT, was applied to the test data supplied by LM. The accuracy of the predicted flight paths, called aircraft trajectories, was measured by comparing them to the actual paths flown by the aircraft. The accuracy of predicting conflicts was measured by comparing the predictions to the conflicts which occurred in the test scenario. The test scenario was created from recorded air traffic by time shifting. The time shifting created artificial conflicts to be tested. This report describes in detail the metrics used and the processing steps required in providing the numerical results. Detailed examples are given of trajectories and conflicts.

For the FAT RFR, LM used two analysis tools to measure strategic conflict alerts – the URET Offline Problem Determination (OFPD) and the ERAM Offline Problem Determination (EOPD). The conflict alert analysis results from OFPD and EOPD have been compared to the CPAT results. They differ in a number of cases. The differences are caused by differences in the details of the definitions of the metrics. Overall, the results of the study confirm that ERAM failed, as

reported in the FAT RFR. The procedures developed for this study will be applied in the future to ERAM test scenarios as the ERAM software problems are corrected.

ERAM failed the vertical trajectory accuracy requirements[1]. Two examples have been included in the report which illustrate that ERAM can do much worse than URET. In the first example[2], URET generates 12 trajectories with an average altitude error of 2022 feet while ERAM generates 18 trajectories with an average altitude error of 6307. The numbers for the second example[3], are 13 and 1789 for URET and 20 and 5840 for ERAM.

In the study of aircraft to aircraft conflicts, the CPAT tools were utilized to tabulate conflict prediction time as well as missed alert and false alert data. ERAM is required to perform at least as well as URET. Statistical hypothesis tests were applied and are described in Appendices C and D. The results (PASS or FAIL) were in agreement with the RFR results obtained by LM[4]. ERAM passed the conflict prediction warning time requirements, but failed the missed alert and false alert requirements. In the study of aircraft to airspace conflicts, the CPAT tools were again utilized to tabulate the missed alert and false alert data, and this data was subjected to the statistical analysis as well. ERAM failed both the missed alert and false alert requirements[5]. Examples were selected to illustrate the processing by URET and by ERAM for trajectory generation, strategic aircraft-to-aircraft conflict alerts, and strategic aircraft-to-airspace alerts. Detailed analyses of the examples show the differences between URET and ERAM.

The LM analysis tools, OFPD, and EOPD, were compared to the CPAT analysis tools. The comparisons found that most of the conflict predictions results matched, but some discrepancies between the tools were discovered. These discrepancies are detailed in this report, and documented in the list of Problem Reports (PRs) found in Appendix F. All CPAT tools PRs were fixed, and then CPAT tools were rerun on the RFR dataset. Discrepancies in OFPD and EOPD were reported to LM and PRs were generated for each one. Most of the discrepancies were in the ERAM conflict prediction performance results reported by the EOPD support tool. It is recommended that all PRs be addressed prior to future OFPD/EOPD runs to ensure validity in the conflict prediction performance results.

This detailed study and analysis of ERAM accuracy as defined by the eight metrics selected illustrates both the strengths and weaknesses of ERAM and its evaluation tools, OFPD and EOPD, and points to areas where improvements are needed. Continued analysis of ERAM accuracy performance is needed as the various problems in ERAM are fixed by LM. It is also recommended that additional accuracy metrics be applied to the ERAM data to obtain a better picture of ERAM accuracy. As a result, CPAT has developed and implemented a set of additional metrics. They have been described in this report and applied to the RFR test data. It is recommended that these metrics be used in future ERAM testing. At the time of the FAT/RFR runs, ERAM's accuracy in aircraft trajectory and strategic conflict notification functions were significantly degraded as compared to the legacy URET system. The methodology and metrics presented here continue to be used to help determine if these upgrades of ERAM and EOPD have indeed addressed the problems. These activities will be subjects of future reports.

---

[1] Listed in Table 1
[2] See 3.1.4.1
[3] See 3.1.4.2
[4] See Table 20
[5] See Table 20

Table of Contents

1 INTRODUCTION 1

1.1 PURPOSE 1
1.2 BACKGROUND 1
1.3 SCOPE 2
1.4 DOCUMENT ORGANIZATION 2

2 DESCRIPTION OF THE RFR ACCURACY TEST 4

2.1 CONTRACTUAL METRICS - REQUIREMENTS 4
2.2 INPUT TRAFFIC SCENARIOS 4
2.3 OUTPUT ACCURACY DATA 6

3 ANALYSIS OF RFR RESULTS USING FAA TOOLS 7

3.1 ALTITUDE ACCURACY 7
3.1.1 Requirements 7
3.1.2 Metrics 7
3.1.2.1 Level or in Transition 7
3.1.2.2 Window Classification 8
3.1.2.3 The Trajectory Altitude Accuracy Metric 8
3.1.3 Measurement Results 8
3.1.3.1 URET Altitude Modeling Accuracy 8
3.1.3.2 ERAM Altitude Modeling Accuracy 9
3.1.4 Examples 9
3.1.4.1 GEN0019 9
3.1.4.2 AIR0003 15
3.2 STRATEGIC CONFLICT ALERTS – AIRCRAFT TO AIRCRAFT 25
3.2.1 Aircraft to Aircraft Conflict Alerts 25
3.2.2 Ground Truth from the Radar Track Data 25
3.2.3 CPAT Processing Steps 26
3.2.4 Alert Type Definitions 26
3.2.5 Analysis of URET Aircraft to Aircraft Conflict Alert Data 27
3.2.5.1 Matched Conflicts (Valid Alerts) 27
3.2.5.2 Unmatched Conflicts (Missed Alerts) 27
3.2.5.3 Unmatched Alerts (False Alerts) 27
3.2.5.4 Discards 27
3.2.6 Analysis of ERAM Aircraft to Aircraft Alert Data 28
3.2.6.1 Matched Conflicts (Valid Alerts) 28
3.2.6.2 Unmatched Conflicts (Missed Alerts) 28
3.2.6.3 Unmatched Alerts (False Alert) 28
3.2.6.4 Discards 28
3.2.7 Examples 29
3.2.7.1 Valid Alert Greater Than 10 Minutes Warning (ERD1879-C1) 29
3.2.7.2 Valid Alert Less Than 10 Minutes Warning (ERD1879-C2) 32
3.2.7.3 Missed Alert (ERD1879-C3) 35
3.2.7.4 False Alert (ERD1879-C4) 38
3.3 STRATEGIC CONFLICT ALERTS – AIRCRAFT TO AIRSPACE 40
3.3.1 Aircraft to Aircraft Conflict Alerts 40

3.3.2 Ground Truth from the Radar Track Data ... *41*
3.3.3 Analysis of URET Aircraft to Airspace Conflict Alert Data ... *41*
3.3.3.1 Matched Conflicts (Valid Alerts) ... 41
3.3.3.2 Unmatched Conflicts (Missed Alerts) ... 41
3.3.3.3 Unmatched Alerts (False Alerts) ... 41
3.3.3.4 Discards ... 41
3.3.4 Analysis of ERAM Aircraft to Airspace Conflict Alert Data ... *42*
3.3.4.1 Matched Conflicts (Valid Alerts) ... 42
3.3.4.2 Unmatched Conflicts (Missed Alerts) ... 42
3.3.4.3 Unmatched Alerts (False Alert) ... 42
3.3.4.4 Discards ... 42
3.3.5 Examples ... *43*
3.3.5.1 Missed Alert (ERD1879-C5) ... 43
3.3.5.2 False Alert (ERD1879-C6) ... 44
3.4 ERAM RUN 4 AND 5 ACCURACY RESULTS USING CPAT TOOLS ... 46

4 COMPARISON OF FAA AND CONTRACT DEVELOPER ANALYSES ... 48

4.1 COMPARISON OVERVIEW ... 48
4.2 COMPARISON PROCEDURES ... 49
4.2.1 Comparing Actual Conflicts' Prediction Results ... *50*
4.2.2 Comparing Falsely Predicted Conflicts ... *51*
4.3 CPAT VERSUS OFPD ANALYSIS ... 52
4.3.1 URET Aircraft-to-Aircraft Conflict Prediction Comparison ... *52*
4.3.1.1 URET Prediction Comparison of Actual AC-AC Conflicts ... 53
4.3.1.2 URET Prediction Comparison of False AC-AC Conflicts ... 53
4.3.2 URET Aircraft-to-Airspace Conflict Prediction Comparison ... *54*
4.3.2.1 URET Prediction Comparison of Actual AC-AS Conflicts ... 54
4.3.2.2 URET Prediction Comparison of False AC-AS Conflicts ... 55
4.4 CPAT VERSUS EOPD ANALYSIS ... 56
4.4.1 ERAM Aircraft-to-Aircraft Conflict Prediction Comparison ... *56*
4.4.1.1 ERAM Prediction Comparison of Actual AC-AC Conflicts ... 56
4.4.1.2 ERAM Prediction Comparison of False AC-AC Conflicts ... 57
4.4.2 ERAM Aircraft-to-Airspace Conflict Prediction Comparison ... *58*
4.4.2.1 ERAM Prediction Comparison of Actual AC-AS Conflicts ... 58
4.4.2.2 ERAM Prediction Comparison of False AC-AS Conflicts ... 60

5 TEST METRICS FOR TRAJECTORY AND STRATEGIC ALERT PROCESSING 62

5.1 TRAJECTORY PREDICTION METRICS ... 62
5.2 CONFLICT PREDICTION METRICS ... 64
5.2.1 Statistical Approach of Comparing Conflict Predictions ... *65*
5.2.2 Issues for Different Input Scenarios ... *68*
5.2.3 URET-to-ERAM and ERAM-to-URET Comparison of Conflict Predictions ... *69*

6 CONCLUSIONS ... 73

LIST OF ACRONYMS/ABBREVIATIONS ... 75

LIST OF SELECTED AIRPORT IDENTIFIERS ... 77

REFERENCES ... 78

APPENDIX A – PROCESSING TO ANALYZE CONFLICT PREDICTIONS ... 80

PROCESS A – LOAD CONFLICTS AND BUILD NOTIFICATION SETS ... 80

PROCESS B – VALID AND MISSED ALERT PROCESSING 80
PROCESS C – FALSE ALERT PROCESSING 81

APPENDIX B – TRAJECTORY EXAMPLES 85

FLIGHT 1 85
FLIGHT 2 85
FLIGHT 3 85

APPENDIX C – STATISTICAL TEST FOR ALTITUDE AND CONFLICT DATA 93

NORMAL DISTRIBUTION 93
APPLICATION OF THE NORMAL DISTRIBUTION 93
EXAMPLE – UPPER BOUND ON THE URET SPECIFICATION DATA – ALTITUDE PREDICTION ACCURACY 94
EXAMPLE – Z STATISTIC ON THE MISSED ALERT RATE 95
EXAMPLE – Z STATISTIC ON THE FALSE ALERT RATE 95

APPENDIX D – STATISTICAL TEST FOR WARNING TIME DATA 97

CALCULATION OF THE T STATISTIC 97
EXAMPLE – WARNING TIMES 98

APPENDIX E – OVERVIEW OF CONFLICT PREDICTION PERFORMANCE COMPARISON 99

APPENDIX F – LIST OF PROBLEM REPORTS (PRS) IN OFPD, EOPD, AND CPAT TOOLS. 100

List of Figures

Figure 1: ERAM Trajectory #1 for Flight GEN0019 10
Figure 2: ERAM Trajectory #8 for Flight GEN0019 11
Figure 3: ERAM Trajectory #9 for Flight GEN0019 11
Figure 4: ERAM Trajectory #10 for Flight GEN0019 12
Figure 5: ERAM Trajectory #11 for Flight GEN0019 12
Figure 6: ERAM Trajectory #16 for Flight GEN0019 13
Figure 7: URET Trajectory #1 for Flight GEN0019 14
Figure 8: URET Trajectory #3 for Flight GEN0019 14
Figure 9: URET Trajectory #8 for Flight GEN0019 15
Figure 10: ERAM Trajectory #1 for Flight AIR0003 18
Figure 11: ERAM Trajectory #9 for Flight AIR0003 19
Figure 12: ERAM Trajectory #18 for Flight AIR0003 19
Figure 13: ERAM Trajectory #19 for Flight AIR0003 20
Figure 14: URET Trajectory #1 for Flight AIR0003 21
Figure 15: URET Trajectory #2 for Flight AIR0003 21
Figure 16: URET Trajectory #9 for Flight AIR0003 22
Figure 17: URET Trajectory #13 for Flight AIR0003 22
Figure 18: Flight Paths of Flights AIR0001 and AIR0027 30
Figure 19: AIR0001 and AIR0027 at Start of Conflict 31
Figure 20: AIR0001 and AIR0027 at End of Conflict 31
Figure 21: Flight Paths of Flights AIR0007 and AIR0015 33
Figure 22: AIR0007 and AIR0015 at Start of Conflict 34
Figure 23: AIR0007 and AIR0015 at End of Conflict 34
Figure 24: Flight Paths of Flights AIR0011 and AIR0013 36
Figure 25: AIR0011 and AIR0013 at Start of Second Conflict 37
Figure 26: AIR0011 and AIR0013 at End of Second Conflict..... 37
Figure 27: Flight Paths of Flights AIR0005 and AIR0009 39
Figure 28: Flight Path of Flight AIR0025 44
Figure 29: Flight Path of Flight GEN0021 45
Figure 30: Conflicts detected by EOPD and CPAT tools. 48
Figure 31: Conflict Prediction Comparison Process. See following tables for matrix examples. 49
Figure 32: Overlap of Trajectory Modeling 62
Figure 33: Distributions of System Differences per Flight Mean Errors 64
Figure 34: Process A - Loads conflict and build notification sets for processing in Process B and C. 82
Figure 35: Process B - Evaluate all conflicts and determine all Valid and Missed Alerts. 83
Figure 36: Process C - Evaluate all unmatched notification sets and determine False and Discard Alerts. 84
Figure 37: URET - Altitude profile and initial trajectory for Flight 1..... 86
Figure 38: ERAM - Altitude profile and initial trajectory for Flight 1 86
Figure 39: URET - Altitude profile and a later trajectory for Flight 1 87
Figure 40: ERAM - Altitude profile and a later trajectory for Flight 1 87
Figure 41: URET - Altitude profile and initial trajectory for Flight 2..... 88
Figure 42: ERAM - Altitude profile and initial trajectory for Flight 2 88
Figure 43: URET - Altitude profile and a later trajectory for Flight 2 89
Figure 44: ERAM - Altitude profile and a later trajectory for Flight 2 89
Figure 45: URET - Altitude profile and initial trajectory for Flight 3..... 90
Figure 46: ERAM - Altitude profile and initial trajectory for Flight 3 90
Figure 47: URET - Altitude profile and a later trajectory for Flight 3 91

Figure 48: ERAM - Altitude profile and a later trajectory for Flight 3 91
Figure 49: This flowchart illustrates how AC-AC and AC-AS conflict predictions results for actually occurring conflicts and falsely predicted conflicts were compared between CPAT tools and OFPD. 99
Figure 50: This flowchart illustrates how AC-AC and AC-AS conflict predictions results for actually occurring conflicts and falsely predicted conflicts were compared between CPAT tools and EOPD. 99

List of Tables

Table 1: ERAM Accuracy Results for Run 4 and 5 5
Table 2: Altitude Modeling Accuracy 9
Table 3: ERAM Altitude Prediction Window Summary for GEN0019 15
Table 4: ERAM Altitude Prediction Windows for GEN0019 - Maximum Errors 16
Table 5: URET Altitude Prediction Window Summary for GEN0019 17
Table 6: URET Altitude Prediction Windows for GEN0019 - Maximum Errors 17
Table 7: ERAM Altitude Prediction Window Summary for AIR0003 23
Table 8: ERAM Altitude Prediction Windows for AIR0003 - Maximum Errors 23
Table 9: URET Altitude Prediction Window Summary for AIR0003 24
Table 10: URET Altitude Prediction Windows for AIR0003 – Maximum Errors 24
Table 11: Alert Type Definitions 26
Table 12: URET False Alert Rates 27
Table 13: URET Discard Rates 27
Table 14: ERAM False Alert Rates 28
Table 15: ERAM Discard Rates 28
Table 16: URET False Alert Rates 41
Table 17: URET Discard Rates 42
Table 18: ERAM False Alert Rates 42
Table 19: ERAM Discard Rates 43
Table 20: ERAM Accuracy Results Using CPAT Tools 46
Table 21: The four comparison combinations for analyzing conflict prediction performances.... 50
Table 22: An example of the matrix containing actual conflict prediction result counts from CPAT and LM. 51
Table 23: An example of the matrix containing false conflict prediction result counts from CPAT and LM. 53
Table 24: Matrix containing URET's actual AC-AC conflict prediction result counts as detected by CPAT and LM. 54
Table 25: Matrix containing URET's false AC-AC conflict prediction result counts from CPAT and LM. 54
Table 26: Matrix containing URET's actual AC-AS conflict prediction result counts as detected by CPAT and LM. 55
Table 27: Matrix containing URET's false AC-AS conflict prediction result counts from CPAT and LM. 55
Table 28: Matrix containing ERAM's actual AC-AC conflict prediction counts as detected by CPAT and LM. 58
Table 29: Matrix containing ERAM's false AC-AC conflict prediction result counts from CPAT and LM. 58
Table 30: Matrix containing ERAM's actual AC-AS conflict prediction result counts as detected by CPAT and LM. 59
Table 31: Conflict start and end time information of conflicts between GEN0017 and SUA7. ... 59
Table 32: Matrix containing ERAM's false AC-AS conflict prediction result counts from CPAT and LM. 61
Table 33: Trajectory Prediction Context Statistics 63
Table 34: Paired Error Statistical Results 64
Table 35: Conflict Prediction Result's Main Reason Codes 65
Table 36: Comparison of Two Run's Resulting Alert and Conflict Event Combinations 66
Table 37: Conflict Prediction Comparison Program Evaluation Codes 67
Table 38: Run Results for Comparison of ERAM and URET 70

Table 39: Detailed Reason Codes for Comparison Runs of ERAM and URET 71
Table 40: Comparative Counts Between ERAM and URET Comparison Runs..... 71
Table 41: Statistical Results for False Alert Comparison of ERAM and URET Runs 72
Table 42: Statistical Results for Missed Alert Comparison of ERAM and URET Runs 72
Table 43: Altitude Errors for the URET Trajectories..... 85
Table 44: Altitude Errors for the ERAM Trajectories..... 92
Table 45: List and status of problem reports (PRs) in OFPD, EOPD, and CPAT tools..... 100

1 Introduction

1.1 Purpose

The Federal Aviation Administration (FAA) is developing a new Air Traffic Control (ATC) system to replace the existing Host Computer System (HCS) in the en route domain. The Host system is used by all twenty en route ATC Centers in the continental United States. The new system, called ERAM (for En Route Automation Modernization), is being developed by the Lockheed Martin Corporation. As documented in the FAA's *Test and Evaluation Master Plan*, the ERAM Test Program is required to ensure key operational issues are verified [WJHTC/ACB-550, 2003]. These issues are organized as "Critical Operational Issues" (COI's). The first critical operational issue (COI 1.0) requires that ERAM supports ATC operations with at least the same effectiveness as the current system. Therefore, the performance of the radar track subsystem in ERAM must be as good as the performance of the existing Host radar tracking. To determine this, a baseline performance of the Host is required to provide performance standards to later compare to ERAM.

This technical note documents the accuracy performance obtained by measuring the ERAM system during the formal Run-For-Record (RFR) in August 2007. This includes measuring the accuracy of ERAM's Flight Data Processing's (FDP) aircraft trajectory predictions and Conflict Probe Tool's (CPT) aircraft-to-aircraft and aircraft-to-airspace strategic conflict predictions. These functions are currently performed by the legacy User Request Evaluation Tool (URET). Therefore, the study compares the accuracy of ERAM's predictions to the legacy URET system.

To drive this study, four test scenarios were created by the FAA's Conflict Probe Assessment Team (CPAT) and supplied to ERAM's development contractor, Lockheed Martin (LM), as government furnished property (GFP). LM ran these scenarios through both the existing or legacy URET system and then its replacement, ERAM.

This study is one of several being conducted by the Automation Metrics Test Working Group (AMTWG). The studies are described in the "ERAM Automation Metrics and Preliminary Test Implementation Plan," published in June 2005 (WJHTC/ACB-330, 2005).

1.2 Background

The FAA's ERAM Test Group formed the Automation Metrics Test Working Group (AMTWG) in August 2003. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. The targeted system functions are Surveillance Data Processing (SDP), Flight Data Processing (FDP), Conflict Probe Tool (CPT), and the Display System (DS) modules. The metrics are designed to measure the performance of ERAM. They also are designed to measure the performance of the legacy En Route automation systems in operation today. Many of the metrics will allow comparison of the functionality in ERAM to similar functionality in the HCS and URET (User Request Evaluation Tool) legacy systems.

The project was divided into key phases: first a metrics identification process was performed. A list of approximately one hundred metrics was generated by the AMTWG and mapped to the Air Traffic services and capabilities found in the Blueprint for the National Airspace System Modernization 2002 Update (FAA, 2002). This took place during most of Fiscal Year 2004 and initial metrics results were published in June 2004 in the document, "ERAM Automation Metrics

Progress Report of the Automation Metrics Test Working Group" (WJHTC/ACB-550, 2004). Next, an implementation-planning phase was performed. In this step, the identified metrics were prioritized for more detailed refinement during 2005 and 2006.

The report "ERAM Automation Metrics and Preliminary Test Implementation Plan," (WJHTC, 2005) documents the implementation-planning phase. It lists these metrics, gives the rational for selecting them, and provides a high level description on how the highest priority metrics will be measured. The Implementation Plan provides the metric's traceability to the basic controller decisions, ERAM Critical Operational Issues (COIs), and the development contractor's Technical Performance Measurements (TPMs). The categories of high priority metrics are: (1) SDP radar tracking, (2) SDP tactical alert processing, (3) FDP flight plan route expansion, (4) FDP aircraft trajectory generation, (5) CPT strategic aircraft-to-aircraft conflict prediction, (6) CPT aircraft-to-airspace conflict prediction, (7) additional system level metrics, and (8) DS human factor and performance metrics.

Currently the AMTWG is analyzing the performance of different aspects of the FAA's en route air traffic control system. The AMTWG is refining and then applying the previously defined metrics to the legacy systems. This work is being documented in a series of metric reports, each covering one of the ERAM modules discussed above, SDP, FDP, CPT, and DS. These reports are being published in multiple drops to provide the ERAM Test Team on-time information. The drops coincide with the approaches used to implement the metrics.

This technical note covers two of the ERAM modules: FDP and CPT functions. It documents the accuracy of the legacy URET trajectory and strategic conflict predictions and its replacement functions in ERAM.

1.3 Scope

A representative sample of en route air traffic was recorded in March 2005 at the Washington Air Route Traffic Control Center (ARTCC), referred to as ZDC. It contains approximately five hours of flight data and approximately 2200 flights. There were two sets of scenarios generated for this study. The first was un-altered and used to test the accuracy of the trajectory predictions in the URET and replacement in ERAM's FDP sub-system. For the second scenario, the flight times were adjusted to create conflicts to be processed. This adjustment procedure is described in detail in (Paglione et al, 2003b). For the later scenario approximately 250 to 230 conflicts were induced to test the strategic conflict predictions in the URET and replacement in ERAM's CPT sub-system. These scenarios are described in more detail in Section 3.

This technical note documents the results from the scenarios and legacy URET and ERAM systems that took place at the formal ERAM RFR testing in August 2007. It first includes results of the ERAM system in terms of the FDP trajectory and CPT accuracy requirements. Next, the technical note documents a comparison of results from the URET Offline Problem Determination (OFPD) tool and ERAM Offline Problem Determination (EOPD) tool to the FAA's internal CPAT accuracy tools, referred to as TrajTools and ConflictTools. Finally, this technical note presents a set of test metrics and their implementation applied to the RFR runs that goes beyond the simply testing the existing requirements but validating the performance of the system.

1.4 Document organization

This technical note is organized into the following primary sections. Section 2 describes the Run for the Record (RFR) acceptance test performed by Lockheed Martin for the FAA. The data analyzed in this study is the data obtained from this test. The accuracy of ERAM and URET in

predicting aircraft flight altitudes and aircraft-to-aircraft and aircraft-to-airspace conflicts is analyzed in Section 3. Detailed examples are presented in this section. A comparison of the FAA's CPAT data analysis and Lockheed Martin's data analysis is given in Section 4. A number of accuracy metrics other than those used in this report are available to be applied to ERAM. They are described in Section 5. Conclusions are in Section 6, followed by a list of acronyms and abbreviations and a set of references. The data reduction and analysis processing done for this study is described in Appendix A. Examples of trajectory errors are presented in Appendix B. The statistical tests used are defined in Appendices C and D. An overview of the conflict prediction and performance comparisons is given in Appendix E and problems found in the analysis tools are listed in Appendix F.

2 Description of the RFR Accuracy Test

URET and ERAM compute trajectories or future flight paths of the aircraft in the en route airspace and use the trajectories to predict conflicts, both aircraft-to-aircraft conflicts and aircraft-to-airspace conflicts. The predictions are for strategic conflicts – between 5 minutes and 20 minutes into the future. The URET system complements the HCS's flight data processing, generating its own trajectory and strategic conflict predictions. In ERAM, these functions are performed by the FDP and CPT subsystems.

The ERAM Test Program has different contractual levels of requirement testing. This study will focus on a set from the highest level, referred to the as the A-level accuracy requirements that measure the accuracy of ERAM in predicting aircraft altitude profiles and strategic conflicts. As referred to earlier as COI 1.0, a critical objective of the testing program is to ensure the replacement ERAM system produces functions with equivalent or improved performance over its legacy systems in URET and the HCS. As a result, these particular requirements concentrate on directly comparing the legacy HCS and URET systems to its replacement in ERAM. This is accomplished by inputting the same operational data, formatted appropriately to drive each system, and compare the resulting predictions.

The formal testing program produced results to verify the status of eight contractual requirements. The following Section 2.1 lists these requirements and results from the RFR test as documented in the FAA memorandum [17]. The subsequent Section 2.2 describes the input data that drove this test, referred to as the input traffic scenarios.

2.1 Contractual Metrics - Requirements

The numerical values obtained for the eight contractual metrics in the formal RFR testing in September 2007 are listed in Table 1 in Section 2.2. The FDP9389 and FDP9390 metrics are described in Section 3.1.2. Definitions of the other metrics can be found in [2]. The FDP9389 and FDP9390 results are based on the FAA's TrajTools (formally provided as GFP for the RFR test). The ERD1879 results are based on the development contractor's test tools called: OFPD and EOPD for URET and ERAM respectively. These requirements will be described in detail later in Sections 3.1.1, 3.1.2, 3.2.1, and 3.3.1 for aircraft-to-aircraft and aircraft-to-airspace conflict predictions respectively. Of the eight metrics, the first two are for altitude prediction and the other six are for conflict prediction. The required values have been obtained from the accuracy performance of URET [10].

2.2 Input Traffic Scenarios

For the testing of the requirements presented in Section 2.1, the FAA's CPAT created test scenarios based on two sources of recorded field data and airspace adaptation data. CPAT obtained ATC HCS data from the Host Air Traffic Management Data Distribution System (HADDS). It includes data recorded over a several hours from the Washington ARTCC (ZDC) on March 17, 2005. This data consists of ATC messages (such as flight plan, flight plan amendment, and interim altitude messages) and radar surveillance track messages produced by the legacy HCS tracker. The second set of data consists of the untracked input radar target surveillance data obtained from the En Route Radar Intelligent Tool (ERIT) recorded over the same time period. This ERIT data consists of the radar surveillance position reports associated to the flights by beacon code and Search Real Time Quality Control (SRTQC) reports that provide

the scan locations of the individual radars. The beacon associated position reports are the main input to ERAM's radar tracker in its Surveillance Data Processing (SDP) sub-system, which drive the FDP and CPT functions being evaluated. Since this recorded data does not normally contain aircraft-to-aircraft conflict events, the flights were time shifted to induce conflicts for test purposes, thus creating two test scenarios: a non-time shifted scenario and a time shifted scenario, each containing the same flights. Besides being a required input for the SDP tracker, the SRTQC reports are also critical component in time shifting the flights.

CPAT used these two scenarios to generate input files for URET and for ERAM test runs:. For URET this consists of a single file containing ATC and track messages. This represents the legacy HCS data that typically drives URET in the field. For ERAM, the input scenarios consist of two files, one containing the ATC messages and the other containing the radar target messages from ERIT. This results in the following four sets of scenario files that were provided to the development contractor as GFP to drive the respective RFR tests:

1. one non-time shifted URET scenario file
2. two non-time shifted ERAM scenario files
3. one time shifted URET scenario file
4. two time shifted ERAM scenario files

Table 1: ERAM Accuracy Results for Run 4 and 5

Requirement Number	Description	Req. (URET)	ERAM	RFR Status
FDP9389	Vertical trajectory accuracy – altitude modeling for level flight	0.0016	0.0088	FAIL
FDP9390	Vertical trajectory accuracy – altitude modeling for transitioning flight	0.1431	0.1785	FAIL
ERD1879-C1	CPT's aircraft-to-aircraft conflict prediction warning time, look-ahead > 10 min.	854 seconds	740 seconds	PASS
ERD1879-C2	CPT's aircraft-to-aircraft immediate conflict prediction warning time, look-ahead < 10 min.	104 seconds	128 seconds	PASS
ERD1879-C3	CPT's aircraft-to-aircraft missed conflict alert rate	0.025	0.067	FAIL
ERD1879-C4	CPT's aircraft-to-aircraft false conflict alert rates	0.16,0.28, 0.007, 0.005	Meas. Error	FAIL
ERD1879-C5	CPT's aircraft-to-airspace missed conflict alert rate	0.02	0.062	FAIL
ERD1879-C6	CPT's aircraft-to-airspace false conflict alert rates	0.14, 0.007, 0.01, 0.003, 0.003	0.08, 0.01, 0.002, 0.001, 0.03	FAIL

The URET files were formatted into an ASCII pipe delimited version of the HCS Common Message Set (CMS) [23]. This was developed during the URET Testing Program by CPAT and forms the main input into URET's GWH Driver, which emulates a HCS driving the various URET systems with the traffic data in the CMS scenario.

For ERAM, a mode of the ATCoach simulator is invoked that reads the ATC clearances in one file and the radar data in a separate file, injecting them into ERAM appropriately, emulating the operational data flow.

2.3 Output Accuracy Data

Once the air traffic scenarios are injected into URET and ERAM, the predictions need to be captured and input into the various test tools for analysis. These include CPAT's data reduction and analysis tools, referred to as TrajTools and ConflictTools, and the formal test tools OFPD and EOPD. Both URET and ERAM have their own system analysis recording (SAR) capabilities. The binary files that are produced are parsed with a set of scripts developed and implemented by the development contractor. This results in three output files per scenario run, including (1) the trajectory file, (2) the converted route file, and (3) the predicted alert file.

The trajectory file contains each four dimensional trajectory prediction in the form of a sorted listing of the trajectory's predicted positions in time, stereographic x-y coordinates, altitude, and ground speed. The route file contains the horizontal expansion of the fixes in the flight plan amendments. Finally the alert file contains the recordings of the predicted aircraft-to-aircraft and aircraft-to-airspace notifications. These events are posted into a series of records listing the aircraft identification codes, times of that the notification was first posted (added) to the ATC displays, updated, and finally removed in the form of a delete record.

For trajectory accuracy metrics (e.g. FDP9389 and FDP9390), CPAT's TrajTools are employed by both the development contractor and the FAA. The input data for these tools besides the air traffic scenario itself are the trajectory and route files produced by URET and ERAM. For the conflict prediction related metrics, CPAT's ConflictTools are run after the TrajTools, building on its results. It requires the additional alert file described above. The development contractor normally uses only OFPD and EOPD directly on the URET and ERAM SARs to take its measurements.

3 Analysis of RFR Results Using FAA Tools

The results presented in Section 2.1 in Table 1 are based on the trajectory processing results of the FAA's TrajTools. These tools are the only software available for providing the measurements. The following Sections 3.1, 3.2, 3.3, and 3.4 provide much more detail in the calculation of these metrics and the results from the RFR run. In Section 2.1, the strategic conflict prediction results are based on the development contractor OFPD and EOPD tools. Due to the various technical problems uncovered in these tools during the RFR and beyond, the following Sections 3.1 and 3.2 provide the processing descriptions and results based on CPAT's ConflictTools.

3.1 Altitude Accuracy

This section provides descriptive details on the two altitude modeling requirements for ERAM. As the name implies, these requirements measure the accuracy of ERAM's trajectory modeling in the vertical dimension. The following sections will define the requirements, metrics, and results from the RFR.

3.1.1 Requirements

ERAM is required, just as URET was, to predict the altitudes of the aircraft to within +/- 500 feet when the aircraft is flying level and to within +/- 1500 feet when the aircraft is climbing or descending. These requirements are number FDP9389 and number FDP9390. The specification is larger for aircraft in altitude transition because of the inherent uncertainty of the location of the Top of Descent (TOD), the Bottom of Climb (BOC), and so forth. Often the initiation of these altitude transition maneuvers are at the pilot's discretion and therefore much more difficult to accurately predict.

3.1.2 Metrics

The accuracy of the altitude prediction is determined by comparing the predicted values with the actual values as obtained from the aircraft's radar track. Each aircraft in the test scenario has one track and several trajectories (predicted flight paths). The aircraft positions for the tracks are interpolated to coincident 10 second intervals, while the trajectory positions are interpolated only if needed to match the coincident measurement times. Measurements are made on the trajectories on a point by point basis. The difference in altitudes between the trajectory point and the time coincident track point is calculated for every sampled point.

3.1.2.1 Level or in Transition

Since the accuracy requirements are different for level and non-level flight, it is necessary to determine, for each sampled point, whether the aircraft is flying level or not. This is done by sampling a set of contiguous points, a window of points, and then determining whether to label the points as all level or all in transition. The phase of flight of each track point is calculated by applying a heuristic algorithm, described in detail in [Paglione 2006a]. The determination of whether the entire window of points is labeled level or transition is a result of applying a set of rules. For a window to be labeled as level, the data must satisfy several requirements. If any of these requirements are not met, the window is labeled as in transition. In the processing of the track data previous to the altitude comparisons, every track point is labeled as being either a level point or a point in transition. The following sub-sections describe the rules in detail.

3.1.2.1.1 Window Selection

A track and a trajectory are sampled as a sequence of windows. A window is a set of contiguous time coincident data points from both track and trajectory. A window is a parameter time (780 seconds) long. Successive windows are generated every parameter sample time (30 seconds) along the track.

3.1.2.1.2 Window Truncation

Any one of the following events will truncate the particular sample window. The event causes all the subsequent measurements in the window to be discarded.

Truncation events:

1. An outbound handoff
2. A clearance
3. A data point out of vertical adherence by a parameter (500 feet) amount
4. A gap in track data larger than a parameter (120 seconds) amount

3.1.2.1.3 Rules for Level or in Transition

The requirements for a window of data points to be a level window are

1. Every track point in the window must be labeled as level
2. The altitude of every track point in the window must be within 500 feet of the assigned altitude (that is the data point must be in adherence)
3. Every track point less than (or equal to) 5 minutes before the start of the window and after the end of the window must be labeled as level

Otherwise the window is an in transition window. If there are less than 5 minutes worth of data before or after the window, the data that is available is used.

3.1.2.2 Window Classification

The altitude of each data point in sample window is compared to the actual altitude the aircraft flew at the same time. If the altitude of any point in the window differs by more than the allowed value (500 feet for level flight and 1500 feet for climbing or descending flight) from the actual value, the window is classified as being in violation of the altitude accuracy requirement. Otherwise the window is declared to be in compliance. Thus, the measurement is the determination that maximum vertical error for the sample window is above a threshold.

3.1.2.3 The Trajectory Altitude Accuracy Metric

The metric is the percentage or ratio of the windows which are in compliance. Table 1 above gives the required ratio of the windows that are required to be in compliance – 0.0016 for level flight and 0.1431 for transition flight.

3.1.3 Measurement Results

The output data from the legacy URET system and from the new ERAM system for the non-time shifted scenario was post processed by CPAT to obtain the altitude trajectory modeling accuracy.

3.1.3.1 URET Altitude Modeling Accuracy

The non-time shifted scenario has 2270 flights which produce 127,587 sample windows of which 41,858 have level trajectory data and 85,729 have transition data (climbing or descending). 66 of the level windows are inaccurate making an error rate of 66/41858 = 0.0016 for level flight.

12,268 of the climbing / descending windows are inaccurate making an error rate of 12268/85729=0.1431.

3.1.3.2 ERAM Altitude Modeling Accuracy

The same scenario when processed by ERAM yielded the following results. 2199 flights produced 123,183 sample windows of which 41,555 have level cruise data and 81,628 have transition data. 367 of the level windows have one or more data points outside the required +/ 500 foot accuracy making an error rate of 367/41555 = 0.0088 for level flight. 14,572 of the climbing / descending windows have one or more data points outside the +/- 1500 feet tolerance making an error rate of 14572/81628 = 0.1785. These numbers are presented in the following Table 2.

The ERAM test scenario had fewer aircraft than the URET scenario because ERAM was unable to build trajectories for some of the flights[6].

Table 2: Altitude Modeling Accuracy

NUMBER PROCESSED	URET DATA	ERAM DATA
Flights	2270	2199
Windows – Total	127587	123183
Windows – Level	41858	41555
Windows – Transition	85729	81628
Error Windows – Level	66	367
Error Rate – Level	0.0016	0.0088
Error Windows – Transition	12268	14572
Error Rate – Transition	0.1431	0.1785

Using the normal distribution approximation, the 90% confidence interval bounds for the level flight altitude prediction error are 0.81% - 0.96%. For transition altitudes the confidence interval bounds on the error rate are 17.63% - 18.07%. Since both of these error rates are larger than the corresponding error rates on URET, ERAM failed these two altitude prediction accuracy tests.

3.1.4 Examples

Two examples have been selected to illustrate the altitude prediction errors. Both examples are descents which are particularly difficult to predict. Additional examples are given in Appendix B. Three flights were picked that had very large altitude prediction errors.

3.1.4.1 GEN0019

In the scenario, this aircraft, a Learjet flying from Saint Petersburg FL to Morristown NJ, descends from Flight Level (FL) 390 to 19,746 in series of steps, leveling off at FL360, FL320, FL280, and FL240.

[6] The fact that ERAM did not build trajectories for <u>all</u> the flights that the legacy URET system did, can skew the test results. The RFR statistical tests do consider sample size as described in the Appendix C, but further discussion on the dropped flights is beyond the scope of this document.

3.1.4.1.1 ERAM Trajectory Data

The ERAM track data for this flight extend from FL390 at 67,000 seconds UTC (Coordinated Universal Time or Greenwich Mean Time) to 19,746 feet at 68,230 seconds UTC and has four step-down interim altitudes at FL360, FL320, FL280, and FL240.

ERAM generates 18 trajectories. The average vertical error for all of the active portions of the trajectories is 6307 feet. The first 7 trajectories descend immediately to a single step, leveling off at FL210. See the altitude profile for Trajectory #1 in Figure 1. Successive trajectories (1 through 7) start later and later at the current track point and also level off later at FL210. Trajectory #8 (
Figure 2), starting at FL 360, where the aircraft is cruising in level flight, continues the level flight and then descends directly with no leveling off steps. Trajectory #9 (
Figure 3) descends and levels of at FL320 and then descends directly. Trajectory #10 (Figure 4) continues level briefly at FL320 and descends, leveling off at FL290, and then descends directly. Trajectories #11 (Figure 5) through #15 descend the aircraft to FL240 and then directly descend to the destination airport. Trajectories #16 (Figure 6) through #18 descend the aircraft to FL130 and then directly descend to the airport.

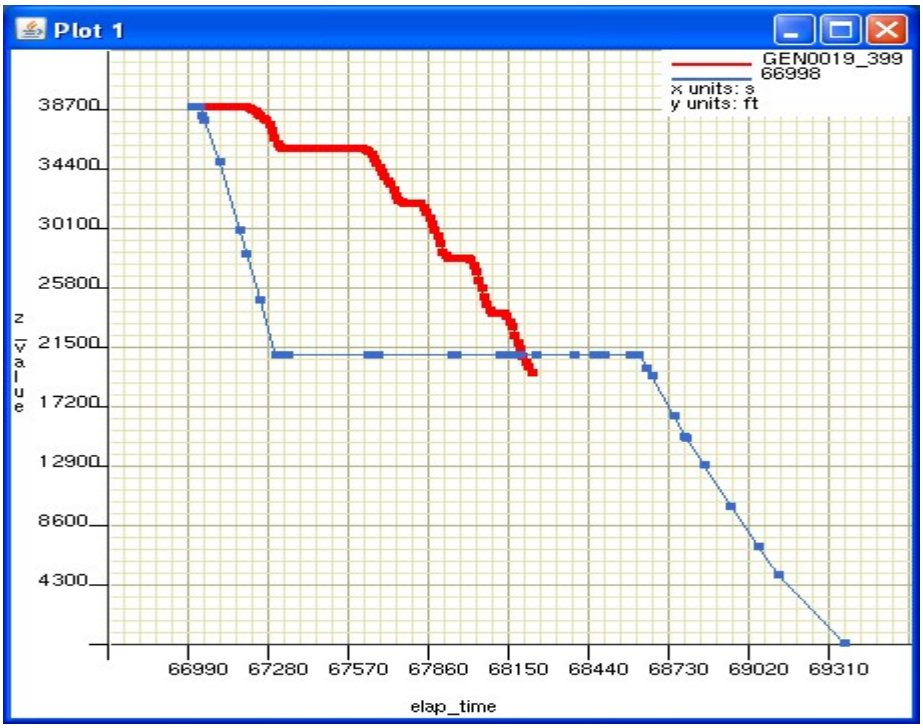

Figure 1: ERAM Trajectory #1 for Flight GEN0019

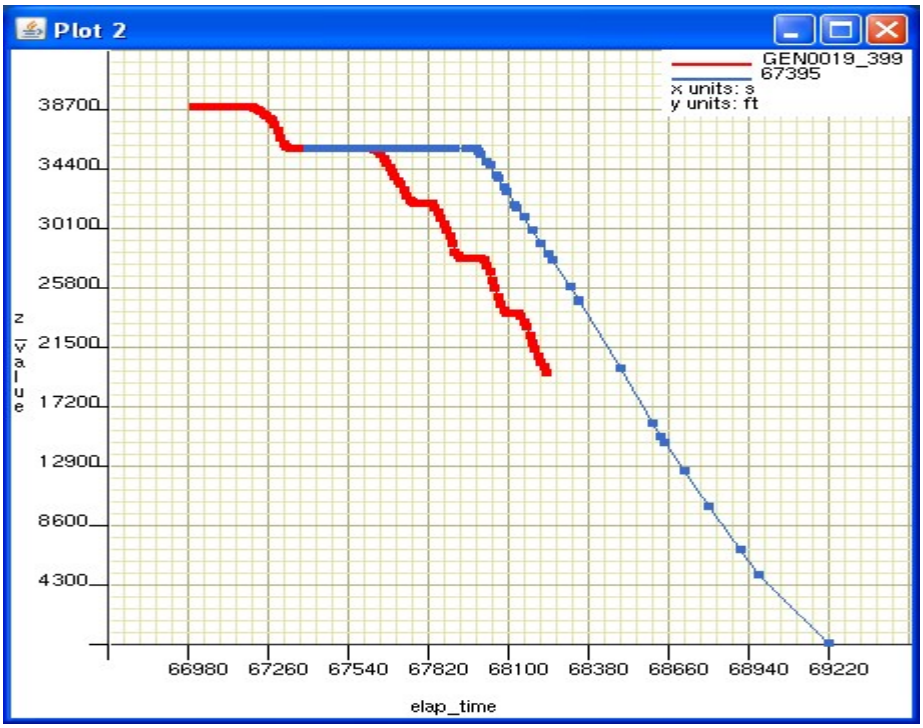

Figure 2: ERAM Trajectory #8 for Flight GEN0019

`

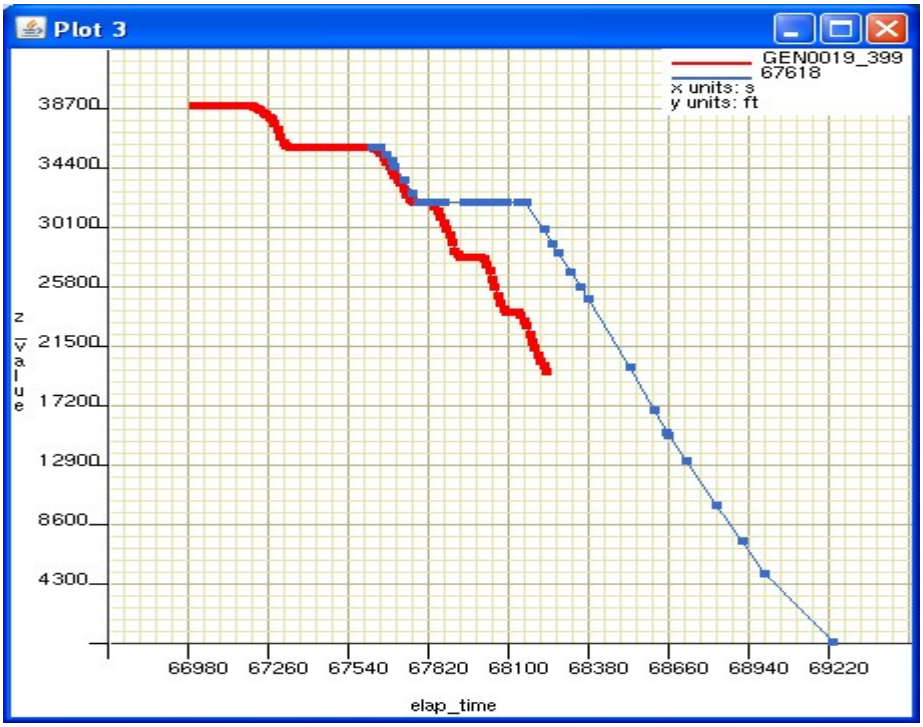

Figure 3: ERAM Trajectory #9 for Flight GEN0019

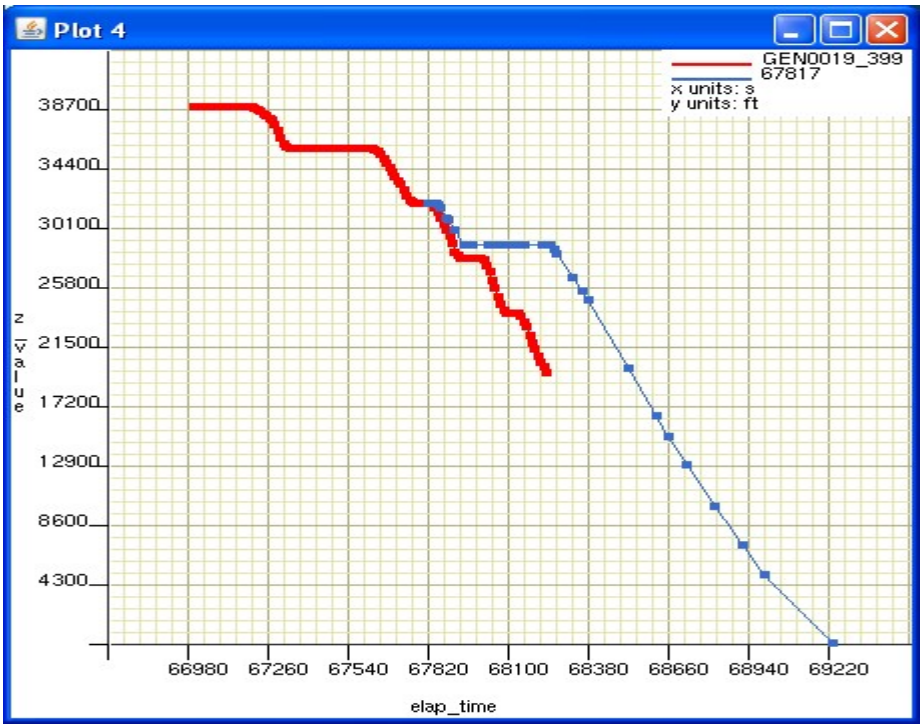

Figure 4: ERAM Trajectory #10 for Flight GEN0019

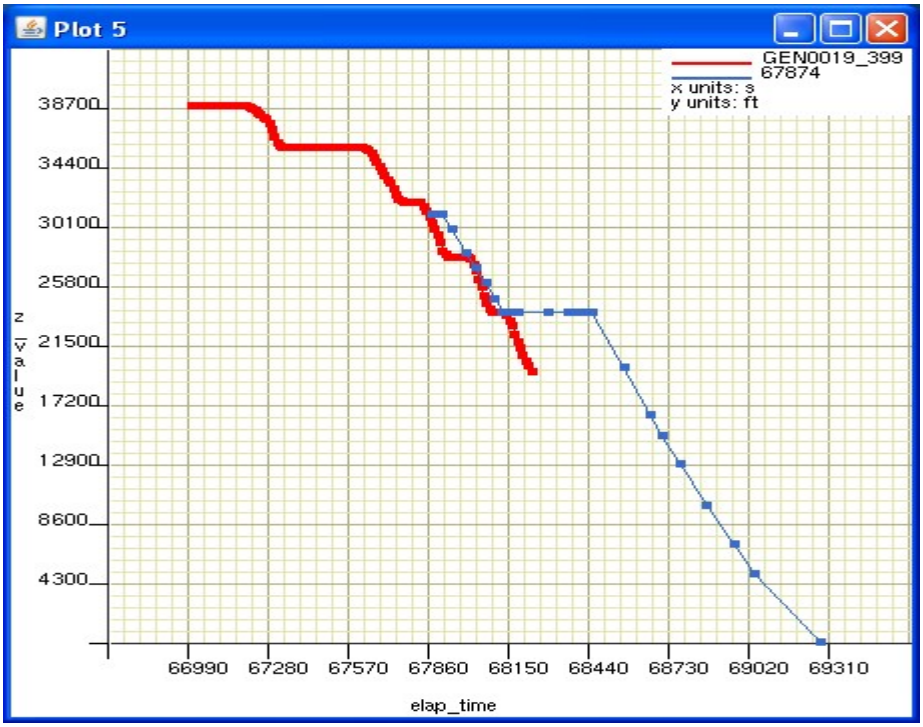

Figure 5: ERAM Trajectory #11 for Flight GEN0019

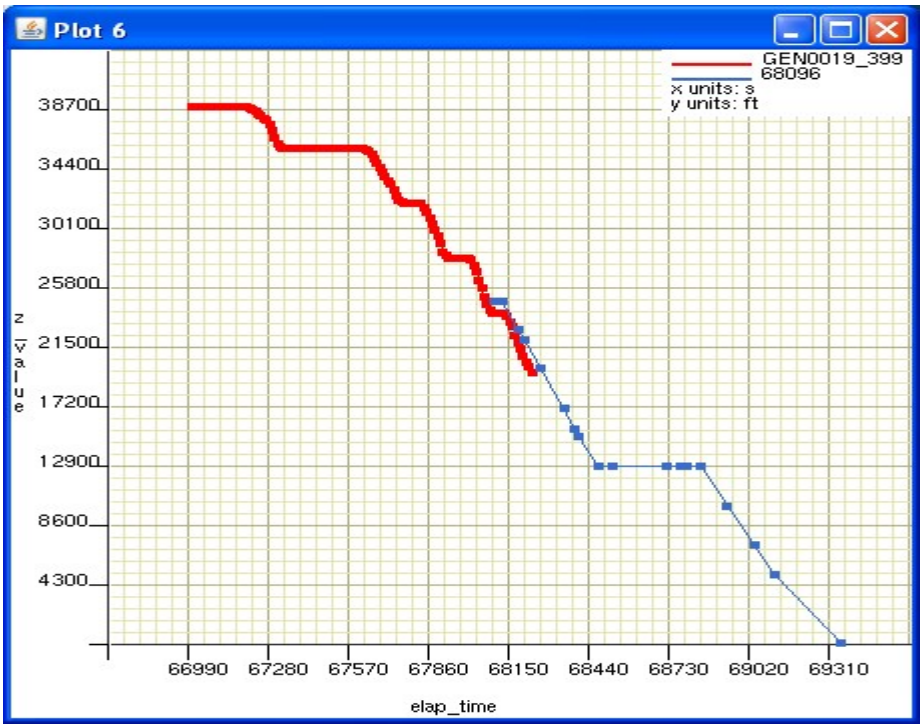

Figure 6: ERAM Trajectory #16 for Flight GEN0019

3.1.4.1.2 URET Trajectory Data

URET track data for this flight extend from FL390 at 67,000 seconds UTC to 13,000 feet at 68,620 seconds UTC. URET has more track data than ERAM. The track has 5 leveling off steps.

URET generates 12 trajectories. The average vertical error for all of the active portions of the trajectories is 2022 feet. The first trajectory (Figure 7) correctly predicts 2 of the interim descent altitudes but misses 3 and is a slightly late with the descent. The third trajectory (Figure 8) predicts 4 of the 5 interim altitudes and continues to be a bit late with the descent. The eighth trajectory (Figure 9) misses a short leveling off segment but fairly accurately predicts the descent. The 4 trajectories after the eighth duplicate the later sections of the eighth.

3.1.4.1.3 Altitude Metric Data for Flight GEN0019

The altitude prediction accuracy metrics described in Section 3.1.2 were applied to the trajectories generated for this flight by ERAM and by URET. The sampled window data obtained is given in the following tables.

3.1.4.1.3.1 ERAM Altitude Metric Data

The following two tables give the results of applying the metric to the ERAM data. The tables show that 39 sample windows were obtained from the URET sampled trajectories for this flight. Of the 39 windows, all were transition windows, and of the 39, 14 had at least one predicted altitude report in error by more than 1500 feet.

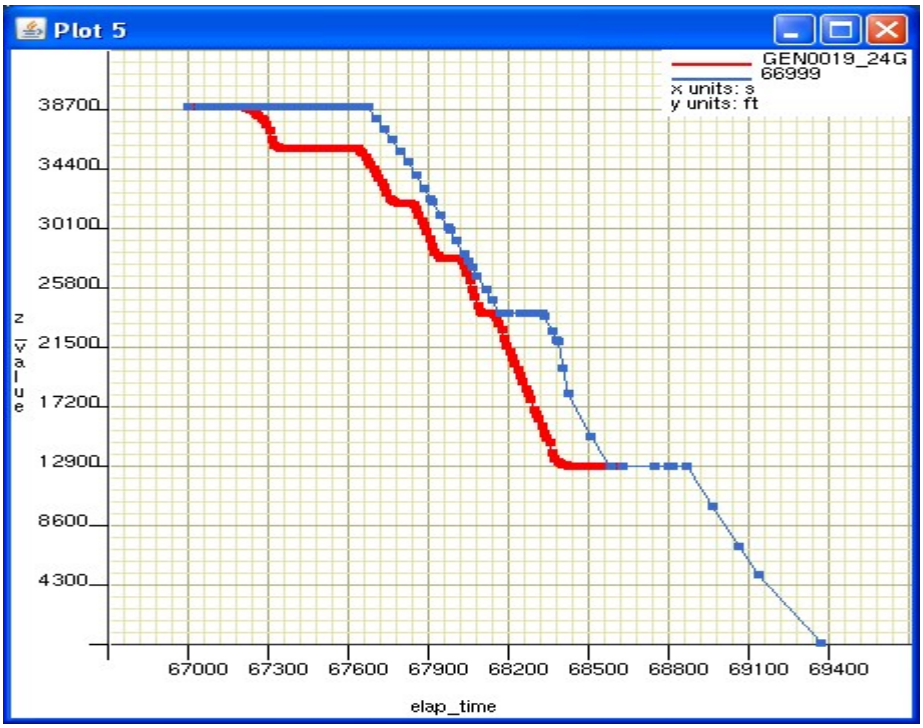

Figure 7: URET Trajectory #1 for Flight GEN0019

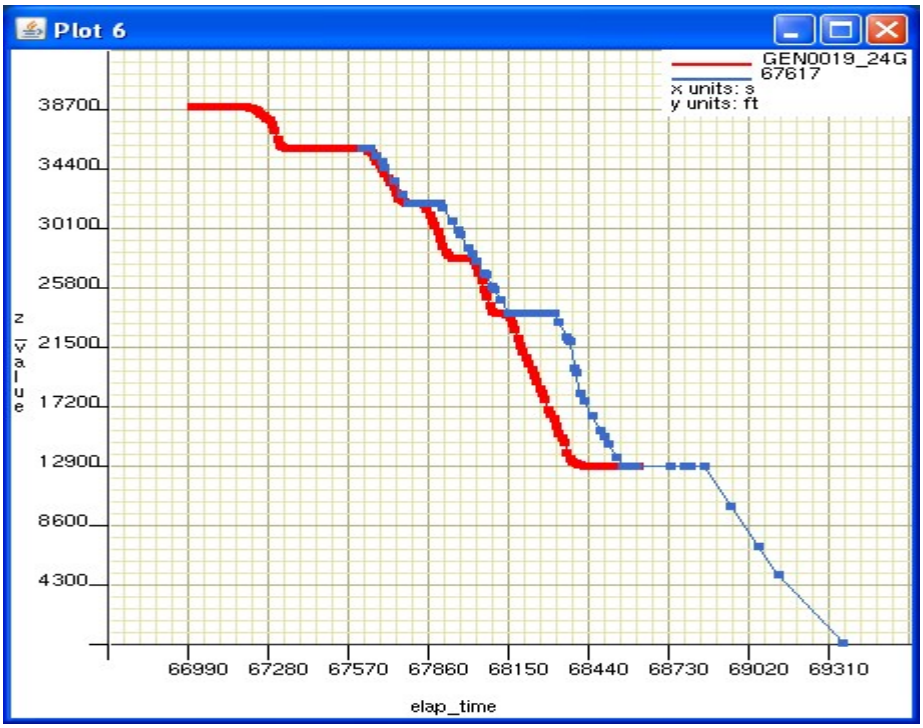

Figure 8: URET Trajectory #3 for Flight GEN0019

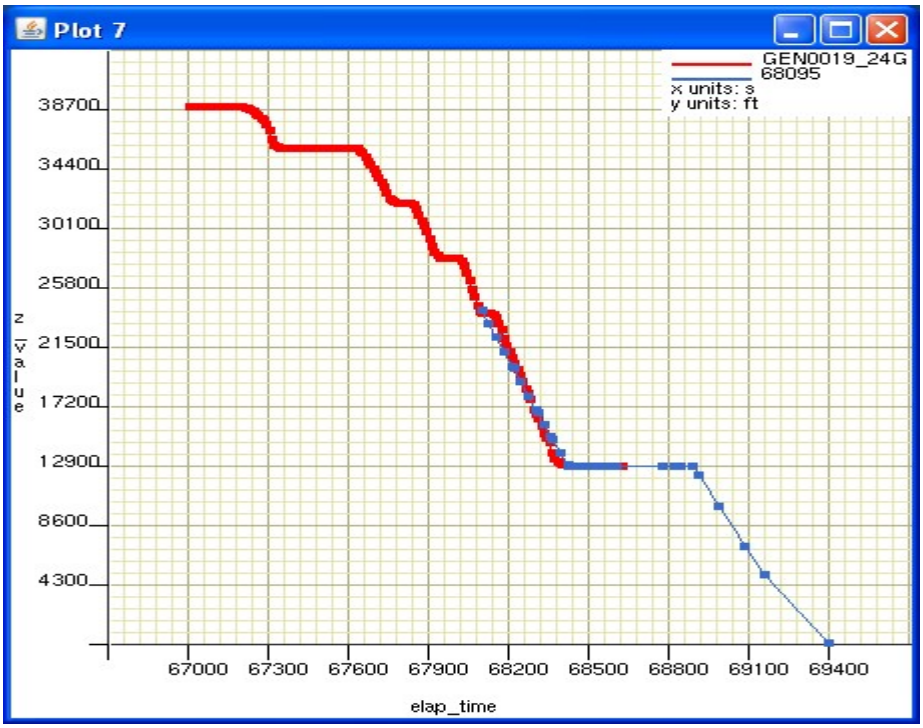

Figure 9: URET Trajectory #8 for Flight GEN0019

Table 3: ERAM Altitude Prediction Window Summary for GEN0019

Number of Sample Windows	Number of Level Sample Windows	Number of Trans. Sample Windows	Number of Level Windows >500 ft error	Number of Trans. Windows >1500 ft error
39	0	39	0	14

3.1.4.1.3.2 URET Altitude Metric Data

The following two tables give the results of applying the metric to the URET data. The tables show that 40 sample windows were obtained from the URET sampled trajectories for this flight. Of the 40 windows, all were transition windows, and of the 40, 9 had at least one predicted altitude report in error by more than 1500 feet.

3.1.4.2 AIR0003

In the scenario this aircraft, a Canadair Regional Jet flying from Greenville SC to Washington/Dulles, climbs to FL280 and then descends from FL280 with interim holding altitudes at FL250, FL220, 10,000 feet, 5,000 feet, and 4,000 feet.

Table 4: ERAM Altitude Prediction Windows for GEN0019 - Maximum Errors

Sample#	SampleTime	WindowType	MaxVert
1	67040	trans:	8056.777
2	67070	trans:	5381.983
3	67100	trans:	5381.983
4	67130	trans:	1015.585
5	67160	trans:	1015.585
6	67190	trans:	15000
7	67220	trans:	15000
8	67250	trans:	15000
9	67280	trans:	10817.01
10	67310	trans:	8049.286
11	67340	trans:	8049.286
12	67370	trans:	8049.286
13	67400	trans:	0
14	67430	trans:	0
15	67460	trans:	0
16	67490	trans:	0
17	67520	trans:	0
18	67550	trans:	0
19	67580	trans:	0
20	67610	trans:	0
21	67640	trans:	705.771
22	67670	trans:	705.771
23	67700	trans:	705.771
24	67730	trans:	705.771
25	67760	trans:	595.3891
26	67790	trans:	0
27	67820	trans:	629.9887
28	67850	trans:	629.9887
29	67880	trans:	2592.738
30	67910	trans:	1350
31	67940	trans:	426
32	68000	trans:	1368.345
33	68030	trans:	1368.345
34	68060	trans:	1368.345
35	68090	trans:	900
36	68120	trans:	1362.694
37	68150	trans:	1763.471
38	68180	trans:	1525
39	68210	trans:	1525

Table 5: URET Altitude Prediction Window Summary for GEN0019

Number of Sample Windows	Number of Level Sample Windows	Number of Trans. Sample Windows	Number of Level Windows >500 ft error	Number of Trans. Windows >1500 ft error
40	0	40	0	9

Table 6: URET Altitude Prediction Windows for GEN0019 - Maximum Errors

Sample#	SampleTime	WindowType	MaxVert
1	67040	trans:	0
2	67070	trans:	0
3	67100	trans:	0
4	67130	trans:	0
5	67160	trans:	0
6	67190	trans:	1165.3236
7	67220	trans:	1165.3236
8	67250	trans:	1165.3236
9	67280	trans:	1165.3236
10	67310	trans:	650
11	67340	trans:	0
12	67370	trans:	0
13	67400	trans:	0
14	67430	trans:	0
15	67460	trans:	0
16	67490	trans:	0
17	67520	trans:	0
18	67550	trans:	0
19	67580	trans:	0
20	67610	trans:	0
21	67640	trans:	483.7526
22	67670	trans:	483.7526
23	67700	trans:	483.7526
24	67730	trans:	483.7526
25	67760	trans:	483.7526
26	67790	trans:	0
27	67820	trans:	325.5415
28	67850	trans:	325.5415
29	67880	trans:	1623.1245
30	67910	trans:	1623.1245
31	67940	trans:	1623.1245
32	67970	trans:	1623.1245
33	68000	trans:	1623.1245
34	68030	trans:	2059.581
35	68060	trans:	2059.581
36	68090	trans:	710.2432

37	68120	trans:	1397.3925
38	68150	trans:	1672.4245
39	68180	trans:	1672.4245
40	68210	trans:	361.9154

3.1.4.2.1 ERAM Trajectory Data

The ERAM track data for this flight extend from climbing at 72,330 seconds UTC at 24,400 feet to descending at 73750 seconds UTC at 10,400 feet. There is a short cruise segment at FL330 and two step-down interim altitudes at FL250 and FL220.

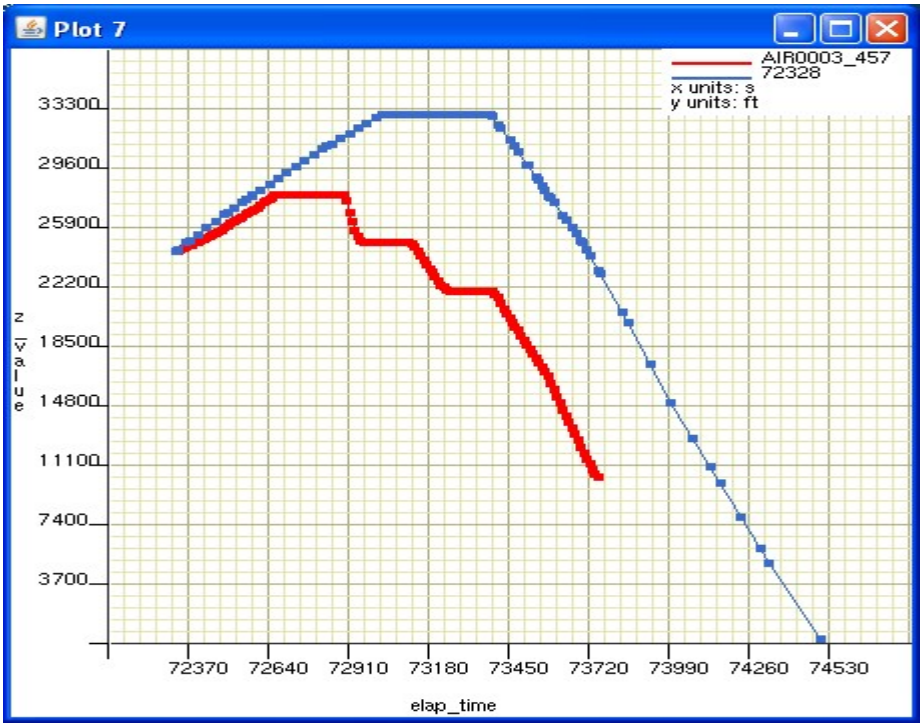

Figure 10: ERAM Trajectory #1 for Flight AIR0003

ERAM generates 20 trajectories. The first trajectory (Figure 10), built at 72328 seconds UTC, climbs the aircraft to FL330 (too high) and then soon after descends (late) the flight to airport level without any interim holding altitudes. The average vertical error is 5840 feet. The following trajectories, #2 though #9 (Figure 11) through #18 (Figure 12) keep predicting that the aircraft is going to climb to a higher altitude, although this higher altitude keeps getting lower as the aircraft descends. Finally, with trajectory #19 (Figure 13), ERAM predicts that the aircraft will continue descending.

3.1.4.2.2 URET Trajectory Data

The URET track data for this flight extend from an altitude of 24,300 feet at a time of 72330 seconds UTC to an altitude of 3900 feet at a time of 74350 seconds UTC. The aircraft initially climbs to FL 280 and then descends with interim altitudes at flight levels 250 and 220, 10,000 feet, 5,000 feet, and 4,000 feet.

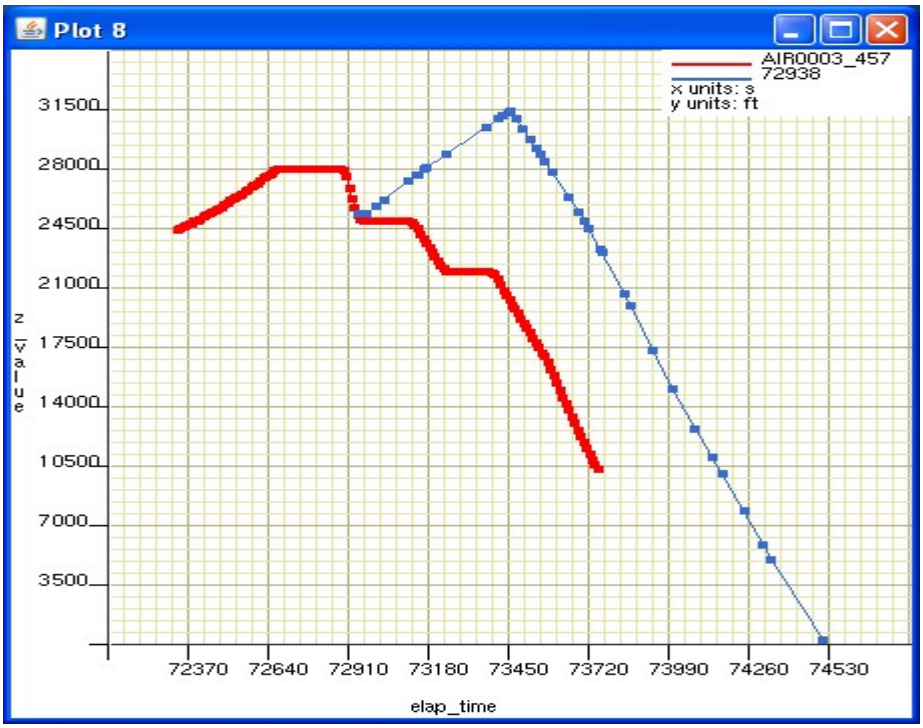

Figure 11: ERAM Trajectory #9 for Flight AIR0003

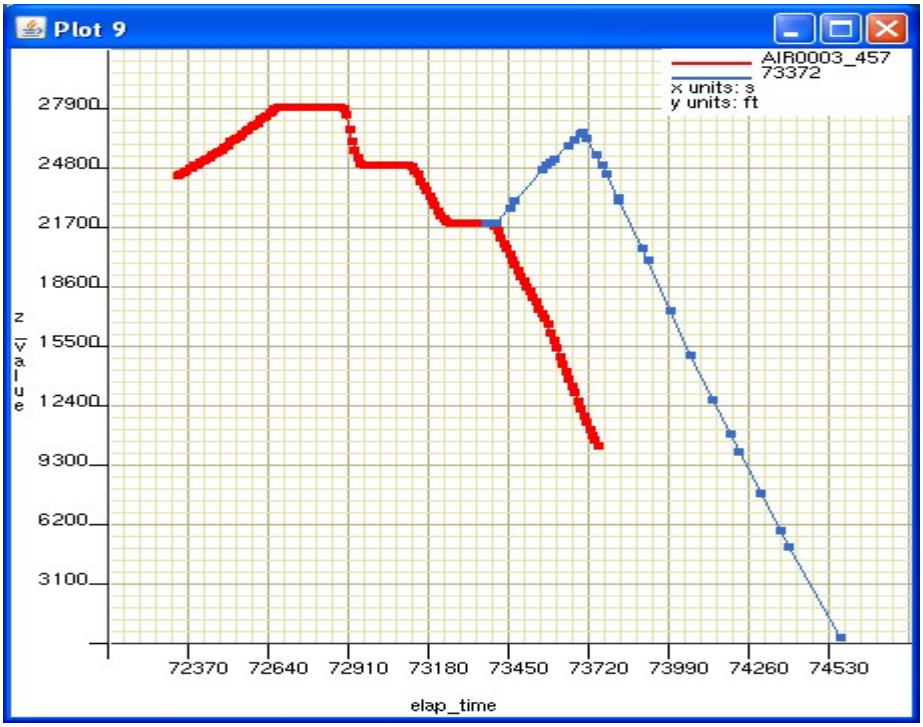

Figure 12: ERAM Trajectory #18 for Flight AIR0003

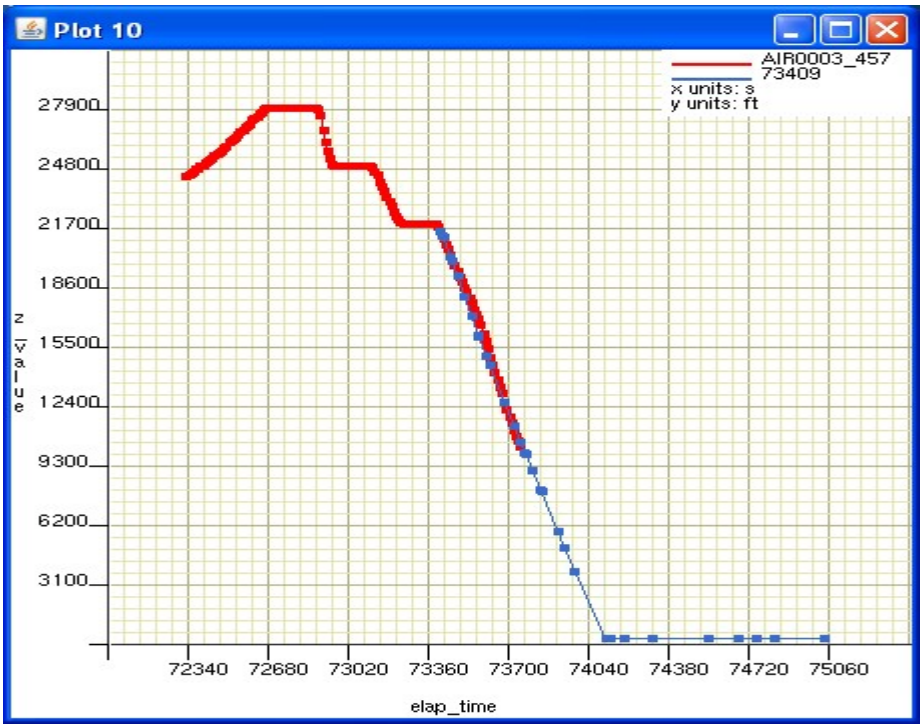

Figure 13: ERAM Trajectory #19 for Flight AIR0003

URET generates 13 trajectories. The average vertical error for all of the active portions of the trajectories is 1789 feet. The first trajectory (Figure 14) initially climbs the aircraft to 29,800 feet and then descends the aircraft, with interim steps at FL 250 and 10,000 feet. It misses the interim altitudes at FL 220, 5,000 feet and 4,000 feet. All of the later trajectories predict only descents. Trajectories #2 (Figure 15) through #8 also miss the same interim altitudes and descend late from FL 250. Trajectory #9 (Figure 16) predicts the interim altitude at FL 220 and descends correctly from FL 250 and FL 220, but stays at 10,000 feet too long, before descending directly towards the destination airport, skipping the interim altitudes at 5,000 and 4,000 feet. Trajectories #10, #11, and #12 all follow the altitude profile of trajectory #9. The last trajectory, #13 (Figure 17), which starts below 10,000 feet at 9500 feet, continues the descent below 10,000 feet, skips the interim altitudes at 5,000 feet and 4,000 feet, and levels off at 312 feet (the Dulles airport elevation).

3.1.4.2.3 Altitude Metric Data for Flight AIR0003

The altitude prediction accuracy metrics described in Section 3.1.2 were applied to the trajectories generated for this flight by ERAM and by URET. The sampled window data obtained is given in Table 7 though Table 10.

3.1.4.2.3.1 ERAM Altitude Metric Data

Table 7 and Table 8 give the results of applying the metric to the ERAM data. The tables show that 41 sample windows were obtained from the URET sampled trajectories for this flight. Of the 41 windows, all were transition windows, and of the 41, 12 had at least one predicted altitude report in error by more than 1500 feet.

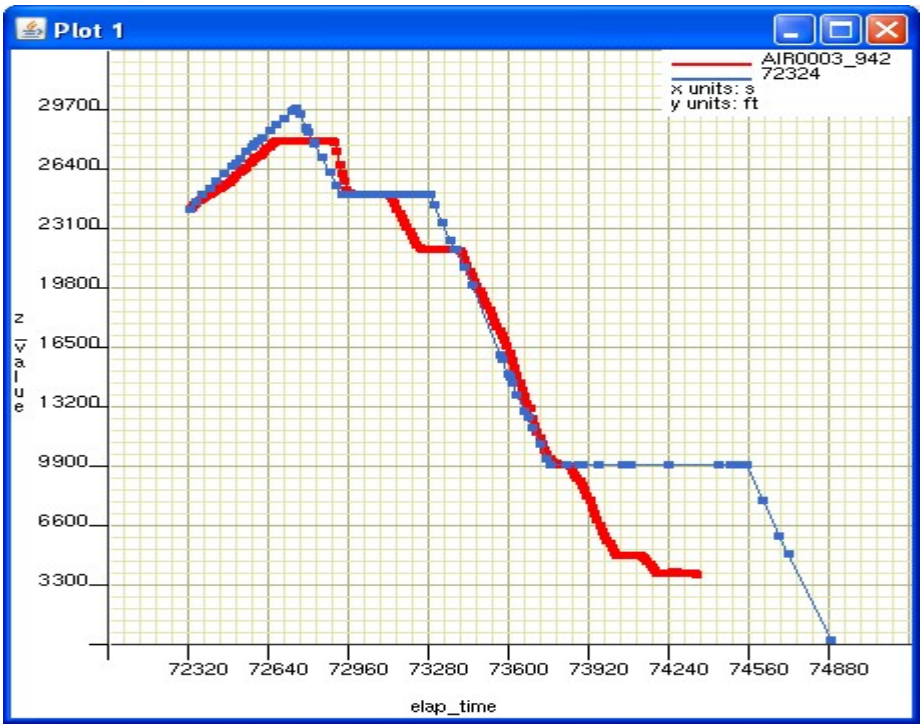

Figure 14: URET Trajectory #1 for Flight AIR0003

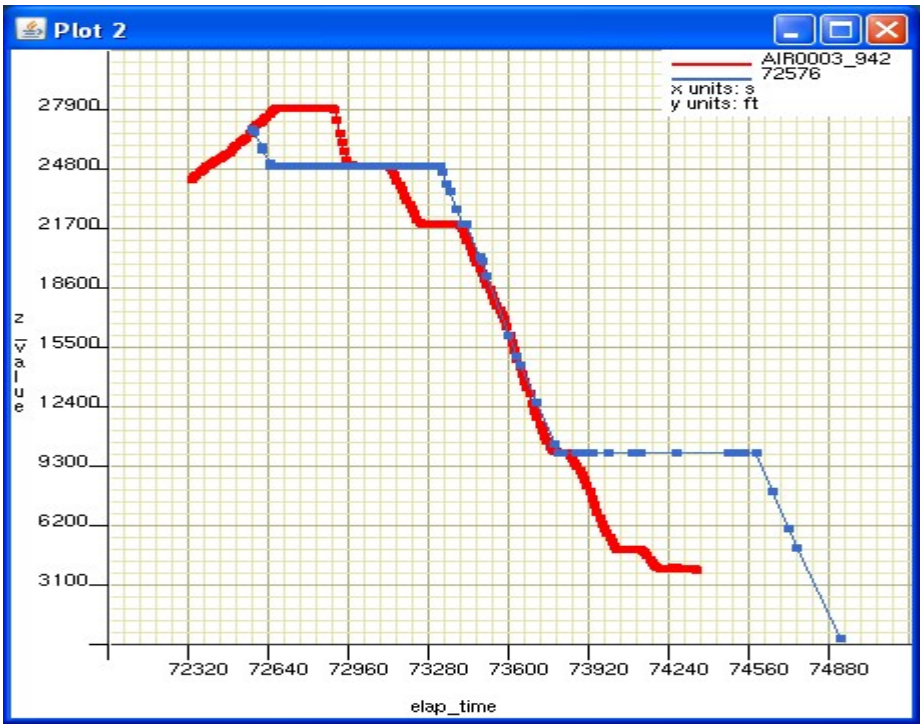

Figure 15: URET Trajectory #2 for Flight AIR0003

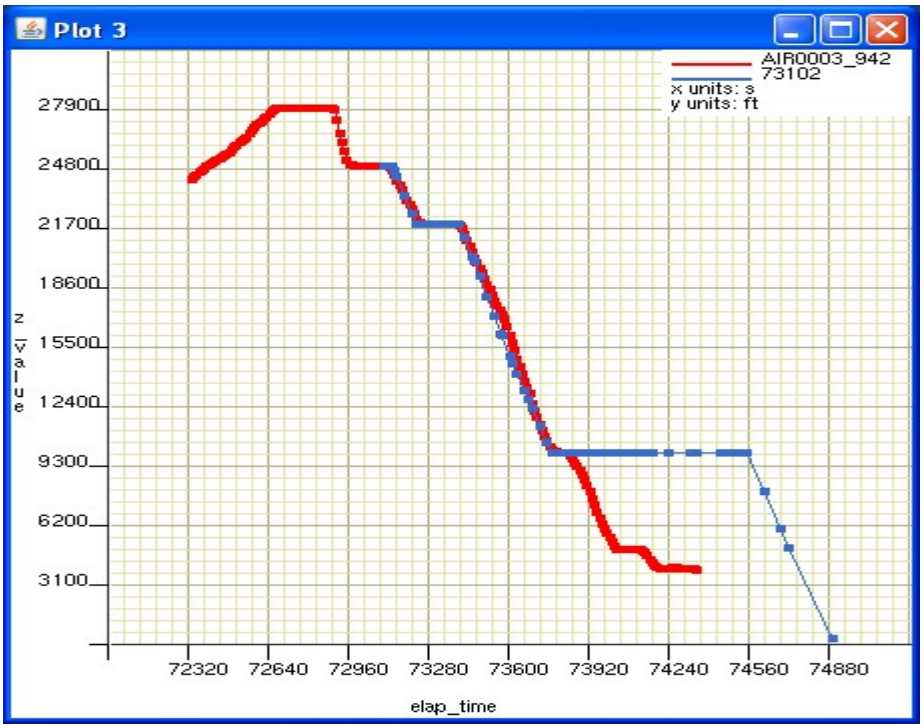

Figure 16: URET Trajectory #9 for Flight AIR0003

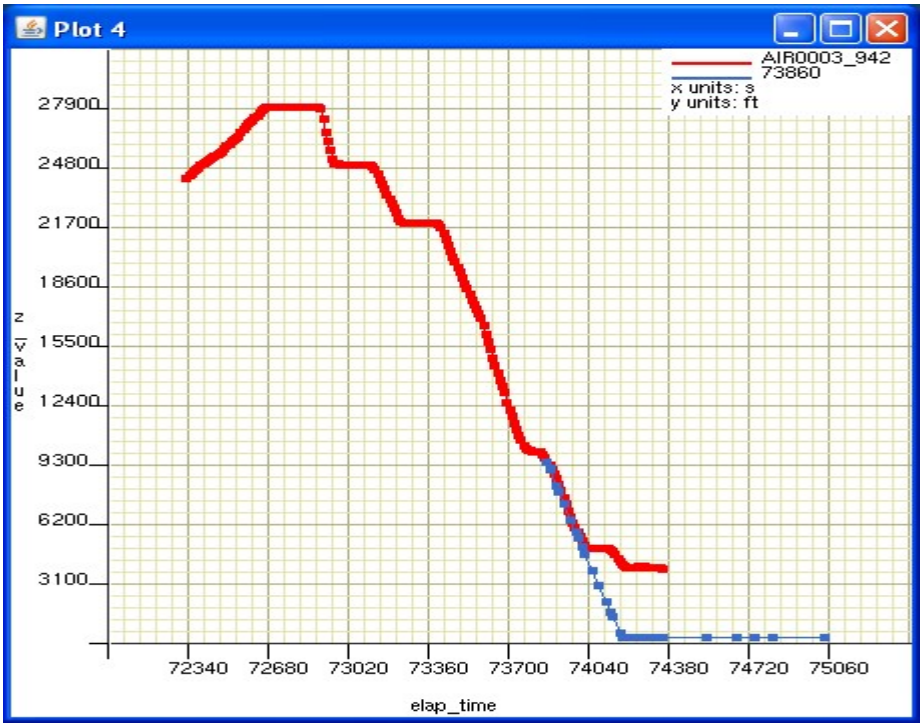

Figure 17: URET Trajectory #13 for Flight AIR0003

Table 7: ERAM Altitude Prediction Window Summary for AIR0003

Number of Sample Windows	Number of Level Sample Windows	Number of Trans. Sample Windows	Number of Level Windows >500 ft error	Number of Trans. Windows >1500 ft error
41	0	41	0	12

Table 8: ERAM Altitude Prediction Windows for AIR0003 - Maximum Errors

Sample#	SampleTime	WindowType	MaxVert
1	72370	trans:	955.4267
2	72400	trans:	955.4267
3	72430	trans:	955.4267
4	72460	trans:	955.4267
5	72490	trans:	955.4267
6	72520	trans:	955.4267
7	72550	trans:	955.4267
8	72580	trans:	1018.8371
9	72610	trans:	1018.8371
10	72640	trans:	1018.8371
11	72670	trans:	1018.8371
12	72880	trans:	4785.8992
13	72910	trans:	4398.2417
14	72940	trans:	2072.7343
15	72970	trans:	1509.7338
16	73000	trans:	1509.7338
17	73030	trans:	1509.7338
18	73060	trans:	208.5984
19	73090	trans:	208.5984
20	73120	trans:	5905.6393
21	73150	trans:	4828.1626
22	73180	trans:	4041.0703
23	73210	trans:	3197.0542
24	73240	trans:	2251.3375
25	73270	trans:	1468.0707
26	73300	trans:	1468.0707
27	73330	trans:	1468.0707
28	73360	trans:	1468.0707
29	73390	trans:	5156.1993
30	73420	trans:	500.8651
31	73450	trans:	500.8651
32	73480	trans:	500.8651
33	73510	trans:	500.8651
34	73540	trans:	1132.5599
35	73570	trans:	1132.5599
36	73600	trans:	1132.5599
37	73630	trans:	1132.5599

38	73660	trans:	1132.5599
39	73690	trans:	1132.5599
40	73720	trans:	1132.5599
41	73750	trans:	1132.5599

3.1.4.2.3.2 URET Altitude Metric Data

The following two tables give the results of applying the metric to the URET data. The tables show that 40 sample windows were obtained from the URET sampled trajectories for this flight. Of the 41 windows, all were transition windows, and of the 41, 3 had at least one predicted altitude report in error by more than 1500 feet.

Table 9: URET Altitude Prediction Window Summary for AIR0003

Number of Sample Windows	Number of Level Sample Windows	Number of Trans. Sample Windows	Number of Level Windows >500 ft error	Number of Trans. Windows >1500 ft error
41	0	41	0	3

Table 10: URET Altitude Prediction Windows for AIR0003 – Maximum Errors

Sample#	SampleTime	WindowType	MaxVert
1	72370	trans:	917.0276
2	72400	trans:	917.0276
3	72430	trans:	917.0276
4	72460	trans:	917.0276
5	72490	trans:	917.0276
6	72520	trans:	917.0276
7	72550	trans:	917.0276
8	72580	trans:	2991
9	72610	trans:	2991
10	72640	trans:	2253.0353
11	72670	trans:	788.4306
12	72880	trans:	687.9566
13	72910	trans:	687.9566
14	72940	trans:	687.9566
15	72970	trans:	54
16	73000	trans:	0
17	73030	trans:	0
18	73060	trans:	0
19	73090	trans:	0
20	73120	trans:	222.6932
21	73150	trans:	222.6932
22	73180	trans:	222.6932
23	73210	trans:	222.6932
24	73240	trans:	58
25	73270	trans:	0

26	73300	trans:	0
27	73330	trans:	0
28	73360	trans:	0
29	73390	trans:	311.3241
30	73420	trans:	311.3241
31	73450	trans:	240.9872
32	73480	trans:	109.1978
33	73510	trans:	105.387
34	73540	trans:	372.9521
35	73570	trans:	372.9521
36	73600	trans:	372.9521
37	73630	trans:	372.9521
38	73660	trans:	372.9521
39	73690	trans:	372.9521
40	73720	trans:	372.9521
41	73750	trans:	327.9451

3.2 Strategic Conflict Alerts – Aircraft to Aircraft

LM ran both URET and ERAM using the CPAT supplied, time shifted scenarios. CPAT analyzed the conflict prediction performance of URET and ERAM using the output data files supplied by LM This section presents the CPAT analysis results. Later sections will compare the CPAT results to LM's OFPD and EOPD results.

3.2.1 Aircraft to Aircraft Conflict Alerts

A pair of aircraft are said to be in conflict whenever they are close together both horizontally, that is less than five nautical miles in en route airspace, and vertically, less than 1000 feet. A strategic conflict alert, as distinct from a tactical conflict, is a conflict predicted to occur more than five minutes and less than 20 minutes into the future.

URET and ERAM are required to predict aircraft to aircraft conflicts. When they do so correctly, the event is called a valid alert. When they do not predict a conflict which actually occurs, the event is called a missed alert. When they predict a conflict but none occur, the event is labeled a false alert. In a test scenario there are artifacts which require that a conflict or an alert be discarded and not counted in the performance calculations.

It is useful to define an encounter between two aircraft to evaluate false alerts. Two aircraft have an encounter when they get close together but not close enough to have a conflict. An encounter is characterized by the minimum separation of the aircraft.

3.2.2 Ground Truth from the Radar Track Data

The actual aircraft to aircraft conflicts in the time shifted scenario were determined from the radar track data. The conflicts for the URET test run were determined from the radar track data produced by the Host. The conflicts for the ERAM test run were determined from the radar track data produced by ERAM. The obvious errors in the radar track data in the scenarios were fixed and then small gaps in data were interpolated. Then the actual conflicts in the scenario were determined by checking the distances between all of the aircraft.

3.2.3 CPAT Processing Steps

To use the CPAT processing tools it was first necessary to convert the LM supplied output data into the formats required by the CPAT tools. This was done and then the data was run through the tools. The processing steps required for the data analysis are described in Section 4, Appendix A – Processing to Analyze Conflict Predictions, and Appendix E – Overview of Conflict Prediction Performance Comparison.

3.2.4 Alert Type Definitions

Appendix A describes the logic and processing involved in classifying conflicts and encounters. There are four basic classifications; Valid Alerts, Missed Alerts, False Alerts, and Discards. The following table is a short summary of the various results defined within the four basic classification types. This summary can be used for reference during examination of the aircraft-to-aircraft and aircraft-to-airspace analysis and examples that follow. Each sub-classification sdoes not provide the source of the error but does provide some insight into each error.

Table 11: Alert Type Definitions

GENERAL TYPE	DETAILED TYPE	DESCRIPTION
VALID	STD_VA	Standard Valid Alert
	LATE_VA	Late Valid Alert caused by a "pop-up" conflict[7]
MISSED	NO_CALL_MA	Missed Alert due to no call (no alert at all)
	LATE_MA	Late Missed Alert, an alert presented within the MWT[8]
FALSE	STD_FA	Standard False Alert
	RETRACT_FA	Retracted False Alert – Alert removed earlier than its predicted conflict start time
DISCARDS	NO_CALL_DISCARD	Missed Alert discarded because of out of range adherence age[9]
	NO_TRK_FA_DISCARD_A	Track data not available
	NO_TRK_DISCARD_B	PCST not after aircraft track data start time(s)
	NO_TRK_DISCARD_C	PCST not before aircraft track data end time(s)
	NO_TRK_DISCARD_D	No track data at PCST (gap in track data)
	NO_ADHER_FA_DISCARD	Out of range adherence age at PCST
	CLR_FA_DISCARD_A	Primary clearance(s) near Alert end time
	CLR_FA_DISCARD_B	Secondary clearance(s) near Alert end time
	CLF_FA_DISCARD	Alert start time after ACST

[7] A pop-up conflict occurs when either aircraft (1) is near the start of its track data, (2) is near an interim altitude, or (3) the conflict starts soon after an air traffic clearance.
[8] MWT is the minimum warning time requirement threshold. For the RFR, it was set to 5 minutes.
[9] Adherence age is the length of time that the current position has been within a threshold laterally and vertically to the automation's known clearances. If the adherence age is beyond a threshold time, the conflict can be safely excused for strategic predictions – more tactical tools would be relevant.

3.2.5 Analysis of URET Aircraft to Aircraft Conflict Alert Data

The time shifting of the radar track data created aircraft to aircraft encounters and conflicts in the scenario. CPAT processing identified and characterized the encounters and conflicts. There are 2275 aircraft, 20,860 encounters and 249 conflicts in the scenario.

3.2.5.1 Matched Conflicts (Valid Alerts)

Within the URET data, there were 50 STD_VA results and 183 LATE_VA results. Of these 233 Valid Alerts, 22 provided more than 10 minutes (300 seconds) of alert time, averaging 909.5 seconds. The remaining 211 provided an average of 154.5 seconds of alert time.

3.2.5.2 Unmatched Conflicts (Missed Alerts)

Within the URET data, there were 2 NO_CALL_MA results and 4 LATE_MA results, for a total of 6 Missed Alerts. Dividing the number of Missed Alerts by the sum of Missed Alerts plus Valid Alerts resulted in a Missed Alert rate of 0.0251, or 2.51%.

3.2.5.3 Unmatched Alerts (False Alerts)

Within the URET data, there were 376 STD_FA results and 276 RETRACT FA results for a total of 652 False Alerts. The rate of False Alerts was determined independently over a number of "bins" of minimum encounter distance in NM, by dividing the number of False Alerts by the number of encounters for each bin. Table 12 below contains the URET False Alert rates for the different bins.

Table 12: URET False Alert Rates

Range	False Alerts	False Alert Rate
0 < d <10	532	0.152 (15.2%)
10 < d <15	62	0.0291 (2.91%)
15 < d <23	32	0.00759 (0.759%)
d > 23	26	0.00236 (0.236%)

3.2.5.4 Discards

Within the URET data, 1,653 Discard conditions were identified as tabulated in Table 13.

Table 13: URET Discard Rates

Discard Type	Number
NO_TRK_FA_DISCARD_A	0
NO_TRK_FA_DISCARD_B	253
NO_TRK_FA_DISCARD_C	555
NO_TRK_FA_DISCARD_D	6
NO_ADHER_FA_DISCARD	483
CLR_FA_DISCARD_A	376
CLR_FA_DISCARD_B	0
CFL_FA_DISCARD	35
NO_CALL_DISCARD	7
LATE_DISCARD	3

3.2.6 Analysis of ERAM Aircraft to Aircraft Alert Data

The time shifting of the radar track data created aircraft to aircraft encounters and conflicts in the scenario. CPAT processing identified and characterized the encounters and conflicts. There are 2227 aircraft, 19,752 encounters and 230 conflicts in the scenario.

3.2.6.1 Matched Conflicts (Valid Alerts)

Within the ERAM data, there were 41 STD_VA results and 163 LATE_VA results. The 204 Valid Alerts had 14 that provided more than 10 minutes (300 seconds) alert time, averaging 994.9 seconds. The remaining 190 provided an average of 144.4 seconds of alert warning time.

3.2.6.2 Unmatched Conflicts (Missed Alerts)

Within the ERAM data, there were 12 NO_CALL_MA results and 2 LATE_MA results, for a total of 14 Missed Alerts. Dividing the number of Missed Alerts by the sum of Missed Alerts plus Valid Alerts resulted in a Missed Alert rate of 0.0642 or 6.42%.

3.2.6.3 Unmatched Alerts (False Alert)

Within the ERAM data, there were 552 STD_FA results and 473 RETRACT_FA results for a total of 1025 False Alerts. The rate of False Alerts was determined independently over a number of "bins" of minimum encounter distance in NM, by dividing the number of False Alerts by the number of encounters for each bin. Table 14 below contains the ERAM False Alert rates for the different bins.

Table 14: ERAM False Alert Rates

Range	False Alerts	False Alert Rate
0 < d <10	745	0.230 (23.0%)
10 < d <15	129	0.0634 (6.34%)
15 < d <23	66	0.0162 (1.62%)
d > 23	85	0.00820 (0.820%)

3.2.6.4 Discards

Within the ERAM data, 2,295 Discard conditions were identified as tabulated in Table 15.

Table 15: ERAM Discard Rates

Discard Type	Number
NO_TRK_FA_DISCARD_A	50
NO_TRK_FA_DISCARD_B	275
NO_TRK_FA_DISCARD_C	515
NO_TRK_FA_DISCARD_D	269
NO_ADHER_FA_DISCARD	742
CLR_FA_DISCARD_A	430
CLR_FA_DISCARD_B	0
CFL_FA_DISCARD	55
NO_CALL_DISCARD	9
LATE_DISCARD	3

3.2.7 Examples

This section presents examples tailored towards each of the 4 aircraft-to-aircraft conflict testing requirements in the ERAM A-Level Test Plan (ERD1879-C1/4). In each example, the results obtained with the ERAM track data correlate to the relevant test requirement, with the results obtained with URET track data not necessarily agreeing.

It is also important to remind the reader that these are examples taken directly from the test data used in the RFR and are not real events only based on real traffic data altered by time-shifting.

3.2.7.1 Valid Alert Greater Than 10 Minutes Warning (ERD1879-C1)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.2.7.1.1 Example Description

In this simulation example, Flight AIR0001 is a Boeing MD-80 series aircraft flying from Palm Beach FL to Laguardia airport in New York, with intermediate fixes at PERMT, ILM, PXT and KORRY3. Flight AIR0027 is an Airbus A300 series aircraft flying from Orlando FL to Boston MA, with intermediate fixes at CHS, JFK, and ORW3. Figure 18 depicts the flight paths of these two aircraft immediately before and after the conflict that occurred. The labeled black squares are major airports. Their names are listed at the end of the report (List of Selected Airport Identifiers, page 77. Note that although Figure 18 utilizes URET track data to plot the aircraft positions, an equivalent figure utilizing ERAM track data is virtually identical.

3.2.7.1.2 URET Results

Track data for these two aircraft within Washington Center airspace started at 81330 seconds into the scenario. At this point AIR0001 was flying on a heading of 35 degrees at an assigned altitude of 35000 feet. At this same scenario time, AIR0027 was flying a heading of 5 degrees at its assigned altitude of 35000 feet. AIR0001 remained on a heading within a heading close to 5 degrees up until the start of the conflict, and both aircraft remained at their assigned altitude of 35000 feet thru the duration of the conflict. At 82090 seconds into the scenario, AIR0027 began a slight turn toward the north, reaching a heading of 10 degrees. At 82190 seconds, AIR0027 started turning back southward reaching a heading of 40 degrees at 82300 seconds. AIR0027 remained on a heading close to 40 degrees up to the start of the conflict. At 82420 seconds, the conflict began when the two aircraft closed to 4.96 (nautical) miles and 0 feet. Closest approach, horizontally, occurred at 82490 seconds with a separation of 1.82 miles and 0 feet. At 82530 seconds, minimum separation requirements were regained when horizontal separation increased to 5.10 miles, ending the conflict.

Scenario track data related to the conflict was processed graphically using the Wolverine Software Proof graphics animation package[10]. Figure 19 is a Proof screen capture at the time of conflict start, and Figure 20 is a screen capture at the time of conflict end.

Briefly, the Proof graphics screen captures are divided into three parts. The upper left section presents the scenario time, referenced to seconds after Midnight, as well as a 24 hour clock. The upper right section depicts the horizontal plane path of the two aircraft. Circles are drawn around each aircraft if they are closer than 5 nm (conflict) or 30 nm (encounter), depending on which data base is being utilized. The lower section of the Proof screen presents state data for each

[10] Post processing simulation animation software

aircraft (lower left/right sections), relative distances and bearings between the two aircraft (lower center sections), and a graphical depiction of the vertical plane path of the two aircraft versus time. For each aircraft, the colored line in the vertical path display represents actual aircraft altitude, and the black line represents assigned altitude.

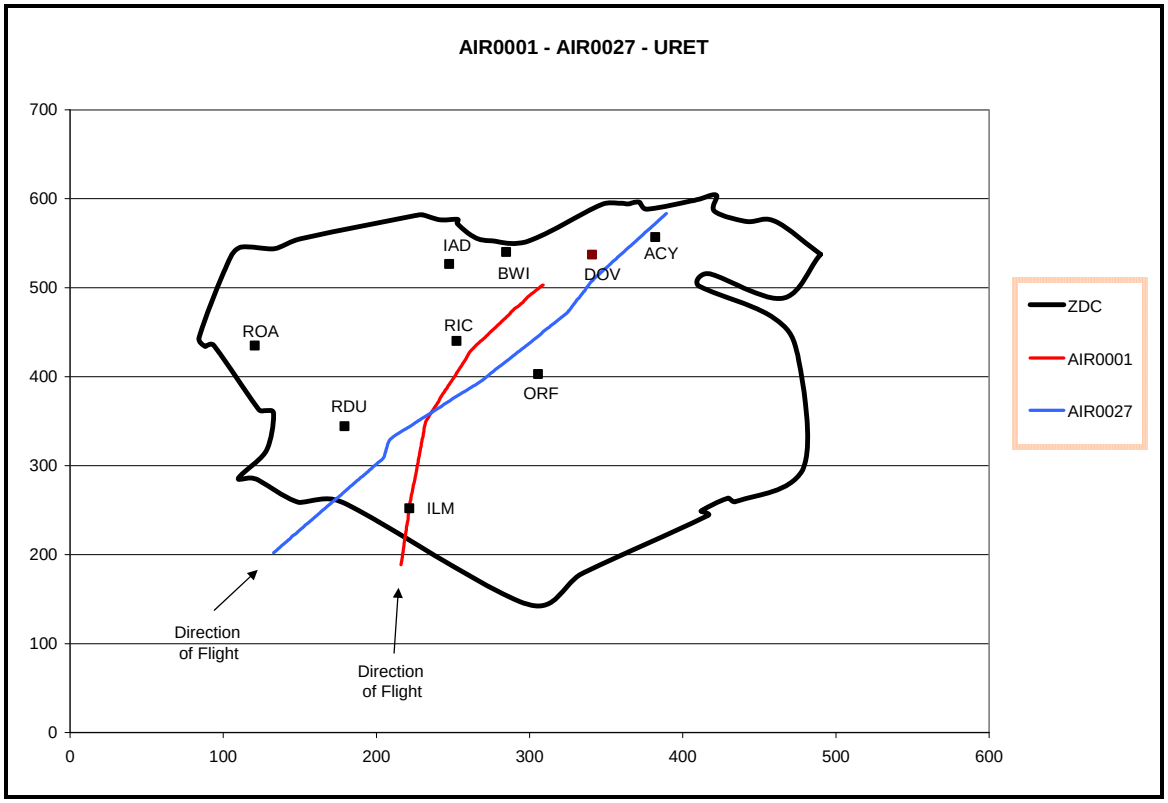

Figure 18: Flight Paths of Flights AIR0001 and AIR0027

As a result of this conflict, URET generated a notification set (NS). The first entry in this NS, the ADD alert, was generated at 81328 seconds into the scenario, and represents the earliest notification by URET of the pending conflict. The actual start of the conflict was at 82420 seconds into the scenario, so the actual warning time provided was 1092 seconds. As the alert was issued and remained active prior to the start of the conflict, this example was placed in the STD_VA category.

3.2.7.1.3 ERAM Results

ERAM track data for both aircraft was first available at 81330 seconds into the scenario. The flight paths of the two aircraft up thru the conflict were virtually identical using ERAM track data as when using URET track data. The conflict began at 82420 seconds when the two aircraft closed to 4.53 miles and 0 feet. Closest approach, horizontally, occurred at 82480 seconds with a separation of 2.79 miles and 0 feet. At 82540 seconds, minimum separation requirements were regained when horizontal separation increased to 5.09 miles, ending the conflict.

ERAM issued an alert for the conflict at 81327 seconds into the scenario, and continued updates for the alert until the conflict was over. The actual start of the conflict was at 82420 seconds into the scenario, so the actual warning time provided was 1093 seconds. As the alert was issued and

remained active prior to the start of the conflict, this example was placed in the STD_VA category as well.

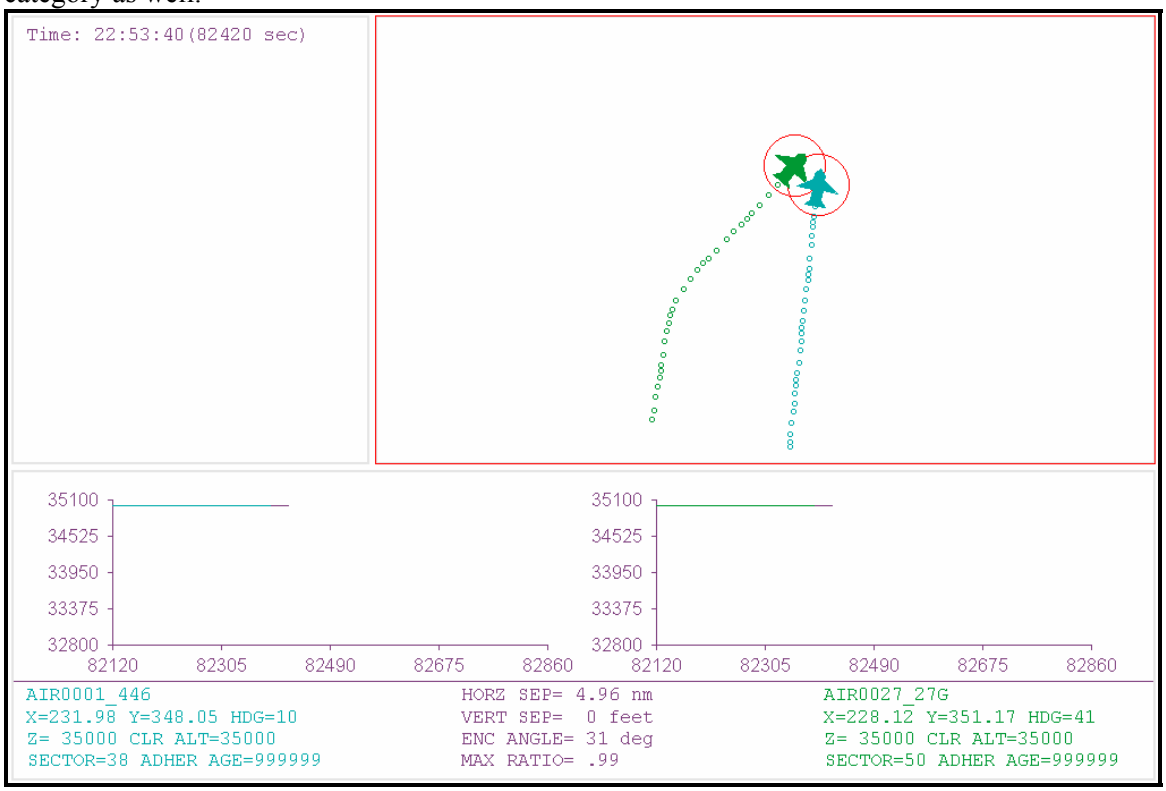

Figure 19: AIR0001 and AIR0027 at Start of Conflict

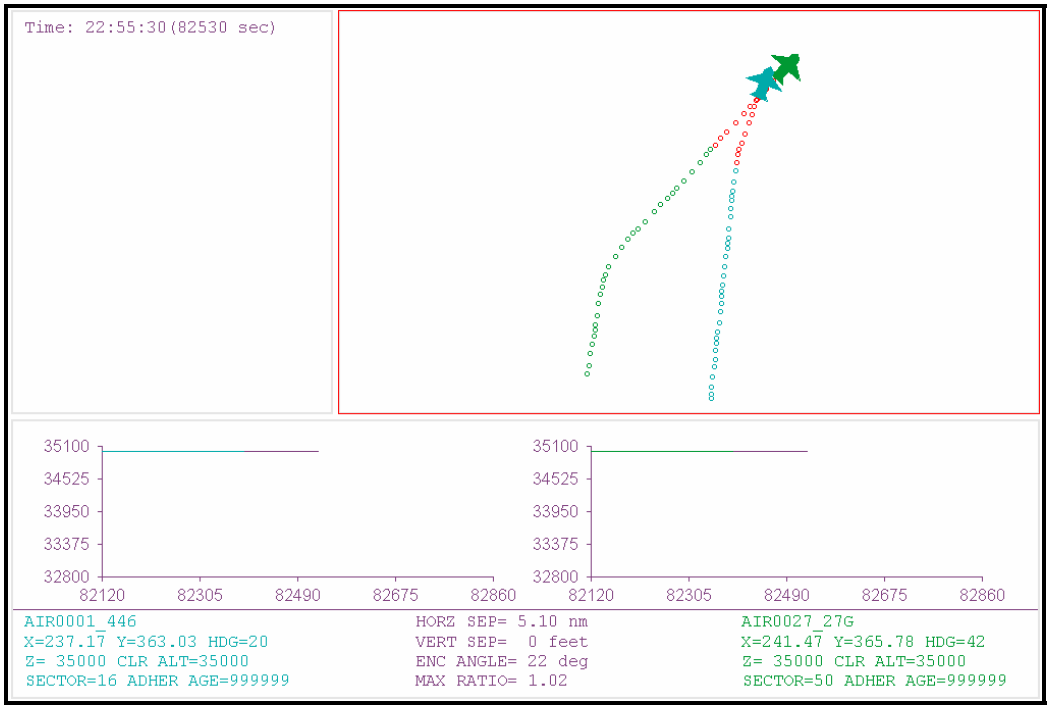

Figure 20: AIR0001 and AIR0027 at End of Conflict

3.2.7.2 Valid Alert Less Than 10 Minutes Warning (ERD1879-C2)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.2.7.2.1 Example Description

In this simulation example, Flight AIR0007 is a Boeing 757 aircraft flying from Palm Beach FL to Newark NJ, with intermediate fixes at RDU, FLOPS, and FAK. Flight AIR0015 is a Dassault Falcon 2000 aircraft flying from Charlotte NC to Boston MA, with intermediate fixes at MERIL, FLOPS and FAK. Figure 21 depicts the flight paths of these two aircraft immediately before and after the conflict that occurred. Plotted aircraft positions are virtually identical using either URET or ERAM track data – URET track data is used here.

3.2.7.2.2 URET Results

Track data for the two aircraft first became available at 66570 seconds into the scenario. At this point, AIR0007 was flying level at an assigned altitude of 35000 feet, and AIR0015 was climbing thru 15000 feet towards an assigned altitude of 33000 feet. As seen in Figure 21, the aircraft horizontally began on a roughly right-angle converging path, then maneuvered to an in-trail configuration by 67300 seconds, being separated by 15+ miles at that point. Subsequently, horizontal separation between the two aircraft slowly decreased until minimum horizontal separation was lost at 68210 seconds.

At 66573 seconds, AIR0015 was issued an interim altitude assignment of 23000 feet, followed by another interim assignment of 25000 feet at 66682 seconds. At 66891 seconds the second interim assignment was briefly removed, then at 66905 another interim altitude of 29000 feet was assigned to AIR0015. At 67044, the interim altitude was removed, and AIR0015 reverted to its original assigned altitude of 33000 feet. AIR0007 remained level at 35000 feet until being issued an assigned altitude of 30000 feet at 66672 seconds. After all of the above altitude assignments, at 67350 seconds into the scenario, AIR0007 was flying level at an assigned altitude of 30000 feet, and AIR0015 was flying level at an assigned altitude of 33000 feet. Horizontal separation between the two aircraft at this point was 16+ miles and decreasing.
At 67919 seconds, AIR0015 was issued an interim altitude clearance to 31000 feet and began its descent, reaching 31000 feet at 68150 seconds. At this point, the two aircraft were separated by 1000 feet and 5.91 miles.

As noted earlier, minimum horizontal separation between the two aircraft was lost at 68210 seconds. At 68271 seconds, AIR0007 was issued an altitude clearance of 27000 feet, and began its descent at 68330 seconds. At 68304 seconds, AIR0015 was issued an interim altitude clearance of 24000 feet, and also began its descent at 68330 seconds. With horizontal separation already below 5 miles, vertical separation dropped below 1000 feet and the conflict between the two aircraft began at this 68330 second point. AIR0015 maintained a higher descent rate than did AIR0007, and by 68430 seconds, minimum vertical separation was regained, ending the conflict. At 68509 seconds, AIR0015 was assigned an interim altitude clearance of 13000 feet, then at 68558 seconds, AIR0007 was issued an interim altitude clearance of 8000 feet. Both aircraft began descending, but minimum vertical separation was not lost prior to the end of track data for the two aircraft.

Scenario track data related to the conflict was processed graphically using the Wolverine Software Proof graphics animation package. Figure 22 depicts the paths of the two aircraft at the beginning of the conflict condition, and Figure 23 depicts the paths at the end of the conflict.

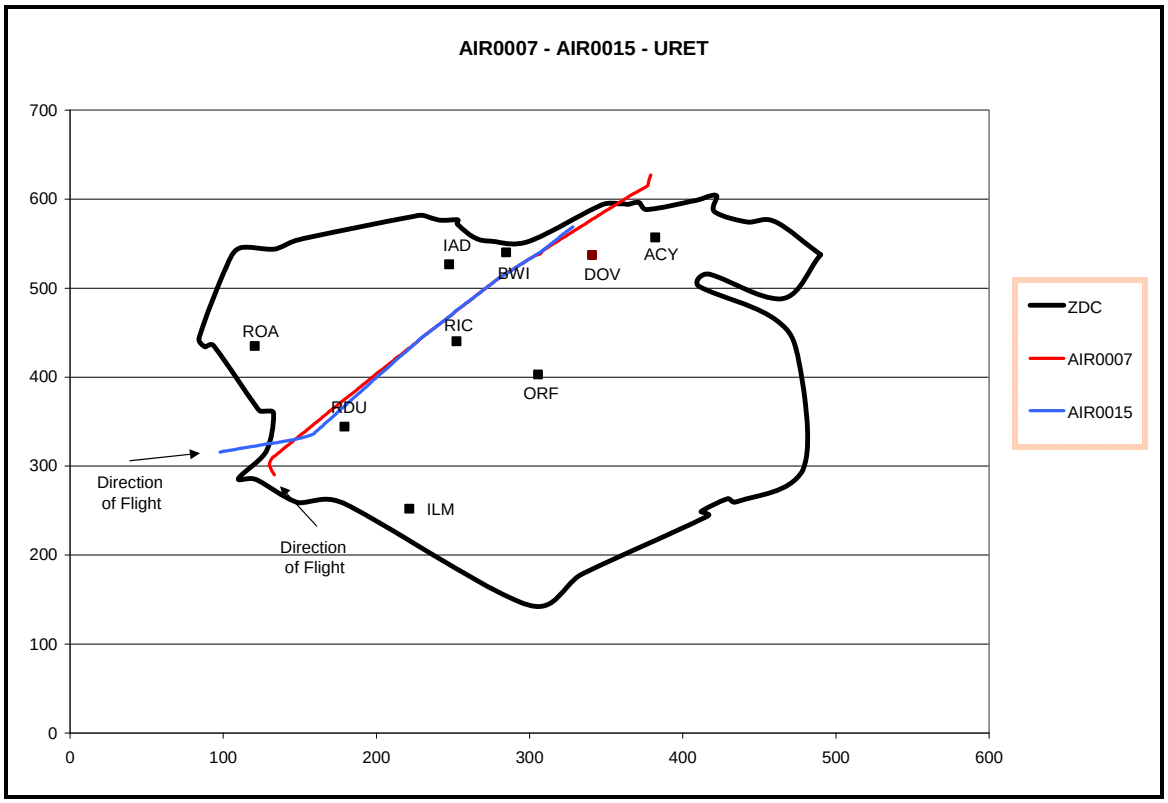

Figure 21: Flight Paths of Flights AIR0007 and AIR0015

The URET system issued two notification sets for the aircraft pair. The first notification set was issued at 67673 seconds into the scenario, and was updated until being deleted at 68496 seconds. The initial "ADD" alert, as well as subsequent updates thru 68272 seconds were labeled as "masked", in that the conflict prediction was based on "final" altitude clearances for the two aircraft that were not in effect as a result of the previously discussed interim altitude clearances that the two aircraft had been receiving. The notification set update at 68305 seconds was not masked, and was considered valid as a result of the interim altitude clearance to AIR0015 at 68304 seconds that resulted in a predicted conflict between the aircraft being based on current cleared altitudes. As a result, the conflict warning time was only 25 seconds, and a LATE_VA result was assigned.

The second notification set was issued at 68496 seconds and had a predicted conflict start time (PCST) of 68825 seconds into the scenario. As the PCST was after the end of track data for the two aircraft, a NO_TRK_FA_DISCARD_C result was assigned.

3.2.7.2.3 ERAM Results

ERAM track data for both aircraft was first available at 66570 seconds into the scenario. The flight paths of the two aircraft up thru the conflict were virtually identical using ERAM track data as when using URET track data. Additionally, the altitude clearance data for the two aircraft was essentially identical when using either URET or ERAM track data. The conflict began at 68340 seconds when the two aircraft closed to 3.27 miles and 900 feet. At 68440 seconds, minimum separation requirements were regained when vertical separation increased to 1125 feet, ending the conflict.

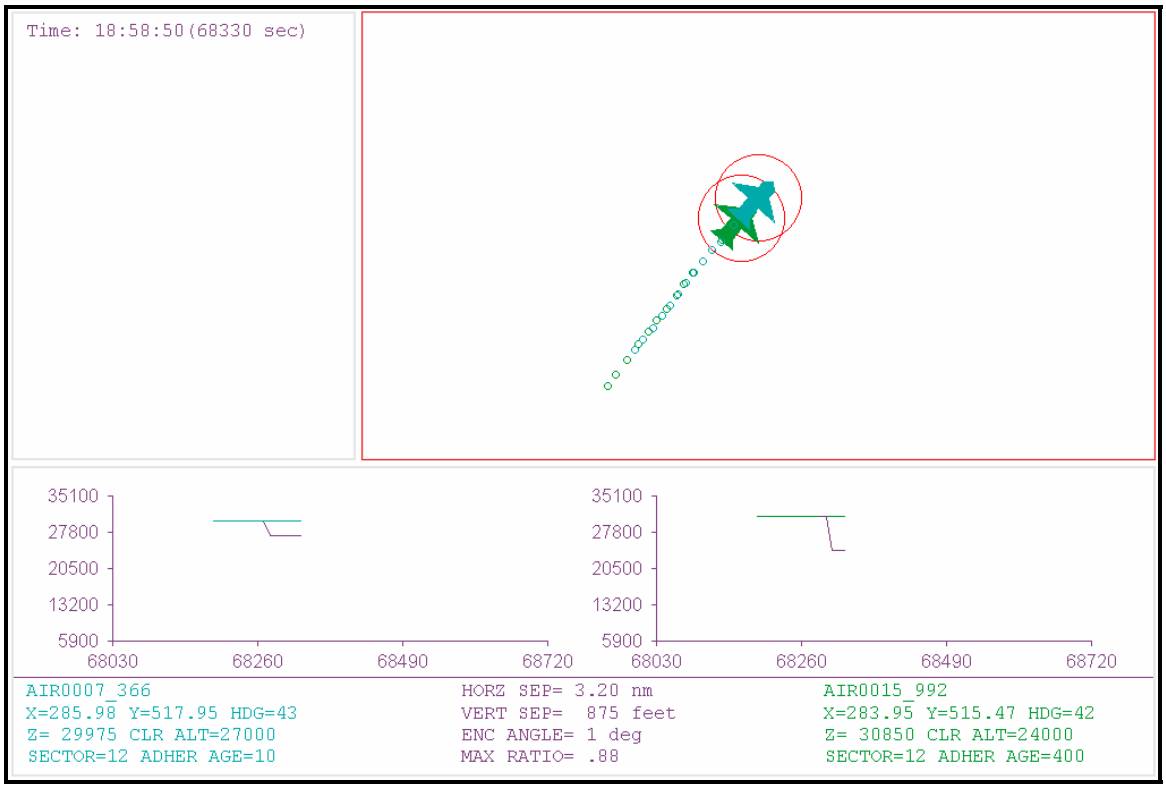

Figure 22: AIR0007 and AIR0015 at Start of Conflict

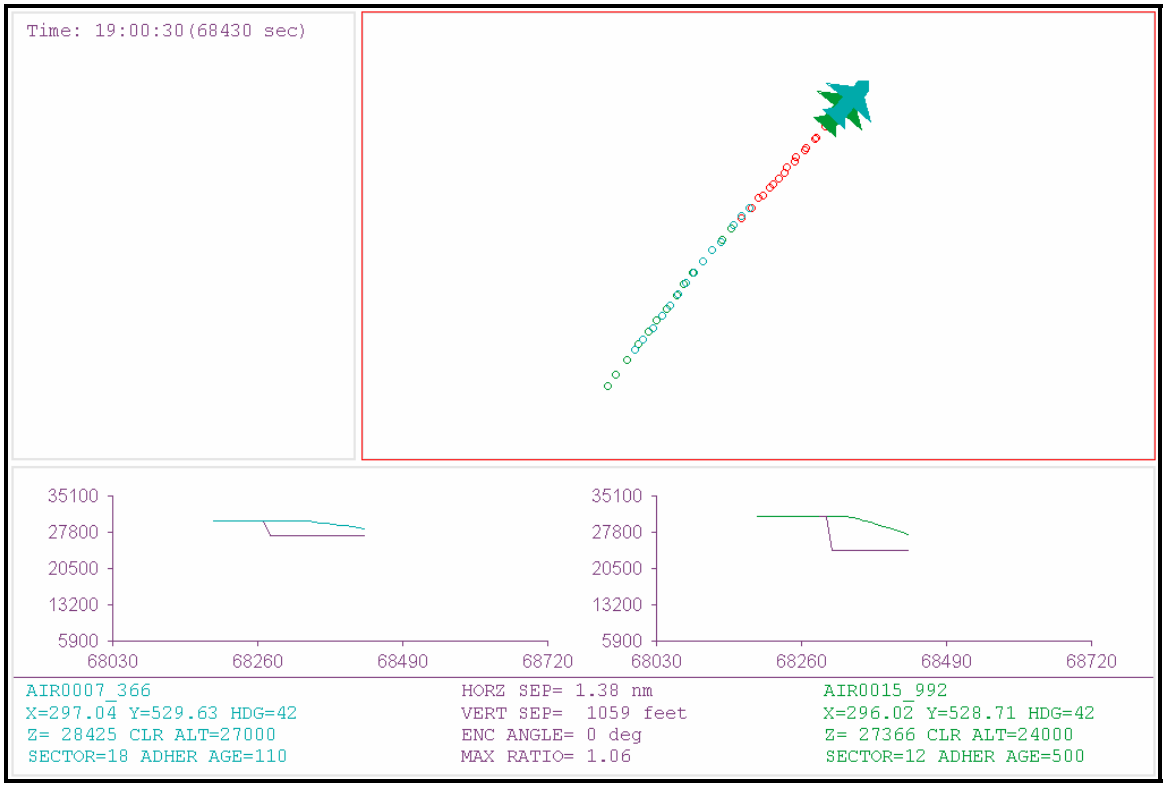

Figure 23: AIR0007 and AIR0015 at End of Conflict

ERAM generated four alerts for this aircraft pair. The first alert was issued at 68045 seconds into the scenario, predicted a conflict starting at 68044 seconds and ending at 68073 seconds, and was deleted at 68073 seconds. As the aircraft were never below minimum separation standards during this alert, a STD_FA False Alert result was generated.

The second ERAM alert was generated at 68306 seconds, was deleted at 68318 seconds, and predicted a conflict starting at 68370 seconds and ending at 68788 seconds. This alert was assigned a NO_ADHER_FA_DISCARD result as the aircraft track data was out of adherence with the aircraft flight plans.

The third ERAM alert was issued at 68329 seconds and predicted a conflict beginning at 68328 seconds and ending at 68370 seconds. This third alert was issued as a result of the interim altitude clearance to AIR0015 at 68306 seconds that resulted in a conflict being predicted based on current altitude clearances. As the actual conflict began at 68340 seconds, conflict warning time was 11 seconds, so a LATE_VA result was assigned.

Following the end of the conflict between the aircraft, updates continued for the third alert. The updates at 68519 seconds and 68542 seconds were "masked" in that the conflict being predicted between the two aircraft was based upon "final" altitude clearances as opposed to the interim altitude clearances that the aircraft had been receiving. The update at 68560 seconds was not masked as the interim altitude clearance to AIR0007 at 68560 seconds resulted in a conflict prediction based upon current cleared altitudes. This update at 68560 seconds was the fourth ERAM alert for the aircraft pair, but given that the alert was issued after the start of the conflict, a CFL_FA_DISCARD result was generated.

3.2.7.3 Missed Alert (ERD1879-C3)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.2.7.3.1 Example Description

In this simulation example, Flight AIR0011 is a Boeing MD-80 series aircraft flying from Atlanta GA to Newark NJ, with intermediate fixes at SPA, CREWE, and DYLIN2. Flight AIR0013 is also a Boeing MD-80 series aircraft flying from Atlanta GA to Bradley Airport in Hartford CT, with intermediate fixes at SPA, PXT, and DPK. Figure 24 depicts the flight paths of these two aircraft immediately before and after the conflict that occurred. Plotted aircraft positions are virtually identical using either URET or ERAM track data – URET track data is used here.

3.2.7.3.2 URET Results

Track data for these two aircraft started at 68460 seconds into the scenario. At this point AIR0011 was flying on a heading of 60 degrees at an assigned altitude of 35000 feet, and AIR0027 was flying a heading of 59 degrees at its assigned altitude of 35000 feet. The two aircraft remained at their assigned altitude of 35000 feet thru the duration of the second conflict situation between them. At the track start time, the two aircraft were already in a conflict situation, being separated by 1.71 miles and 0 feet. The two aircraft remained in an in-trail conflict situation as the aircraft began to diverge at 68670 seconds via AIR0013 changing heading to near 45 degrees until the conflict ended at 68900 seconds when the two aircraft regained minimum separation of 5.37 miles and 0 feet. The aircraft then continued on a 10

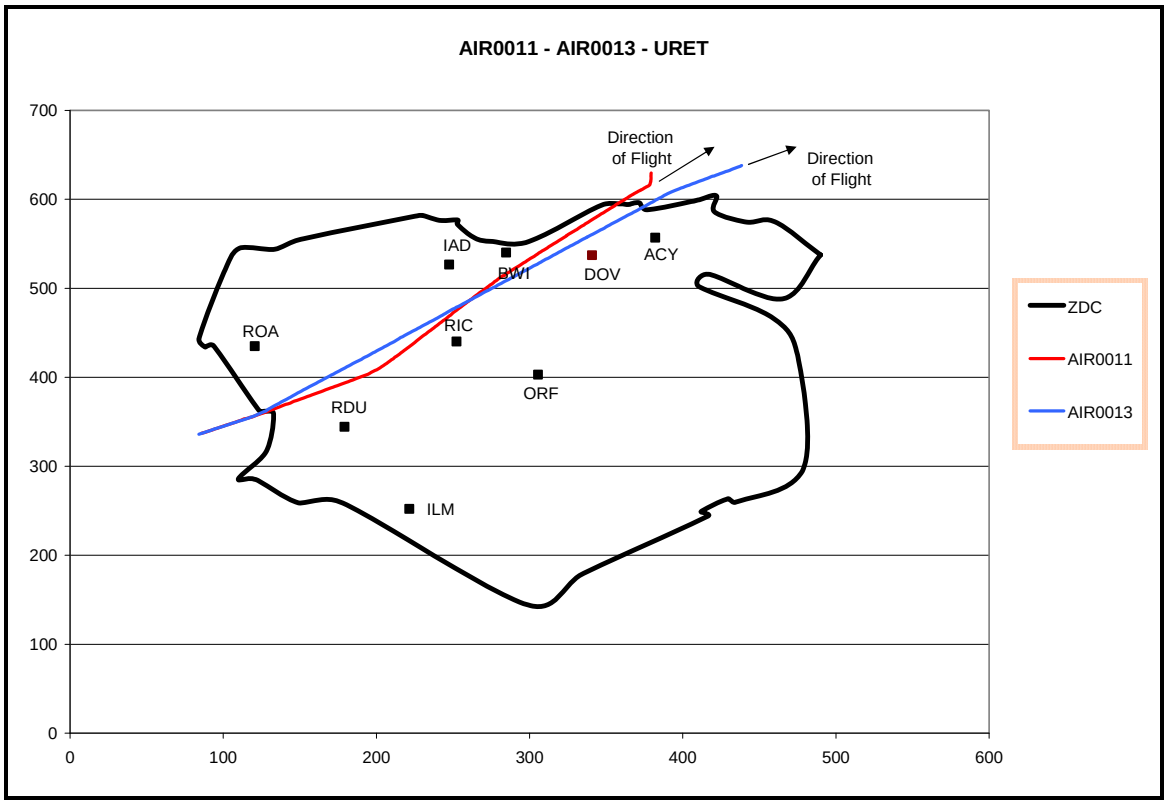

Figure 24: Flight Paths of Flights AIR0011 and AIR0013

degree diverging course until 69200 seconds, when AIR0011 began a gradual turn to a 35 degree heading. By 69370 seconds, AIR0011 reached its 35 degree heading and the aircraft were converging on a 10 degree encounter angle, being separated by 13.45 miles and 0 feet. At 69760 seconds, the aircraft entered their second conflict situation, being separated by 4.91 miles and 0 feet. Closest approach was 4.62 miles and 0 feet at 69770 seconds, and the conflict ended at 69800 seconds with the aircraft being separated by 5.38 miles and 0 feet. Following the end of the second conflict situation, the two aircraft continued on divergent path until the end of simulation track data.

The path of the two aircraft between the end of the first conflict situation and the start of the second can be seen in Figure 24. Scenario track data related to the second conflict was processed graphically using the Wolverine Software Proof graphics animation package. Figure 25 is a Proof screen capture at the time of conflict start, and Figure 26 is a screen capture at the time of conflict end. These two screen captures illustrate the "grazing" nature of the second conflict.

The URET system issued three notification sets for this aircraft pair. The first notification set was issued at 68452 seconds, prior to the start of track data, and continued being updated until 69000 seconds, after the end of the first conflict. This first conflict was a pop-up conflict in that it began prior to the availability of 5 minutes of track data, so the paired Alert and conflict were assigned a LATE_VA result.

The second notification set was issued at 68691 seconds, while the two aircraft were still involved in the first conflict situation, and was updated until 68821 seconds, prior to the end of the first conflict. As this notification set was issued prior to the end of the ongoing conflict, a CFL_FA_DISCARD result was assigned.

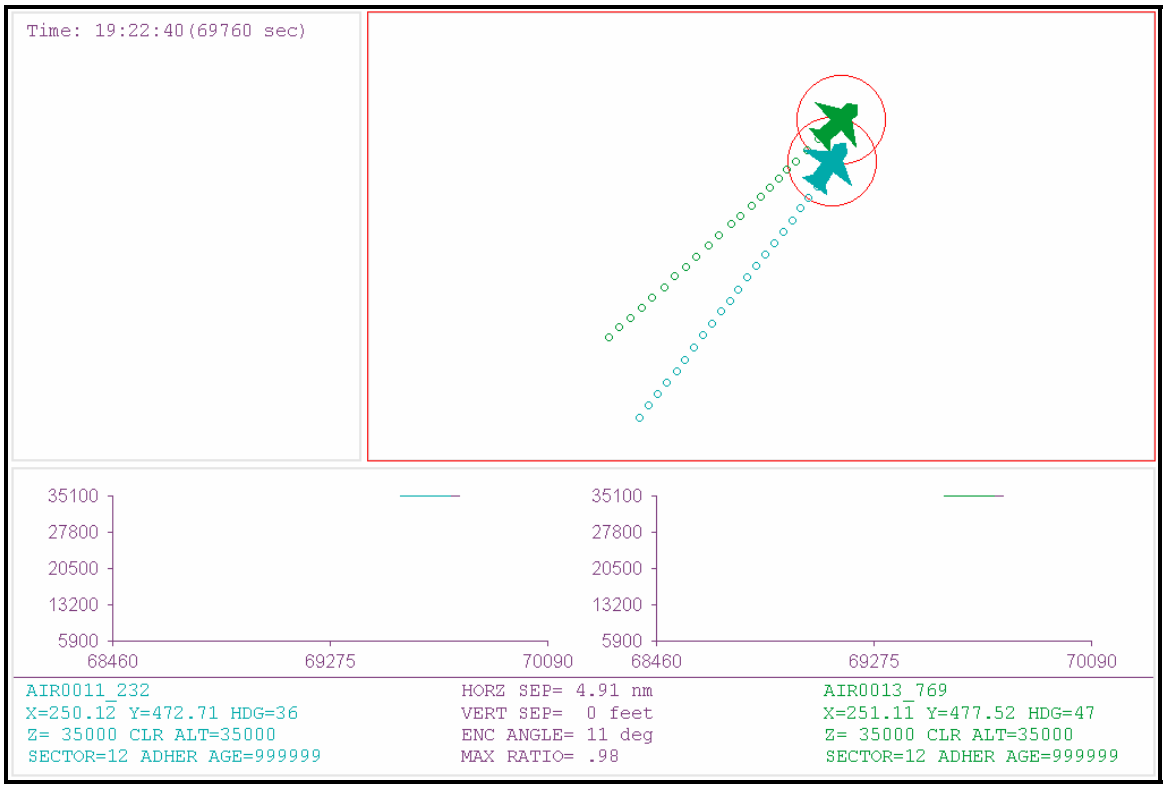

Figure 25: AIR0011 and AIR0013 at Start of Second Conflict

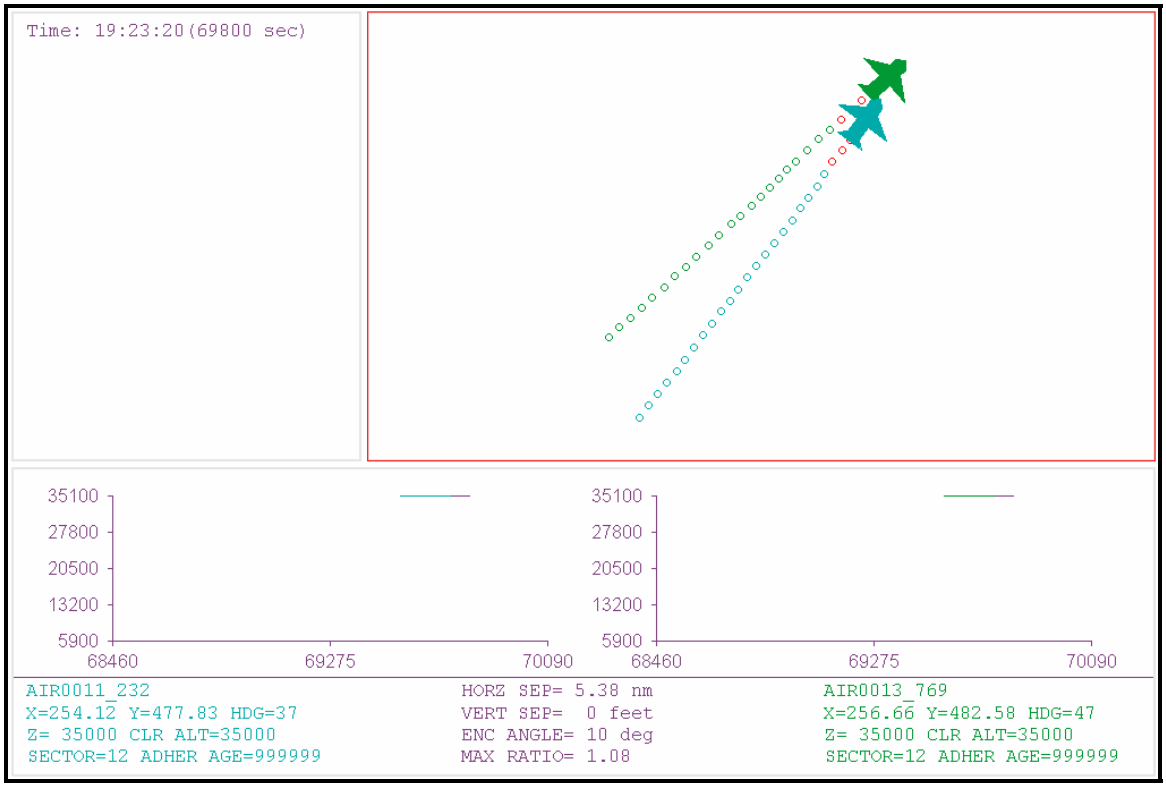

Figure 26: AIR0011 and AIR0013 at End of Second Conflict

The third notification set was issued at 69049 seconds, prior to the start of the second conflict, and was updated until 70217 seconds, after the end of this conflict. With 711 seconds of warning time, a STD_VA result was assigned to this conflict and notification set pairing.

3.2.7.3.3 ERAM Results

ERAM track data for both aircraft was first available at 68470 seconds into the scenario. The flight paths of the two aircraft up thru the start of the second conflict were virtually identical using ERAM track data as when using URET track data. The first conflict began prior to the availability of track data (pop-up), and ended at 68870 seconds with the aircraft separated by 5.3 miles and 0 feet. The second conflict began at 69750 seconds with a separation of 4.87 miles and 0 feet. The second conflict based upon ERAM track data was longer in duration than with using URET data, having a closest approach point at 69810 seconds of 4.55 miles and 0 feet, and a conflict end time at 69890 seconds with separation of 5.06 miles and 0 feet. As observed when using the URET track data, this second conflict was of a "grazing" nature, having an incursion of less than half of a mile into the minimum 5 mile separation distance during the 80 second duration of the conflict.

Prior to track data availability, ERAM issued a single alert for this aircraft pair at 68457 seconds, with updates continuing until 68705 seconds. As this alert was issued for an ongoing conflict beginning before track data availability, the pairing of this alert and conflict was assigned a LATE_VA result.

No ERAM alert was issued for the second conflict between the two aircraft, so a NO_CALL_MA result was assigned.

3.2.7.4 False Alert (ERD1879-C4)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.2.7.4.1 Example Description

In this simulation example, Flight AIR0005 is an Embraer 145 aircraft flying from Jacksonville FL to Newark NJ, with intermediate fixes at SAV, FAK, and DILON2. Flight AIR0009 is a Boeing MD-80 series aircraft flying from Atlanta GA to Norfolk VA, with intermediate fixes at GRD and RDU. Figure 27 depicts the flight paths of these two aircraft immediately before and after the encounter (non-conflict event) that occurred. Plotted aircraft positions are virtually identical using either URET or ERAM track data – URET track data is used here.

3.2.7.4.2 URET Results

Track data for both aircraft first became available at 70590 seconds into the scenario. At this time, AIR0005 was climbing thru 33000 feet on the way to an assigned altitude of 35000 feet, and was flying on a heading of 38 degrees. AIR0009 was flying level at an assigned altitude of 35000 feet on a heading of 62 degrees. By 70640 seconds, AIR0005 had reached its assigned altitude of 35000 feet, resulting in the two aircraft being on a converging path at 35000 feet, with a horizontal separation at that time of 39.9 miles. At 70993 seconds, AIR0009 received an interim altitude assignment of 25000 feet, and began its descent at 71020 seconds with the two aircraft horizontally separated by 14.3 miles. By 71070 seconds, AIR0009 descended thru 34000 feet and regained minimum vertical separation with the two aircraft being separated by 9.7 miles.

AIR0009 continued its descent as the two aircraft crossed horizontally, then diverged. No conflict occurred between these two aircraft.

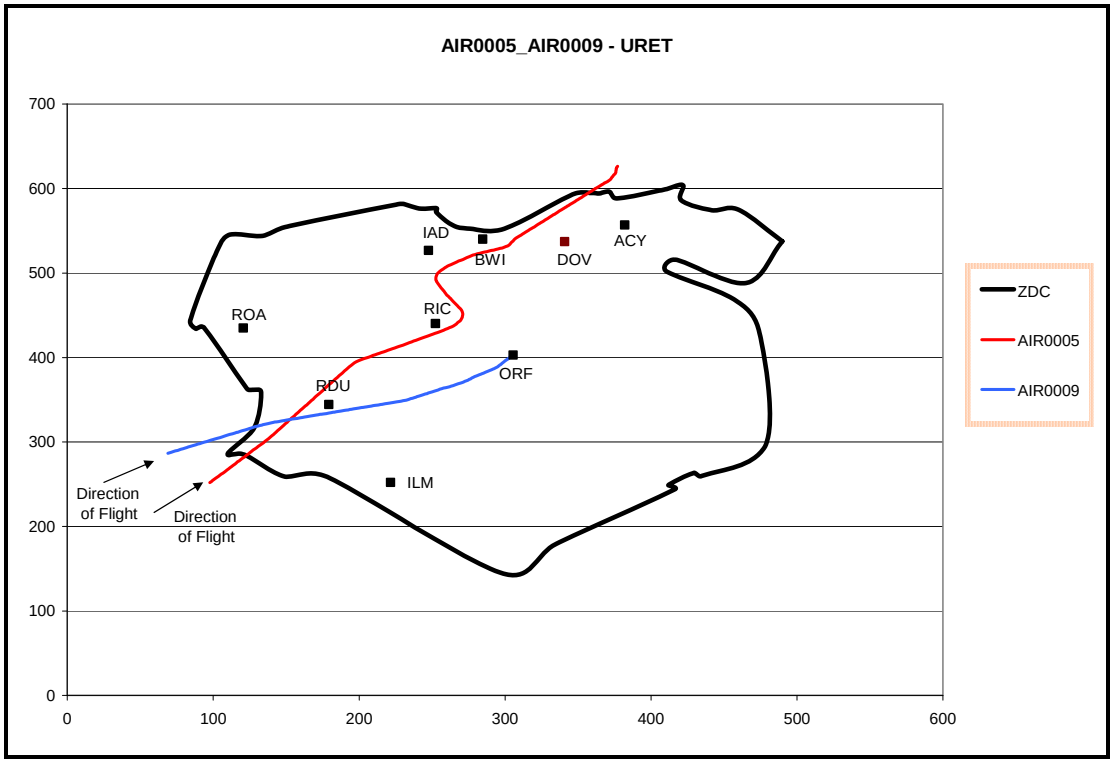

Figure 27: Flight Paths of Flights AIR0005 and AIR0009

The URET System issued an alert for these two aircraft in the form of a Notification Set (NS) having an initial ADD time of 70379 seconds and predicting a conflict start time of 71084 seconds. NS updates were posted thru 70914 seconds that predicted minimum separation distances of less than 8 miles. At 70994 seconds, one second following the AIR0009 interim altitude assignment of 25000 feet, a NS update was issued that predicted a minimum separation greater than 8 miles. Per [5], the first predicted minimum horizontal separation of greater than 8 miles defines the end of the NS.

As the interim altitude assignment to AIR0009 was within the range of five seconds before and one second following the end of the Notification Set, a CLR_FA_DISCARD_A result was assigned for the encounter between these two aircraft.

3.2.7.4.3 ERAM Results

As with the URET analysis above, track data for the two aircraft first became available at 70590 seconds, with AIR0005 climbing to an assigned altitude of 35000 feet on a heading of 38 degrees, and with AIR0009 flying level at its assigned altitude of 35000 feet on a heading of 61 degrees. By 70640 seconds, AIR0005 reached its assigned altitude of 35000 feet, and the two aircraft were converging at that altitude with a horizontal separation of 40.0 miles. At 70995 seconds, AIR0009 received an interim altitude assignment of 25000 feet, and by 71020 seconds had begun its descent with the two aircraft separated by 13.7 miles. By 71070 seconds, AIR0009 had descended thru 34000 feet regaining minimum vertical separation between the aircraft with

horizontal separation having decreased to 9.4 miles. AIR0009 continued its descent as the two aircraft crossed horizontally, then diverged. No conflict occurred between these two aircraft.

ERAM issued an Alert for these two aircraft with an initial ADD time of 70379 seconds, and with a predicted conflict start time of 71067 seconds. Updates for this Alert were issued thru 70995 seconds, predicting minimum horizontal separations less than 8 miles. At 71103 seconds, a DEL delete update was issued, ending the Alert.

The interim altitude update for AIR0009 was issued 108 seconds prior to the end of the Alert. As this was outside of the range of 5 seconds before and one second following the end of the Alert that would qualify for a CLR_FA_DISCARD_A determination, a STD_FA False Alert result was assigned for this aircraft pair.

3.3 Strategic Conflict Alerts – Aircraft to Airspace

LM ran both URET and ERAM using the CPAT supplied, time shifted scenarios. CPAT analyzed the conflict prediction performance of URET and ERAM using the output data files supplied by LM to the FAA. This section presents the CPAT analysis results. Later sections will compare the CPAT results to LM's OFPD and EOPD results.

3.3.1 Aircraft to Aircraft Conflict Alerts

An aircraft is said to be in conflict with a Special Use Airspace (SUA) whenever it is close to that SUA both horizontally, that is less than 3 nautical miles in en route airspace, and vertically, less than 500 feet. A strategic conflict prediction, as distinct from a tactical conflict, is one which will be predicted more than five minutes and less than 20 minutes into the future.

For the purposes of this URET and ERAM testing, five test SUA's were defined within the utilized Washington Center airspace with defined X-Y borders, and with an altitude range for each SUA of 18000 feet to 60000 feet. If an aircraft penetrates, entering the X-Y boundaries of the SUA and altitude range or simultaneously comes within the horizontal and vertical standard separation distances, defined above, a conflict occurs with the SUA.

Both URET and ERAM typically build a buffer around the SUA and detect conflicts when this buffer is penetrated. The buffer extends 500 feet below, 500 feet above, and three miles out from each side of the SUA, matching the separation standards. This buffer zone increases the overall volume of the SUA when looking for conflict situations. For example, if an aircraft is entering an SUA from below, the conflict would begin at 17500 feet rather than at 18000 feet if the SUA was not buffered. When evaluating URET data with the CPAT tools, no buffer is utilized.[11]

URET and ERAM are required to predict aircraft-to-airspace conflicts. When they do so correctly, the event is called a valid alert. When they do not predict a conflict which actually occurs, the event is called a missed alert. When they predict a conflict but none occurs, the event

[11] This was due to an error in the adaptation of URET's test configuration supporting the RFR. The URET test airspaces were not spatially buffered but the ERAM test airspaces were buffered – that is their boundaries were extended by three nautical miles. ERAM airspace conflicts occurred when the buffered SUA was violated and URET when the actual SUA boundary was violated. Thus, ERAM had a slightly larger set of conflicts for the RFR testing (i.e. conflicts included those that entered the buffered area as well).

is labeled a false alert. In a test scenario there are artifacts which require that a conflict or an alert be discarded and not counted in the performance calculations.

It is useful to define an encounter between an aircraft and an SUA to evaluate false alerts. An aircraft is defined to have an encounter with an SUA when it gets close together but not close enough to have a conflict. An encounter is characterized by the minimum separation of the aircraft.

3.3.2 Ground Truth from the Radar Track Data

The actual aircraft-to-airspace conflicts in the time shifted scenario were determined from the radar track data. The conflicts for the URET test run were determined from the radar track data produced by the Host. The conflicts for the ERAM test run were determined from the radar track data produced by ERAM. The obvious errors in the radar track data in the scenarios were fixed and then small gaps in data were interpolated. Then the actual conflicts in the scenario were determined by checking the distances between all of the aircraft and all of the SUA's

3.3.3 Analysis of URET Aircraft to Airspace Conflict Alert Data

The time-shifted radar track data and defined five test SUAs resulted in many aircraft-to-airspace encounters and conflicts available for the test. From the test traffic scenario of 2275 aircraft flights, CPAT processing identified the 4969 encounters and 2606 conflicts.

3.3.3.1 Matched Conflicts (Valid Alerts)

Within the URET data, there were 2275 STD_VA results and 253 LATE_VA results. Of these 2528 Valid Alerts, 2011 provided more than 10 minutes (300 seconds) of alert time, averaging 1575.4 seconds. The remaining 517 provided an average of 277.2 seconds of alert time.

3.3.3.2 Unmatched Conflicts (Missed Alerts)

Within the URET data, there were 22 NO_CALL_MA results and 15 LATE_MA results, for a total of 37 Missed Alerts. Dividing the number of Missed Alerts by the sum of Missed Alerts plus Valid Alerts resulted in a Missed Alert rate of 0.0144, or 1.44%.

3.3.3.3 Unmatched Alerts (False Alerts)

Within the URET data, there were 74 STD_FA results and 146 RETRACT FA results for a total of 220 False Alerts. The rate of False Alerts was determined independently over a number of "bins" of minimum encounter distance in NM, by dividing the number of False Alerts by the number of encounters for each bin. Table 16 below contains the False Alert rates for the different bins.

Table 16: URET False Alert Rates

Range	False Alerts	False Alert Rate
0 < d <7	201	0.0626 (6.26%)
7 < d <9	7	0.0222 (2.22%)
9 < d <11	5	0.0410 (4.10%)
11 < d <16	3	0.0073 (0.73%)
d > 16	4	0.0044 (0.44%)

3.3.3.4 Discards

Within the URET data, 751 Discard conditions were identified as tabulated in Table 17.

3.3.4 Analysis of ERAM Aircraft to Airspace Conflict Alert Data

The time-shifted radar track data and defined five test SUAs resulted in many aircraft-to-airspace encounters and conflicts available for the test. From the test traffic scenario of 2227 aircraft flights, CPAT processing identified the 4826 encounters and 2672 conflicts.

Table 17: URET Discard Rates

Discard Type	Number
NO_TRK_FA_DISCARD_A	0
NO_TRK_FA_DISCARD_B	31
NO_TRK_FA_DISCARD_C	296
NO_TRK_FA_DISCARD_D	0
NO_ADHER_FA_DISCARD	186
CLR_FA_DISCARD_A	126
CLR_FA_DISCARD_B	0
CFL_FA_DISCARD	106
NO_CALL_DISCARD	6
LATE_DISCARD	3

3.3.4.1 Matched Conflicts (Valid Alerts)

Within the ERAM data, there were 2191 STD_VA results and 333 LATE_VA results. Of these 2524 Valid Alerts, 1800 provided more than 10 minutes (300 seconds) alert time, averaging 1523.7 seconds. The remaining 724 provided an average of 294.3 seconds of alert time

3.3.4.2 Unmatched Conflicts (Missed Alerts)

Within the ERAM data, there were 85 NO_CALL_MA results and 9 LATE_MA results, for a total of 94 Missed Alerts. Dividing the number of Missed Alerts by the sum of Missed Alerts plus Valid Alerts resulted in a Missed Alert rate of 0.0642 or 3.59%.

3.3.4.3 Unmatched Alerts (False Alert)

Within the ERAM data, there were 81 STD_FA results and 425 RETRACT_FA results for a total of 506 False Alerts. The rate of False Alerts was determined independently over a number of "bins" of minimum encounter distance in NM, by dividing the number of False Alerts by the number of encounters for each bin. Table 18, listed below, contains the False Alert rates for the various bins.

Table 18: ERAM False Alert Rates

Range	False Alerts	False Alert Rate
0 < d <7	464	0.1476 (14.76%)
7 < d <9	9	0.0321 (3.21%)
9 < d <11	7	0.0522 (5.22%)
11 < d <16	11	0.0262 (2.62%)
d > 16	15	0.0177 (1.77%)

3.3.4.4 Discards

Within the ERAM data, 1543 Discard conditions were identified as tabulated in Table 19.

3.3.5 Examples

This section presents examples tailored towards each of the 2 aircraft-to-airspace conflict testing requirements in the ERAM A-Level Test Plan (ERD1879-C1/4).

Table 19: ERAM Discard Rates

Discard Type	Number
NO_TRK_FA_DISCARD_A	44
NO_TRK_FA_DISCARD_B	57
NO_TRK_FA_DISCARD_C	434
NO_TRK_FA_DISCARD_D	178
NO_ADHER_FA_DISCARD	606
CLR_FA_DISCARD_A	224
CLR_FA_DISCARD_B	0
CFL_FA_DISCARD	124
NO_CALL_DISCARD	44
LATE_DISCARD	10

3.3.5.1 Missed Alert (ERD1879-C5)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.3.5.1.1 Example Description

In this simulation example, Flight AIR0025 is a Boeing 737 series aircraft flying from Baltimore/Washington International Airport, MD to Salt Lake City, UT, with intermediate fixes at AML, FWA, OBK, and LH03. Figure 28 depicts the flight path of this aircraft with respect to the Special Use Airspace within Washington Center. Plotted aircraft positions are virtually identical using either URET or ERAM track data – URET track data is used here.

3.3.5.1.2 URET Results

Track data for AIR0025 first became available at 80160 seconds into the scenario. At this point, the aircraft had departed BWI, and was climbing at 7500 feet on a heading of 265 degrees. The aircraft was within the horizontal boundaries of SUA5, was below the vertical SUA5 floor of 18000 feet, and was within an Automatic Problem Detection Inhibited Area (APDIA) associated with the BWI Terminal Area. The aircraft continued climbing in a generally westerly direction, exiting the horizontal boundaries of SUA5 and entering the horizontal boundaries of SUA4, but still remained within the APDIA even though the aircraft climbed above the SUA floor of 18000 feet. At 80480 seconds, the aircraft exited the APDIA while climbing thru 23000 feet, initiating the conflict with SUA4. At 81050 seconds, the aircraft exited the horizontal boundaries of SUA4 at 36000 feet, ending the conflict.

The URET System issued one alert for this conflict, with a Notification Set ADD time of 80375 seconds, predicting a conflict beginning at 80485 seconds. The actual start of the conflict was at 80480 seconds, so the warning time provided was 105 seconds, less than the Minimum Warning Time (MWT) of 300 seconds. At 80156 seconds and 80424 seconds, interim altitude assignments of 23000 feet and 27000 feet respectively were given to the aircraft. As the times of these ATC clearances were less than 300 seconds prior to Actual Conflict Start Time (ACST), a VA_LATE result was assigned to this conflict.

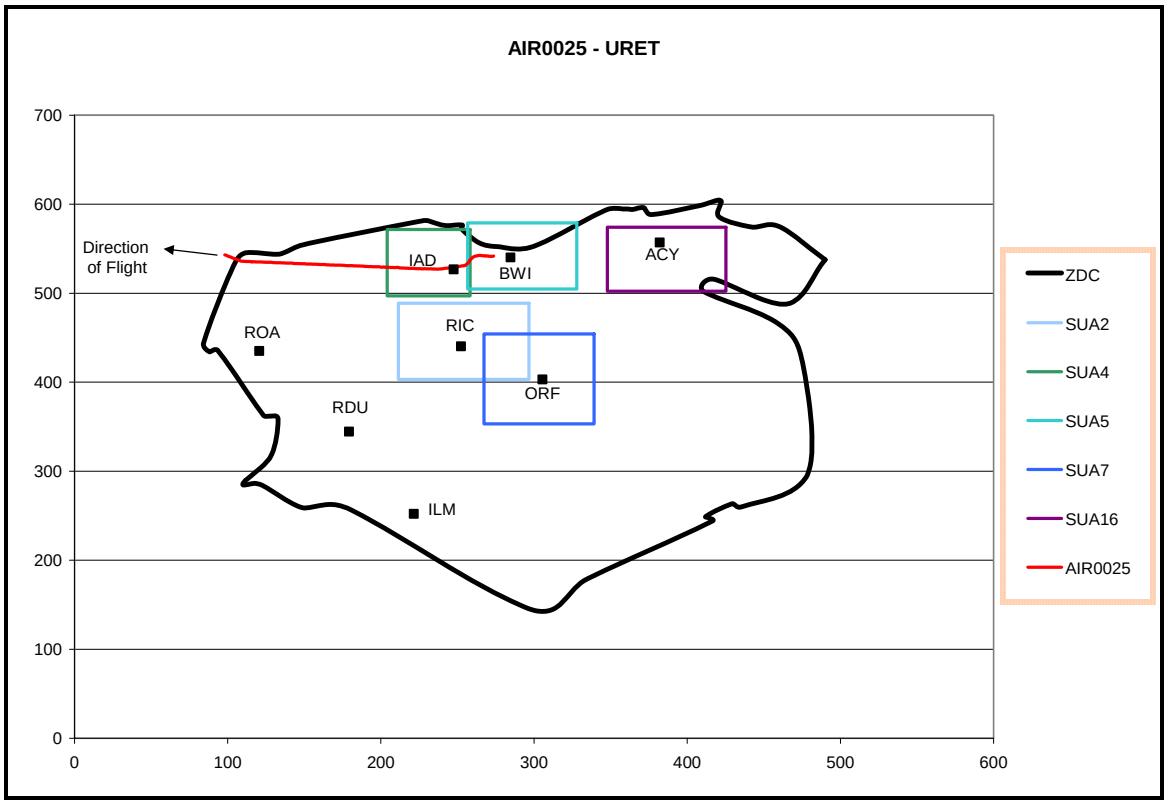

Figure 28: Flight Path of Flight AIR0025

3.3.5.1.3 ERAM Results

The aircraft departure climb path using ERAM track data was very similar to that using URET track data. ERAM track data for the aircraft became available at 80120 seconds as the aircraft was climbing thru 5600 feet on a heading of 281 degrees. The aircraft continued climbing on a generally westerly course, entering then leaving the horizontal boundaries of SUA5, then entering the horizontal boundaries of SUA4 while remaining within the APDIA associated with the BWI Terminal Area. At 80470 seconds, the aircraft exited the APDIA while climbing thru 23000 feet, initiating the conflict with SUA4. At 81070 seconds, the aircraft exited the horizontal boundaries of SUA4 while flying level at 36000 feet, ending the conflict.

As a result of the conflict, ERAM initiated one alert at 80471 seconds, predicting a conflict start time of 80471 seconds. This alert was initiated one second following the Actual Conflict Start Time. As no warning of the conflict was provided, a NO_CALL_MA Missed Alert result was assigned. Additionally, since the alert was after the ACST, a CFL_FA_DISCARD Discard result was also assigned.

3.3.5.2 False Alert (ERD1879-C6)

The following sub-section will describe the sample event and the next two sections will present results for URET and ERAM, respectively.

3.3.5.2.1 Example Description

In this simulation example, Flight GEN0021 is a Beechcraft Super King Air turboprop aircraft flying from Washington Dulles International Airport, VA to Greater Binghamton Airport, NY,

with intermediate fixes at JERES, MICAH, and CFB. Figure 29 depicts the flight path of this aircraft with respect to the Special Use Airspace within Washington Center. URET track data is used for this Figure 29. URET track data was available for a slightly longer period of time during the scenario, resulting in the aircraft flight path extending slightly beyond the northern boundary of SUA4, and past the point of aircraft track handoff.

3.3.5.2.2 URET Results

Track data for the aircraft became available at 78350 seconds into the scenario, with the aircraft climbing out from its departure from Dulles at 6000 feet and on a heading of 12 degrees. At this point, the aircraft was within both the horizontal boundaries of SUA4, and the APDIA associated with the Dulles Terminal Area. The aircraft continued climbing in a general northerly direction within SUA4 and within the APDIA, until 78910 seconds when the aircraft was handed off to the next (New York) Center, and processing of the track terminated. At track handoff, the aircraft was still below 18000 feet, and was still in the APDIA, so no conflict occurred.

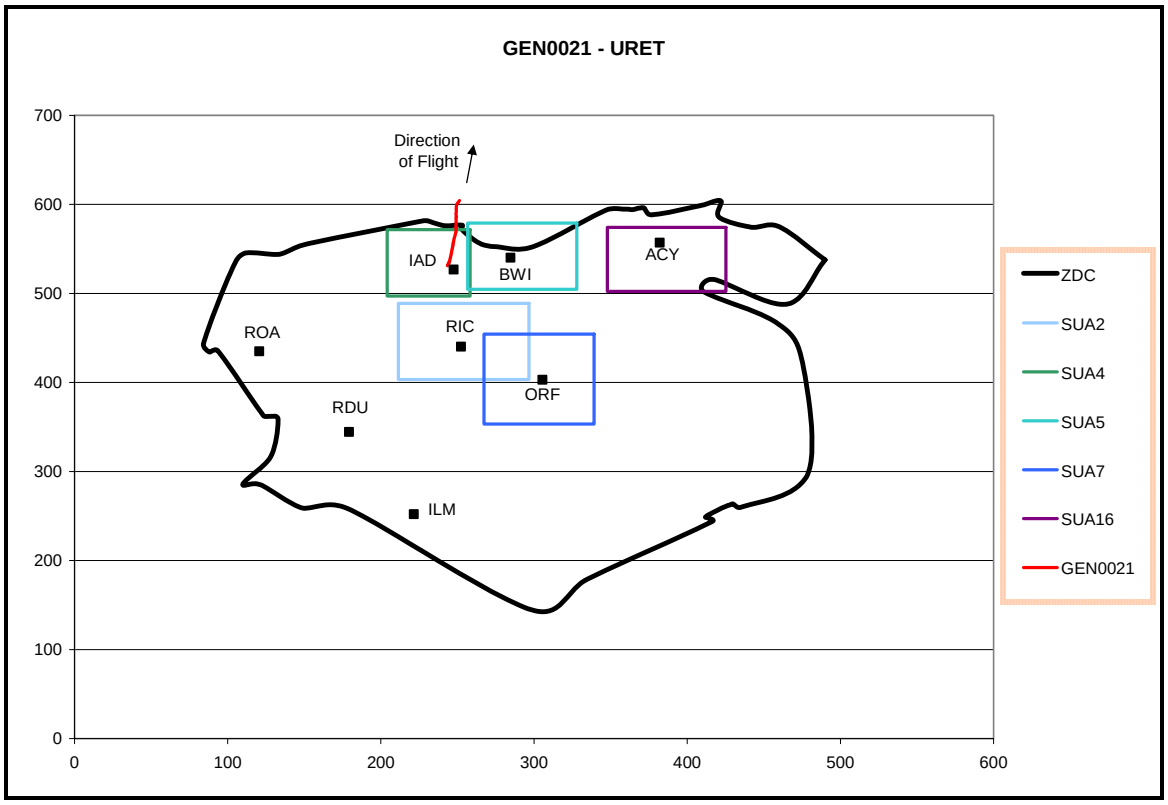

Figure 29: Flight Path of Flight GEN0021

At 78228 seconds, the Host Computer System received a DP Departure Message indicating that the aircraft had departed IAD at an assigned altitude of 21000 feet. At 78347 seconds, an Interim assigned altitude of 16000 feet was given to the aircraft.

The URET System generated one alert for the aircraft/SUA4 encounter, with a Notification Set ADD time of 78230 seconds, predicting a conflict start at 78895 seconds. This Notification Set was deleted at 79041 seconds, after the aircraft track handoff.

As no conflict occurred between the aircraft and SUA4, a STD_FA False Alert result was generated for this aircraft/SUA pair.

3.3.5.2.3 *ERAM Results*

Aircraft track data became available at 78350 seconds into the scenario, with the aircraft climbing thru 6000 feet on a heading of 9 degrees. As with the URET data analysis above, the aircraft continued climbing within the horizontal boundaries of SUA4, and within the APDIA, until 78900 seconds, when aircraft track handoff occurred and track data processing was terminated. At track handoff, the aircraft was below 18000 feet and was within the APDIA, so no conflict occurred.

At 78230 seconds, the ERAM received a DP Departure Message indicating that the aircraft had departed IAD at an assigned altitude of 21000 feet. At 78349 seconds, an Interim assigned altitude of 16000 feet was given to the aircraft.

ERAM initiated one alert at 78349 seconds, predicting a conflict start time of 78762 seconds. The alert was terminated at 78925 seconds, after aircraft track handoff.

As no conflict occurred between the aircraft and SUA4, a STD_FA False Alert result was generated for this aircraft/SUA pair.

3.4 ERAM Run 4 and 5 Accuracy Results Using CPAT Tools

Table 1 summarizes the results of the formal ERAM Run 4 and Run 5 testing for aircraft-to-aircraft and aircraft-to-airspace categories using Lockheed Martin's offline tools OFPD and EOPD. Sections 3.2 and 3.3 contain the tabulation of aircraft-to-aircraft and aircraft-to-airspace results using the various CPAT tools. Table 20 presents the pass/fail statistics for ERAM Run 4 and Run5 using CPAT tools derived data. As with Table 1, pass/fail determination for missed alert and false alert requirements was via the z-statistic analysis described in Appendix C, and that for the conflict prediction warning time requirements was via the t-statistic analysis described in Appendix D.

From Table 20, the ERD1879-C1/C2 conflict prediction warning time requirements passed, and all of the missed alert and false alert requirements failed for the formal ERAM Run 4 and Run 5 testing. Note that two of the five range "bins" for the ERD1879-C6 aircraft-to-airspace false alert requirement passed, but an overall passing grade for this requirement needed all five "bins" to pass.

Table 20: ERAM Accuracy Results Using CPAT Tools

Requirement Number	Description	Req. (URET)	ERAM	RFR Status
ERD1879-C1	CPT's aircraft-to-aircraft conflict prediction warning time, look-ahead > 10 min.	1086 sec	948 sec	PASS
ERD1879-C2	CPT's aircraft-to-aircraft immediate conflict prediction warning time, look-ahead < 10 min.	202 sec	199 sec	PASS
ERD1879-C3	CPT's aircraft-to-aircraft missed conflict alert rate	0.025	0.064	FAIL
ERD1879-C4	CPT's aircraft-to-aircraft false conflict alert rates	0.152 0.029	0.230 0.063	FAIL FAIL

		0.008 0.002	0.162 0.008	FAIL FAIL
ERD1879-C5	CPT's aircraft-to-airspace missed conflict alert rate	0.014	0.036	FAIL
ERD1879-C6	CPT's aircraft-to-airspace false conflict alert rates	0.063 0.022 0.041 0.007 0.004	0.148 0.032 0.052 0.026 0.018	FAIL PASS PASS FAIL FAIL

4 Comparison of FAA and Contract Developer Analyses

This section describes the results of comparing the analyses performed by the contract developer, Lockheed Martin (LM), and the FAA. LM's OFPD (Offline Problem Determination) and EOPD (ERAM Offline Problem Determination) are support tools used to measure the conflict prediction performance of URET and ERAM, respectively. Since OFPD and EOPD tools are used in testing the CPT requirements of ERAM, CPAT considers it useful to validate these results using their own "in-house" tools, referred to as CPAT tools. CPAT tools measures the conflict prediction performance for both URET and ERAM. This comparison is helpful not only for verification, but offers LM with third party performance measurements that allows them to sanity check their own tools.

The analyses involved comparing aircraft-to-aircraft (AC-AC) and aircraft-to-airspace (AC-AS) conflict predictions separately between the different tools. Also, only five test special activity airspaces (SAAs) were used in the analyses to avoid airspace adaptation issues. CPAT defined these test SAAs as simple rectangular volumes and provided them to LM to use during the RFR.

4.1 Comparison Overview

CPAT tools and EOPD (OFPD) both start with the same dataset from ERAM (URET). LM provided the alerts from both URET and ERAM System Analysis Records (SARs) and used scripts to extract and format the alerts in manner which CPAT tools can read. Each application processes the same alerts. Comparing the conflict prediction performance between two separate applications is challenging because the two tools defined their own sets of actually occurring conflicts, and falsely predicted conflicts (FPCs). The two sets are different due to the fact that the input data into the conflict prediction performance evaluation programs differ because logic differences and errors in the software.

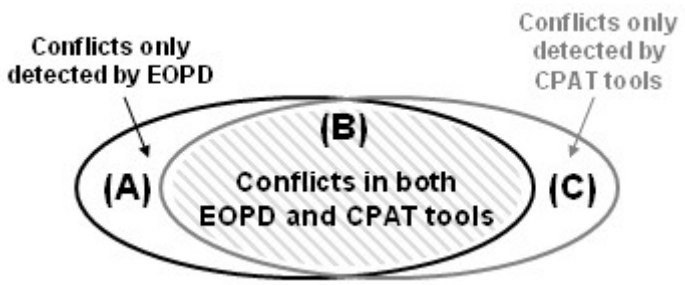

Figure 30: Conflicts detected by EOPD and CPAT tools.

When comparing the conflict prediction results for actually occurring conflicts from both sets of tools most of the conflicts detected by CPAT tools and EOPD should be the same. These mutually detected conflicts appear in Area B of Figure 30. As a result of differences in data processing and problem determination logic some conflicts are only detected by one tool. Area A of Figure 30 contains these conflicts only detected by EOPD, and Area C contains these conflicts only detected by CPAT tools. The conflict prediction results of the summation of Areas A, B, and C are the input into the compare process shown in Figure 31. The result of the comparison quantifies the number of instances where the tools agree or disagreement with whether an actually occurring conflict was classified as VA, MA, or a DISCARD.

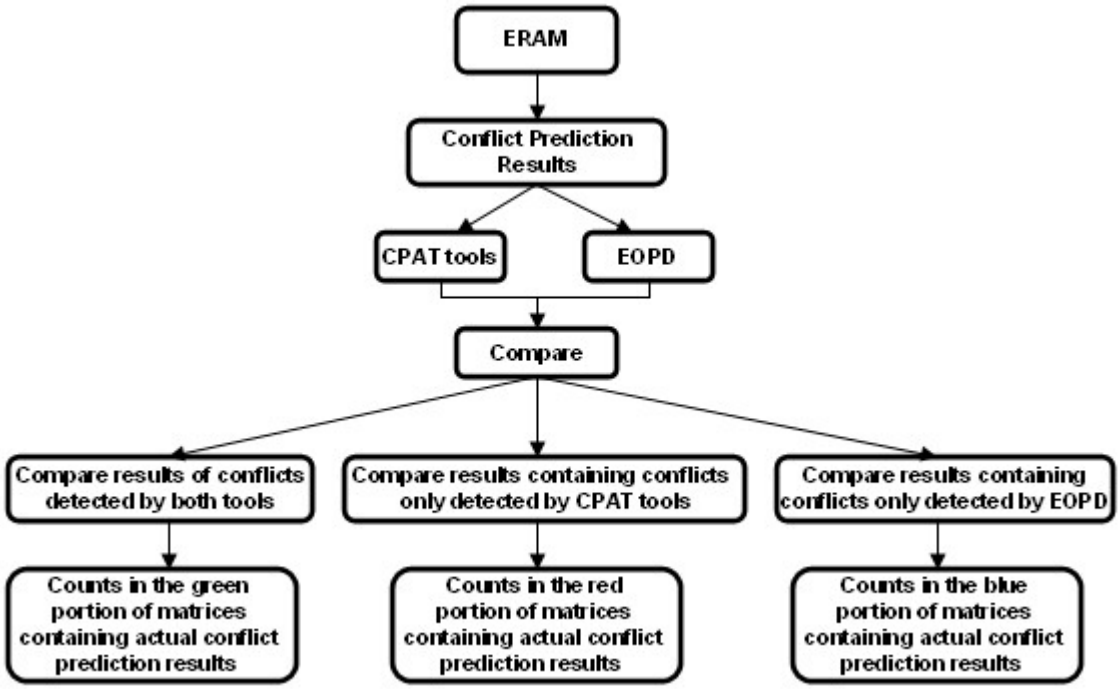

Figure 31: Conflict Prediction Comparison Process. See following tables for matrix examples.

The process of comparing the conflict prediction results of FPCs between CPAT tools and EOPD is analogous to the Venn diagram in Figure 30, illustrating actually occurring conflicts in CPAT tools and EOPD. Between the two tools there is a set of FPCs agreed upon by the two tools (similar to Area B of Figure 30). Also, there is a set of FPCs only considered by EOPD (similar to Area A of Figure 30), and a set only considered by CPAT tools (similar to Area C in Figure 30). The process of comparing the conflict prediction results for FPCs is also very similar to the comparison of conflict prediction results of actually occurring conflicts, except the agreement and disagreements are analyzed based on if the systems discarded the FPC or not.

The entire process of comparing the conflict prediction results of CPAT tools and LM's tools (EOPD and OFPD) is a complex problem. Since LM developed two separate applications to determine the performance of URET and ERAM's CPT, the comparison is split by two into comparing CPAT tools' results for URET and ERAM, separately against OFPD and EOPD. In addition, the comparison is divided by comparing conflict predictions for actually occurring conflicts and FPCs. Lastly, the comparison is split among the conflicts that are between two aircrafts and conflicts between an aircraft and an airspace. The illustration in Figure 31 details the comparison process of comparing the coflict prediction results of actually occurring AC-AC conflicts CPAT tools and EOPD, which is only one phase of the eight total phases of the comparison. The eight phases are illustrated and summarized in Appendix E.

4.2 Comparison Procedures

The comparison of conflict predictions is divided into two groups. The first group, described in described in Section 4.2.1, consists of comparing the conflict prediction results for actual occurring conflicts. The second group, described in Section 4.2.2, consists of comparing notification sets which falsely predicted conflicts.

4.2.1 Comparing Actual Conflicts' Prediction Results

This section describes the comparison of actually occurring conflicts' prediction results between CPAT and LM. The procedure in comparing the conflict prediction performance of conflicts measured by CPAT and LM tools is a two step process. The first step is to process all the conflicts detected by CPAT tools, and identify all conflicts that match LM detected conflicts and do not match LM detected conflicts. Secondly, process all LM conflicts deriving from the first step that did not match a CPAT detected conflict and register as conflicts not detected by CPAT tools. These processes are performed twice for both AC-AC and AC-AS conflicts and described in Table 21. In order to avoid redundancy while explaining the two processes, below, in detail, only EOPD will be used in the explanation of the process, but could be used interchangeably with OFPD.

Table 21: The four comparison combinations for analyzing conflict prediction performances.

System	Conflict Type	Tools Comparison
URET	AC-AC	CPAT tools vs. OFPD
URET	AC-AS	CPAT tools vs. OFPD
ERAM	AC-AC	CPAT tools vs. EOPD
ERAM	AC-AS	CPAT tools vs. EOPD

The first step begins by loading all the conflict and their predictions results for both CPAT tools and EOPD. This data includes actual conflict start time (ACST) and end time (ACET), ACIDs of the aircrafts in conflict, notification time, predicted conflict start time (PCST) and end time (PCET), minimum horizontal separation, minimum vertical separation, and conflict prediction result. Once the data is loaded, every conflict detected by CPAT tools is individually processed for matches within the EOPD conflict set. A matching conflict is when both conflicts detected by CPAT tools and EOPD are between the same aircrafts. Each CPAT tools conflict can match to zero or more EOPD conflicts. In the case of zero matches, the CPAT tools conflict is considered to be undetected by EOPD. If the CPAT conflict matches one EOPD conflict then the information about this match is considered to be detected by both tools. If the CPAT tools conflict matches more than one EOPD conflict, then the EOPD conflict closes in ACST is considered to be the match. As EOPD conflicts become matched with CPAT tools conflicts they are flagged as already been matched. The first step is complete once all CPAT tools conflicts have been processed.

The second step is a simple traversal process. Since some EOPD conflicts may not have matched to a CPAT tools conflict counterpart, all EOPD conflicts not flagged as matching in the previous step are considered to be undetected by CPAT tools.

When classifying the conflict prediction results of EOPD conflicts in terms of how CPAT tools classifies (See Appendix A) the following assumptions are made:

- Conflicts in the *Missed Conflict* section of the EOPD report are classified as NO_CALL_MAs.
- Conflicts in the *Strategic Missed Conflict* section of the EOPD report are classified as LATE_MAs.
- Conflicts in the *Exception to Missed Conflict Notification* section of the EOPD report are classified as NO_CALL_DISCARDs.
- Conflicts in the *Notifications Paired with Invalid Conflict* section of the EOPD report are classified as LATE_DISCARDs.

- Conflicts in the *Average Warning Time Report* section of the EOPD report with warning times greater than or equal to five minutes are classified as STD_VAs.
- Conflicts in the *Average Warning Time Report* section of the EOPD report with warning times less than five minutes are classified as LATE_VAs, hence are assumed to be popup conflicts.

The counts of the different outcomes are manually recorded in a matrix. The matrix provides an overview of how many instances exist for each conflict prediction result outcome. An example of the matrix with description for each cell appears in Table 22. The green portion of the matrix is linked to the conflicts detected by both sets of tools, the red portion contains conflicts detected by CPAT tools, but not EOPD, the blue portion derives from the conflicts detected by EOPD, but not CPAT tools, and the orange portion would contain all aircrafts that were near conflicts (encounters), but are not evaluated in our analyses.

Table 22: An example of the matrix containing actual conflict prediction result counts from CPAT and LM.

LM **CPAT**	**Valid Conflicts**	**Missed Conflicts**	**Discarded Conflicts**	**No Conflict**	**Totals (CPAT)**
Valid Conflicts	Conflicts detected and classified as valid by both set of tools	Conflicts detected by both tools, but CPAT classified as valid and LM classified as missed	Conflicts detected by both tools, but CPAT classified as valid and LM classifies as discarded	Conflicts only detected by CPAT tools and classified as valid	CPAT's Total Number of Valid Conflicts
Missed Conflicts	Conflicts detected by both tools, but CPAT classified as missed and LM classified as valid	Conflicts detected and classified as missed by both set of tools	Conflicts detected by both tools, but CPAT classified missed and LM classified as discarded	Conflicts only detected by CPAT tools and classified as missed	CPAT's Total Number of Missed Conflicts
Discarded Conflicts	Conflicts detected by both tools, but CPAT classified as discarded and LM classified as valid	Conflicts detected by both tools, but CPAT classified as discarded and LM classified as missed	Conflicts detected and classified as discarded by both set of tools	Conflicts only detected by CPAT tools and classified as discarded	CPAT's Total Number of Discarded Conflicts
No Conflict	Conflicts only detected by LM and classified as valid	Conflicts only detected by LM and classified as missed	Conflicts only detected by LM and classified as discarded	Total Encounters (not analyzed)	
Totals (LM)	LM's Total Number of Valid Conflicts	LM's Total Number of Missed Conflicts	LM's Total Number of Discarded Conflicts		

4.2.2 Comparing Falsely Predicted Conflicts

This section describes the comparison of false conflicts' prediction results between CPAT and LM. Similar to comparing actually occurring conflicts in Section 4.2.1, the procedure in comparing the conflict prediction performance of falsely predicted conflicts (FPCs) measured by CPAT tools and LM tools is a two step process. The first step is to process all FPCs from CPAT tools, and identify all FPCs that match FPCs from LM and do not match FPCs from LM. Secondly, process all LM FPCs deriving from the first step that did not match a FPC from CPAT tools and register as FPCs not registered by CPAT tools. These processes are performed twice for both AC-AC and AC-AS FPCs and described in Table 21. In order to avoid redundancy while explaining the two processes, below, in detail, only EOPD will be used in the explanation of the process, but could be used interchangeably with OFPD.

The first step begins by loading all the FPCs and their conflict prediction results for both CPAT tools and EOPD. This data includes ACIDs of the aircrafts, notification time, predicted conflict

start time (PCST) and end time (PCET), minimum horizontal separation, minimum vertical separation, and conflict prediction result. Once the data is loaded, every FPC from CPAT tools is individually processed for matches within the EOPD FPC set. A matching FPC is when both tools determine a predicted conflict between the same aircrafts did not occur. Each FPC from CPAT tools can match to zero or more FPCs from EOPD. In the case of zero matches, the CPAT tools FPC is considered to not have been falsely predicted by EOPD (no call). If the CPAT tools FPC matches one EOPD FPC then the FPC is considered by both tools. If the CPAT tools FPC matches more than one EOPD FPC, then the EOPD FPC closes in notification set start time is considered to be the match, and considered to be a mutual FPC. As EOPD FPCs become matched with CPAT tools FPCs they are flagged as already been matched. The first step is complete once all CPAT tools FPCs have been processed.

The second step is a simple traversal process. Since some FPCs from EOPD may not have matched to a FPC from CPAT tools, all EOPD FPCs not flagged as matching in the previous step are determined to not be classified as FPCs by CPAT tools (no call).

When classifying the prediction results of FPCs from EOPD in terms of how CPAT tools classifies (See Appendix A) the following assumptions are made:

- Conflicts in the *False Conflict Notification* section of the EOPD report are classified as FAs, but whether they are RETRACT_FAs or STD_FAs can not be determined because OFPD does not provided predicted conflict start time (PCST).
- Conflicts in the *Exception to False Conflict Notification* section of the EOPD report are classified as DISCARDs with the following EOPD to CPAT tools mapping:
    - RETRACTED_BY_CLEARANCE_CODE → CLR_FA_DISCARD_A
    - NOT_ADHERENCE_CODE → NO_ADHER_FA_DISCARD
    - AFTER_LAST_CONFLICT_IN_ENCOUNTER → CFL_FA_DISCARD
    - NO_TRACK_5_SEC could not be mapped because CPAT tools consider four different no track exception cases (see Appendix A).

The matrix provides an overview of how many instances exist for each FPC prediction result outcome. An example of the matrix with description for each cell appears in Table 23. The green portion of the matrix is linked to the FPCs determined by both sets of tools, the red portion derives the FPCs determined by CPAT tools, but not EOPD, the blue portion derives from the FPCs determined by EOPD, but not CPAT tools, and the orange portion would contain all the correct no calls.

4.3 CPAT versus OFPD Analysis

The CPT performance measurements of URET are used as the standard in determining whether ERAM passes the RFR, hence validating the results of URET is of utmost importance. OFPD produces prediction results for both aircraft-to-aircraft and aircraft-to-airspace conflicts. In this section we detail the comparison between CPAT tools and OFPD.

4.3.1 URET Aircraft-to-Aircraft Conflict Prediction Comparison

The comparison of URET's AC-AC conflict prediction results is divided into two parts. The first part describes the comparison of the conflict prediction results for actual AC-AC conflicts. The second part details the comparison of conflict prediction results of false AC-AC conflicts.

Table 23: An example of the matrix containing false conflict prediction result counts from CPAT and LM.

LM CPAT	False Alert	Discarded Alert	No FPC	Totals (CPAT)
False Alert	FPCs classified as false alerts by both set of tools	FPCs classified as false alerts by CPAT tools and discarded alerts by LM	FPCs only classified as false alerts by CPAT tools	CPAT's Total Number of False Alerts
Discarded Alert	FPCs classified as false alerts by LM and discarded alerts by CPAT tools	FPCs classified as discarded alerts by both set of tools	FPCs only classified as discarded alerts by CPAT tools	CPAT's Total Number of Discarded False Alerts
No FPC	FPCs only classified as false alerts by LM	FPCs only classified as discarded alerts by LM	Total No Calls	
Totals (LM)	LM's Total Number of False Alerts	LM's Total Number of Discarded False Alerts		

4.3.1.1 URET Prediction Comparison of Actual AC-AC Conflicts

CPAT tools detected a total of 249 AC-AC conflicts and OFPD detected 243 AC-AC conflicts. Of these two sets of conflicts, 238 were detected by both tools having the same aircraft and if not same start time, then very close, therefore considered to the same conflicts. These matched conflicts appear in the green portion of Table 24. Of the 238 conflicts detected by both tools, 232 conflicts had the same conflict prediction result. Of all the conflict matches only one was classified as a VA by CPAT tools and MA by OFPD. Similarly, only one conflict was classified as a MA by CPAT tools and VA by OFPD, therefore the additional MA in each tool nullifies any effect when considering the CPT's AC-AC missed conflict alert rate requirement (ERD1879-C3).

Since CPAT tools detected 238 of the 243 conflicts detected by OFPD, five unmatched conflicts remain. These five conflicts, appearing in the blue portion of Table 24, had three VA, one MA, and one DISCARD conflict prediction result. One reason why five conflicts were not detected by CPAT tools was some conflicts were just shorter than 20 seconds long. Short duration conflicts are sometimes overlooked, especially in 10 second interpolated track data, due to the fact that it is fine line in determining separation exactly agreeing distances stemming from rounding errors and ARTCC handoff considerations (PR number TBD).

CPAT tools detected 11 conflicts, appearing in the red portion of Table 24, in which OFPD failed to detect. We found that these conflicts were only detected by CPAT for similar reasons (rounding errors, etc.) as five conflicts were only detected by OFPD, but in the reverse effect. Overall, there were minor differences between the results of CPAT tools and OFPD for AC-AC conflicts, which was expect since the legacy URET system has been a stable operational system for a while. Therefore, we focused more analyses and detail on the comparison of ERAM's conflict prediction results (see Section 4.4.1.1). However, these differences should be addressed in future usage of OFPD during new releases of ERAM.

4.3.1.2 URET Prediction Comparison of False AC-AC Conflicts

CPAT tools detected a total of 2360 FPCs and OFPD detected a total of 2369 FPCs. Of these totals, both tools detected 2355 of the same FPCs, which appear in the green portion of Table 25. The notification times of these 2355 FPCs generally agreed among the tools, however a few notification times differed by one or two seconds, which resulted in very little impact. Overall the number of FAs and DISCARDs for CPAT tools and OFPD are very close. There were 102 FPCs classified as FAs by CPAT tools, where were discarded by OFPD. Of these 102 FPCs, 41

were discarded by OFPD based on a NO_TRACK_5_SEC exception code. Many of these FPCs had track at the PCST, but were within an APDIA, which is still considered to be valid track therefore not discarded by CPAT tools. OFPD considers no track exist for a flight if it is within an APDIA. This is an error in the logic for discarding FAs, and has been registered as (PR number TBD) by LM. Additionally, 59 FPCs of the 102 FPCs kept as FAs by CPAT tools and discarded by OFPD were discarded based on adherence. In the case where OFPD considered a FPC to be a FA and CPAT tools a DISCARD, there were 100 instances, and within these 100 instances there were almost the same number of FPCs discarded by CPAT tools based on clearances, adherence, and no track data, individually.

Table 24: Matrix containing URET's actual AC-AC conflict prediction result counts as detected by CPAT and LM.

LM CPAT	Valid Alert	Missed Alert	Discarded Alert	No Conflict	Totals (CPAT)
Valid Alert	226	1	0	6	233
Missed Alert	1	4	0	1	6
Discarded Alert	4	0	2	4	10
No Conflict	3	1	1	0	5
Totals (LM)	234	6	3	11	254

There was a total of 14 FPCs detected by OFPD that were not considered FPCs by CPAT tools, which appear in the blue portion of Table 25. Six of the 14 FPCs were not considered because CPAT tools detected actual conflicts to match to the notification sets, and classified the prediction as VAs (these six VAs appear in the red portion of Table 24). The rest of the FPCs were considered to be no calls by CPAT tools, which means CPAT tools did not detect eight of the same notification sets. Furthermore, CPAT tools detected five FPCs in which OFPD did not consider. Three of the five were because OFPD detected actual conflicts in which CPAT tools did not. The remaining two were simply no calls detected by OFPD.

Table 25: Matrix containing URET's false AC-AC conflict prediction result counts from CPAT and LM.

LM CPAT	False Alert	Discarded Alert	No FPC	Totals (CPAT)
False Alert	548	102	2	652
Discarded Alert	100	1605	3	1708
No FPC	4	10	0	14
Totals (LM)	652	1717	5	2374

4.3.2 URET Aircraft-to-Airspace Conflict Prediction Comparison

The comparison of URET's AC-AS conflict prediction results is divided into two parts. The first part describes the comparison of the conflict prediction results for actual AC-AS conflicts. The second part details the comparison of conflict prediction results of false AC-AS conflicts.

4.3.2.1 URET Prediction Comparison of Actual AC-AS Conflicts

CPAT tools detected a total of 2606 AC-AS conflicts and OFPD detected 2600 AC-AS conflicts among the test SAAs used in the RFR. Of these two sets of conflicts, 2588 were detected by both tools having the same aircraft and if not same start time, then very close, therefore considered to be the same conflicts. These matched conflicts appear in the green portion of Table 26. Of the 2588 conflicts detected by both tools, 2577 conflicts had the same conflict prediction result. CPAT tools detected 18 conflicts in which OFPD did not detect, and OFPD detected 12 conflicts

not detected by CPAT tools. Interestingly, all 12 conflicts only detected by OFPD were classified as VAs. Of the 18 conflicts detected only by CPAT tools, two were MAs. In total CPAT tools classified 37 conflicts as MAs, and OFPD classified 33 conflicts as MAs. A total of 30 conflicts were missed by both tools. A total of three conflicts were considered to be MAs by CPAT tools and VAs by OFPD, and three were considered to be DISCARDs by CPAT tools and MAs by OFPD. Furthermore, two were considered to be MAs by CPAT tools and DISCARDs by OFPD, and two conflicts only detected by CPAT tools were MAs. All the counts of different events between OFPD and CPAT tools had only minor differences.

Table 26: Matrix containing URET's actual AC-AS conflict prediction result counts as detected by CPAT and LM.

LM CPAT	Valid Alert	Missed Alert	Discarded Alert	No Conflict	Totals (CPAT)
Valid Alert	2524	0	0	4	2528
Missed Alert	3	30	2	2	37
Discarded Alert	3	3	23	12	41
No Conflict	12	0	0	0	12
Totals (LM)	2542	33	25	18	2618

4.3.2.2 URET Prediction Comparison of False AC-AS Conflicts

CPAT tools detected a total of 965 FPCs and OFPD detected a total of 956 FPCs. Of these totals, both tools detected 949 of the same FPCs, which appear in the green portion of Table 27. The notification times of these 949 FPCs generally agreed among the tools, however one differed by 61 seconds and approximately 25% of the notification times differed by one or two seconds. There were 151 FPCs which both tools classified as FA and 696 FPCs that were discarded by both tools. Furthermore, there were 62 FPCs in which CPAT tools classified as FAs when OFPD discarded the FPCs. The majority of these 62 FPCs were discarded by OFPD because of the no track exception. CPAT tools disagreed with these DISCARDs because it detected that track existed at the PCST, although the track was within an APDIA (see OFPD PR number TBD). Also, there were 40 FPCs in which OFPD classified as FAs when CPAT tools discarded the FPCs. The majority of these FPCs were discarded by CPAT tools because of adherence.

There was a total of seven FPCs detected by OFPD that were not considered FPCs by CPAT tools, which appear in the blue portion of Table 27. Four of the seven FPCs were not considered because CPAT tools detected actual conflicts to match to the notification sets. The rest of the FPCs were considered to be no calls by CPAT tools, which means CPAT tools did not detect three of the same notification sets. Furthermore, CPAT tools detected 16 FPCs in which OFPD did not consider. Twelve of the 16 were because OFPD detected actual conflicts in which CPAT tools did not. All 12 of these actual conflicts detected by OFPD were classified as VA, and appear in the blue portion of Table 26. The remaining four were simply no calls detected by OFPD.

Table 27: Matrix containing URET's false AC-AS conflict prediction result counts from CPAT and LM.

LM CPAT	False Alert	Discarded Alert	No FPC	Totals (CPAT)
False Alert	151	62	7	220
Discarded Alert	40	696	9	745
No FPC	2	5	0	7
Totals (LM)	193	763	16	972

4.4 CPAT versus EOPD Analysis

The CPT performance measurements of ERAM, which detail how the system compares to the legacy system, are used to determine if the system passes requirements for RFR Run 5, hence validating the results of ERAM is critical. EOPD produces prediction results for both aircraft-to-aircraft and aircraft-to-airspace conflicts. In this section we detail the comparison between CPAT tools and EOPD.

4.4.1 ERAM Aircraft-to-Aircraft Conflict Prediction Comparison

The comparison of ERAM's AC-AC conflict prediction results is divided into two parts. The first part describes the comparison of the conflict prediction results for actual AC-AC conflicts. The second part details the comparison of conflict prediction results of false AC-AC conflicts.

4.4.1.1 ERAM Prediction Comparison of Actual AC-AC Conflicts

CPAT tools detected a total of 230 AC-AC conflicts and EOPD detected 226 AC-AC conflicts. Of these two sets of conflicts, 220 were detected by both tools having the same aircraft and if not same start time, then very close, therefore considered to be the same conflicts. These matched conflicts appear in the green portion of Table 28. Of the 220 conflicts detected by both tools, 207 conflicts had the same alert classification; hence 13 disagreements existed among alert classifications within the conflicts detected by both tools.

The most significant of the 13 differences were the seven instances where a conflict was classified as a DISCARD by EOPD and MA by CPAT. These inconsistencies were reported to LM, and after investigating it was discovered that the problem resided in EOPD. One of the seven conflicts is referred to as Incident[12] 1. For this conflict, both tools detect start of conflict at 67610 seconds (18:46:50). CPAT tools determined that both aircrafts are never out of adherence, but EOPD erroneously determined that N1926S went out of adherence, therefore discarding the conflict and classifying as a DISCARD. EOPD miscalled the conflict due to PR_28728; setting first track to ADH_AGE = 1 in error. After further investigation of this conflict, LM determined the actual miss was caused by PR_28096; trajectory build fails with ALGORITHMIC_FAILURE. All of the other six conflicts were misclassified as DISCARD and correctly reclassified as MA under the same circumstances as the conflict between N1926S and USA527.

Two of the 13 differences were instances where a conflict was classified as a VA by EOPD and DISCARD by CPAT. One example of this difference was with Incident 2 where EOPD detected start of conflict at 69070 (19:11:10) and CPAT detected start of conflict at 68970 (19:09:30). After investigating, LM reported the conflict start time was detected later than CPAT due to a gap in track data from 68940 (19:09:00) until 69070 (19:11:10) caused by track unreasonableness. CPAT tools detected all track points to be reasonable. LM reported to CPAT that the unreasonableness in the track detected by EOPD could possibly be a PR. This issue should be verified in the future.

Three of the 13 differences were instance opposite of the previous case where a conflict was classified as a DISCARD by EOPD and VA by CPAT. For Incident 3, CPAT tools and EOPD both detect gap in track from 68590 (19:03:10) to 68730 (19:05:30), and ACST of 68730 (19:05:30). Since the conflict started within 5 minutes of a gap in track, CPAT considered it to be

[12] We use *Incident X* to refer to conflict events between aircraft pairs or an aircraft with an SAA and non-conflict encounter events between aircraft pairs or an aircraft with an SAA, where *X* is used to enumerate the events within the "Comparison of CPAT and LM Analyses" section.

a popup conflict, hence classified the prediction as a LATE_VA (See Appendix A – Processing to Analyze Conflict Predictions). EOPD and CPAT tools are in complete agreement except for conflict prediction result. EOPD failure to label conflicts as a popup should be resolved.

The last difference was with Incident 4. In this case no alert existed for the conflict, but CPAT tools discarded the conflict based on adherence age, and EOPD found no fault in adherence, therefore classified as MA.

EOPD detected a total of six conflicts in which CPAT tools did not detect. These conflicts appear in the blue portion of Table 28. Of the six, three were classified as VA and three were DISCARDs. In four of the six conflicts only detected by EOPD, CPAT detected gaps in track around EOPD's defined ACSTs. These gaps contributed to CPAT tools not detected the conflicts. One of the conflicts, Incident 5, was not detected by CPAT tools because the conflict start time was after the center handoff time (PR number TBD). The closest point of approach of the remaining conflict, Incident 6, CPAT tools did not detect was very close to the minimum horizontal and/or vertical separation standards, therefore rounding errors of different trackers were considered to contribute to the discrepancy.

In addition, CPAT tools detected a total of 10 conflicts[13] in which EOPD did not detect. These conflicts appear in the red portion of Table 28. All the conflicts were classified as VAs and all but two were of short duration (20 seconds). Therefore, these undetected conflicts were not considered critical to the RFR results.

During the RFR, EOPD classified eight conflicts as MAs and passed the missed AC-AC conflict rate requirement (ERD1879-C3). After comparing EOPD's results to CPAT tools, six additional MAs were detected; EOPD gained one in which CPAT discarded, but lost seven as previously described. These six additional missed conflicts caused ERAM to fail the missed AC-AC conflict rate requirement, and this requirement was re-statused during the RFR Post Test Briefing. In the future, EOPD's PR_28728 should be resolved for further ERAM accuracy testing and re-statusing.

4.4.1.2 ERAM Prediction Comparison of False AC-AC Conflicts

CPAT tools detected a total of 3361 FPCs and OFPD detected a total of 3394 FPCs. Of these totals, both tools detected 3355 of the same FPCs, which appear in the green portion of Table 29. The notification times of these 3355 FPCs generally agreed among the tools, however one differed by 23 seconds and very few of the notification times differed by one second. There were 205 FPCs which both tools classified as FAs and 2284 FPCs that were discarded by both tools[14]. Moreover, EOPD classified 49 FPCs as FAs when CPAT tools classified as DISCARDs. Even though EOPD allows these FAs, CPAT tools discarded approximately 50% of these FPCs because of adherence and the other 50% based on no track at PCST. Furthermore, EOPD discarded 817 FPCs in which CPAT tools classified as FAs. This is a very significant discrepancy because it represents approximately 25% of the FPCs detected by both tools. Of the 817 FPCs in question, 723 were discarded based on having an ATC clearance change. This issue

[13] Originally there were 11 conflicts total, but a bug in CPAT tools' flight path data validation existed where spikes in altitude were mishandled causing an extra conflict to be detected. This PR was reported to LM, corrected, then CPAT tools was reran, and the results documented above reflect this upgrade.
[14] These numbers of FAs and DISCARDs are after a bug in CPAT tools was fixed. The bug was in the discarding of FAs based on change in ATC clearance at notification set end time. CPAT tools also required the FA to be retracted (see Appendix A) to be discarded in this situation, which is incorrect. This bug had minor impact of the results.

was reported to LM, and LM confirmed that EOPD has a problem with discarding with the exception of RETRACTED_BY_CLEARANCE_CODE (CLR_FA_DISCARD_A). LM has registered this bug in EOPD as PR_30869. Furthermore, 62 of the 817 FPCs discarded by EOPD only were discarded because of no track at the PCST. After investigating it was found that, like OFPD, EOPD considers when a flight is within an APDIA that no track exists; hence the FA can be discarded, but it is wrong to consider track within an APDIA as having no track. This is an error in the logic for discarding FAs, and has been registered as PR number TBD by LM. Also, 73 of the 817 FPCs discarded by EOPD only were based on adherence. These PRs should be addressed prior to future EOPD runs.

Table 28: Matrix containing ERAM's actual AC-AC conflict prediction counts as detected by CPAT and LM.

LM CPAT	Valid Alert	Missed Alert	Discarded Alert	No Conflict	Totals (CPAT)
Valid Alert	191	0	3	10	204
Missed Alert	0	7	7	0	14
Discarded Alert	2	1	9	0	12
No Conflict	3	0	3	0	6
Totals (LM)	196	8	22	10	236

There was a total of 39 FPCs detected by EOPD that were not considered FPCs by CPAT tools, which appear in the blue portion of Table 29. Ten of the 39 FPCs were not considered because CPAT tools detected actual conflicts to match to the notification sets. The rest of the FPCs were considered to be no calls by CPAT tools, which means CPAT tools did not detect 29 of the same notification sets. Of these 29 only considered by EOPD, one was a FA, the rest were discarded. Furthermore, CPAT tools detected six FPCs in which EOPD did not consider. Five of the six were because EOPD detected actual conflicts in which CPAT tools did not. The remaining FPC was simply a no call detected by EOPD, and classified as a FA by CPAT tools.

Table 29: Matrix containing ERAM's false AC-AC conflict prediction result counts from CPAT and LM.

LM CPAT	False Alert	Discarded Alert	No FPC	Totals (CPAT)
False Alert	205	817	3	1025
Discarded Alert	49	2284	3	2336
No FPC	5	34	0	39
Totals (LM)	259	3135	6	2400

4.4.2 ERAM Aircraft-to-Airspace Conflict Prediction Comparison

The comparison of ERAM's AC-AS conflict prediction results is divided into two parts. The first part describes the comparison of the conflict prediction results for actual AC-AS conflicts. The second part details the comparison of conflict prediction results of false AC-AS conflicts.

4.4.2.1 ERAM Prediction Comparison of Actual AC-AS Conflicts

CPAT tools detected a total of 2672 AC-AS conflicts and EOPD detected 2683 AC-AS conflicts. Of these two sets of conflicts, 2664 were detected by both tools having the same aircraft and if not same start time, then very close, therefore considered to be the same conflicts. These matched conflicts appear in the green portion of Table 30. Of the 2664 conflicts detected by both

tools, 2567 conflicts had the same alert classification; hence 97 disagreements existed among alert classifications within the conflicts detected by both tools.

Table 30: Matrix containing ERAM's actual AC-AS conflict prediction result counts as detected by CPAT and LM.

LM CPAT	Valid Alert	Missed Alert	Discarded Alert	No Conflict	Totals (CPAT)
Valid Alert	2476	29	12	7	2524
Missed Alert	2	81	10	1	94
Discarded Alert	3	11	40	0	54
No Conflict	12	5	2	0	19
Totals (LM)	2493	188	2	8	2691

The most interesting aspect of the 2567 conflicts classified the same by both tools is the discrepancies in the detected ACSTs by both tools. For instance, the greatest ACST difference occurred with Incident 8, which contains an aircraft in a conflict with SUA7. CPAT tools detected the start of this conflict at 74680 (20:44:40), while EOPD detected it at 73650 (20:27:30) (See Table 31). CPAT tools detected that the aircraft was in conflict with SUA7 at 73650, but that it was the end of the previous conflict, because after 73650 the aircraft was within an APDIA for 1030 seconds. This 1030 second gap in conflict processing caused CPAT tools not to merge the conflicts together, but considered to be two conflicts. EOPD detected that the latter conflict started at 73650 for one point, and then merged this point with the conflict detected after the 1030 seconds of being within an APDIA. Therefore, the issue is that EOPD detects the start of the second conflict at the point of entering the APDIA region instead of at its departure. This is an incorrect way of defining conflict durations, and was caused by an error in EOPD's processing. This issue affected 23 conflicts total in the RFR dataset. Also, this issue was confirmed by LM, and registered as a bug in EOPD as PR_30191.

Table 31: Conflict start and end time information of conflicts between GEN0017 and SUA7.

INCIDENT NO.	CPAT_ACST	CPAT_ACET	EOPD_ACST	ACST_DIFF
Incident 7	73490	73650	73480	-10s
Incident 8	74680	74800	73650	-1030s

CPAT tools classified two conflicts as MAs when EOPD classified as VAs. These conflicts were within Incident 9 and 10. In both of these cases EOPD considered the conflict to be a popup, hence classified as a LATE_VA. CPAT considered Incident 9 to be a LATE_MA, therefore not considering the conflict to be a popup like EOPD did. One case where a conflict is considered to be a popup is when the ACST is less than some parametric time (in our analysis 300 seconds) later than the receipt of an ATC clearance. According to CPAT tools, this conflict was very close to a popup, because an ATC clearance was issued exactly 300 seconds[15] prior to the ACST, which means EOPD must use less than or equal to some parametric time. Incident 8 was a case where CPAT tools could not match it with an appropriate notification set. A notification set did exist for this conflict but the start and end times were 80895 (22:28:15) and 80897 (22:28:17), but the ACST was 80890 (22:28:10), hence did not fit in the notification set time window.

[15] After the ATCoach scenario was ran through ERAM, and in the process of creating the CMS scenario for CPAT, lag in the simulation process delayed the ATC clearance by two seconds, therefore putting the clearance on the 300 seconds mark, rather than 302 seconds.

CPAT tools classified 29 conflicts as VAs when EOPD classified as MAs and 12 when EOPD classified as DISCARDs. CPAT tools considered all 41 conflicts to have been popups, which means CPAT tools labeled these conflicts as LATE_VAs. LM confirmed that these 41 popup conflicts should have been considered to be popups by EOPD, and therefore not classified as MAs. This bug in EOPD has been identified and registered by LM as PR_30868. This PR should be addressed prior to future EOPD runs.

There were 11 instances where CPAT tools classified a conflict as a DISCARD and EOPD considered it to be a MA. Furthermore, there were 10 instances of the opposite situation where CPAT tools classified a conflict as a MA and EOPD classified as a DISCARD. In all of these conflicts the tools agreed whether or not there was an appropriate notification set for the conflict; the only discrepancy existed in determining whether the conflict is excused or not based on whether the flight is adhering to its route. This discrepancy could have been another result of EOPD PR_28728, which was described for the ERAM AC-AC conflict comparison in Section 4.4.1.

Additionally, two conflicts[16] were classified as DISCARDs by CPAT tools and VAs by EOPD. One of these conflicts, Incident 11, was of the previously mentioned case in which EOPD incorrectly detected start of conflict prior to entering an APDIA instead of at the time of departure. Incident 12 was the second conflict of discrepancy was labeled as LATE_VA by EOPD and CPAT tools labeled as LATE_DISCARD, which means EOPD considered this conflict to be a popup and CPAT tools did not. CPAT tools and EOPD agree that the conflict starts at 83570 (23:12:50). During the process of determining if a conflict is a popup, CPAT tools registered an interim altitude for FL100 at 83266 (23:07:46), which is just passed the threshold of 300 seconds within the ACST, hence it is not considered to be a popup conflict.

4.4.2.2 ERAM Prediction Comparison of False AC-AS Conflicts

CPAT tools detected a total of 2173 FPCs and EOPD detected a total of 2166 FPCs. Of these totals, both tools detected 2155 of the same FPCs, which appear in the green portion of Table 32. The notification times of these 2155 FPCs generally agreed among the tools, however six differed by a significant amount of time and very few of the notification times differed by one second. Within the six FPCs with large notification set start time differences only one, Incident 14, resulted in a different conflict prediction result where EOPD discarded the FA based on AFTER_LAST_CONFLICT_IN_ENCOUNTER (CFL_FA_DISCARD) and CPAT tools did not discard the FA. CPAT tools detects that the last actual conflict in the encounter starts after the notification set start time, hence does not discard the FA. There were 69 FPCs which both tools classified as FAs and 1593 FPCs that were discarded by both tools**.** There were 62 cases where CPAT tools discarded a FA and EOPD determined no exception for the FA, hence keeping it. The difference here mainly is attributed to differences in adherence age calculation and usage of ATC clearances between CPAT tools and EOPD. CPAT tools determined to discard many of these 62 FAs based on adherence or reception of an ATC clearance at PCST. Furthermore, EOPD discard 431 FAs in which CPAT tools kept. The majority of these FAs were wrongly discarded by EOPD based on clearance at PCST, as it was the case in the previous section with AC-AC FAs. This error was documented as PR_30869. Many others of the 431 FAs only

[16] Originally, there were three AC-AS conflicts in this case, but a discrepancy between CPAT tools and EOPD methodology of matching unused notification sets to conflicts was discovered caused CPAT tools to have an extra missed conflict (Incident 13). EOPD allows a conflict to be paired with a notification set when the ACST is equal to the notification end time, CPAT tools previously did not, but has been modified appropriately.

discarded by EOPD have to due with the previous error, PR number TBD**,** where track within APDIA is considered to be invalid track, therefore incorrectly discarding many FAs.

Of the 11 FPCs only considered by EOPD, appearing in the blue portion of Table 32, seven were matched with actual conflicts and determined to be VAs, therefore four of the remaining FPCs were actual no calls by CPAT tools. Only one of these four was a kept FA, the rest were discarded by EOPD, hence this discrepancy had little affect on the results. Of the 18 FPCs only considered by CPAT tools, appearing in the red portion of Table 32, 13 matched to actual conflicts in the EOPD results, and out of the five remaining FPCs, only two were FAs, therefore only modestly altering the total number of FAs.

Table 32: Matrix containing ERAM's false AC-AS conflict prediction result counts from CPAT and LM.

LM **CPAT**	**False Alert**	**Discarded Alert**	**No FPC**	**Totals (CPAT)**
False Alert	69	431	6	506
Discarded Alert	62	1593	12	1667
No FPC	1	10	0	11
Totals (LM)	132	2034	18	2184

5 Test Metrics for Trajectory and Strategic Alert Processing

The previous sections have presented results based on the metrics defined formally in the ERAM Test Program. However, the AMTWG's main objective is to develop metrics that go beyond the testing program's defined requirements. In this section, metrics will be defined first for trajectory predictions and then strategic conflict predictions in Sections 5.1 and 5.2, respectively.

5.1 Trajectory Prediction Metrics

The trajectory prediction requirements at least at the formal A-level test focused only on the vertical dimension. Thus, the altitude modeling results reported in Table 2 are strictly reporting the error rates in predicting the altitude of the aircraft. The metrics proposed here include all three dimensions, including the lateral side-to-side spatially coincident cross track error, the longitudinal spatially coincident along track error, and time coincident vertical error. These are defined in detail in reference (Paglione, 2007 – AIAA GNC paper).

For comparison of the legacy system to the new ERAM system, it is first necessary to confirm that the two systems modeled the same aircraft. The metrics above are applied only when both systems do indeed generate trajectories for the given aircraft. However, there is no guarantee that either system, due to functional software errors or by design, build a trajectory for every flight at all times. Figure 32 below illustrates a Venn diagram of the three possibilities. Area A in the figure occurs when the legacy system generates a trajectory but new ERAM system does not. Area C is the opposite, where ERAM generates a trajectory that the legacy system does not. Area B includes the flights that have trajectories generated for both systems. As stated above, only area B situations can be compared using the detailed metrics of cross, along, and vertical trajectory errors.

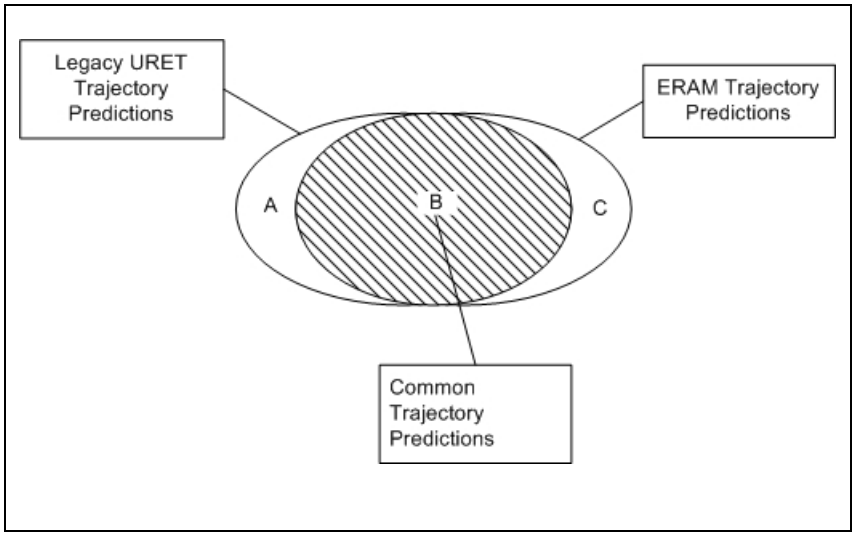

Figure 32: Overlap of Trajectory Modeling

For the non-time shifted RFR run between URET and ERAM, Table 33 provides a listing of a set of statistics that illustrate the context of the later more detailed error measurements. These statistics serve to illustrate the relationships expressed in Figure 32. For this study, ERAM has 29 out of 2275 less flights available for modeling due to tracker related issues, unrelated to the

trajectory prediction function. Due to flight path validation techniques, see reference (Ryan, paper on fpv) for details, ERAM loses another five flights. Next, ERAM, due to unresolved function issues in the software at the time of the RFR testing, did not generate any trajectories for 38 of these flights. URET did the same but only for one flight. This resulted in about a two percent loss in measurements for ERAM. The lost flights are being examined individually with the ERAM Development Contractor's software engineering team.

Table 33: Trajectory Prediction Context Statistics

#	Metric Description	URET	ERAM
1	Number of flights with paired track	2275	2246
2	Number of flight tracks passing flight path validation checks	2275	2241
3	Number of flights without any trajectories available	1	38
4	Number of flights with >= one trajectory measurement	2262	2201
5	Percentage of available flights with trajectory measurement = (4)/(2)	99.43%	98.22%

When correlating the measurements for the same flights between the two systems as illustrated in Figure 32, 2198 flights are common to both systems. These matched flights are Figure 32's area B, while areas A and C are the individual quantity of flights with measurements for URET and ERAM, respectively. These are listed in row (4) in Table 33. Of the 61 flight net difference between the two systems, exactly 64 flights had trajectories and measurements for URET but not ERAM and only 3 for ERAM but not URET. In ERAM's case, several flights had no trajectories available due to problems in either the simulation process or unresolved functional issues in ERAM's trajectory predictor. For URET, the flights missing measurements had very short tracks in the recording and limited availability for measurement.

For the 2198 matched flights, the process begins by application of the Interval Based Sampling Technique (IBST) on each system, ERAM and URET individually (Paglione, 2007 – AIAA GNC). The focus for this application is comparison of the resulting two spatially coincident horizontal measures, cross track and along track errors, and the time coincident vertical error. These originally signed errors are unsigned and sample mean calculated per flight. Next, the flights are paired for each system and the difference calculated. Thus, the mean error of URET for the given flight is subtracted by the same flight's error in ERAM. A positive difference indicates URET had a larger error for the given flight. If negative, ERAM has the larger average error. As described in (Paglione, 2007 – AIAA GNC), the paired statistical test is applied, providing evidence to indicate if the differences for all flights is statistically significant. Figure 33 illustrates the distributions of these differences for all flights for the formal RFR. Both cross track error and along track error are skewed negative, indicating ERAM has more flights with larger errors than URET. Vertical error seems only slightly skewed.

Table 34 provides the tabular statistical results of the application of the paired statistical standardized t-Test. For the cross track error the average difference is about -0.2 nautical miles between the two systems. The standard t-Test is applied to determine if the difference can be considered zero. This is not the case for all three of the metrics below. Along track error is much larger with an average difference of -1.2 nautical miles. Vertical error is statistically significant but borderline not practically significant at -110 feet. Thus, the results of the statistical tests are very consistent with the histograms of the differences' distributions illustrated in Figure 33.

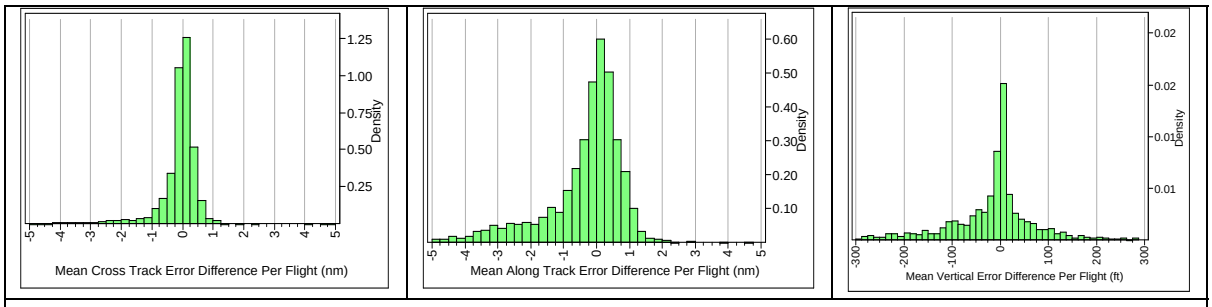

Figure 33: Distributions of System Differences per Flight Mean Errors

Table 34: Paired Error Statistical Results

Description	Cross Track Error (nm)	Along Track Error (nm)	Vertical Error (ft)
Mean Difference	-0.1906	-1.1735	-109.52
Degrees of Freedom	2197	2197	2197
Std Dev of Differences	1.05624	3.50605	491.784
	t-Test Results		
Test Statistic	-8.4589	-15.692	-10.441
Prob > |t|*	<.0001	<.0001	<.0001
Prob > t*	1.0000	1.0000	1.0000
Prob < t*	<.0001	<.0001	<.0001
*Probabilities that are small, less than 0.05, are very statistically significant. Allowing rejection of the hypothesis.			

5.2 Conflict Prediction Metrics

The conflict prediction accuracy was discussed in Section 3.2. However, the analysis focused on the defined requirements of acceptable rates of missed and false alerts. The conflict prediction requirements were determined by measuring the legacy URET system, refreshing the requirements, and then applying the basically the same input traffic scenario into ERAM and measuring its accuracy. To fail the requirement, meant failing a statistical hypothesis test that the rate of missed or false alert events was large enough relative to the requirement that it was unlikely to have occurred due to chance.

There are two main weaknesses in the approach defined in Section 3.2. First, both the rate of missed and false alerts for the legacy system and new system are random variables. The test assumed the requirement as measured from the legacy system was fixed. Second, the errors for both systems were lumped together into a ratio and only net effects were being compared. For example, if the ERAM system had a total of two more missed alerts than the legacy URET system; it is only considering the net difference. In reality for this example, ERAM had four missed alerts that the legacy system had correctly predicted, yet two more missed alerts were generated by the legacy system that ERAM correctly predicted. The test only compared the net quantities of missed alert events. A more sensitive test would compare the same conflict and alert events, reporting all mismatches. Furthermore, it is necessary to identify all the specific error events from a practical standpoint, so software corrections can be made. This section will present a method to identify and statistically compare these events.

Before discussing this comparison as a review of Section3.2, Table 35 lists the individual reason codes for each run's conflict prediction results. The alert type falls into four categories. First, the missed alert and false alerts are the two errors being measured. Next, the valid alerts are the

correct prediction of a conflict and discards are the events excused due to out of adherence situations or other artifacts of the traffic sample being used.

Table 35: Conflict Prediction Result's Main Reason Codes[17]

REASON CODE	ALERT TYPE	REASON DESCRIPTION
STD_VA	Valid Alert	Standard Valid Alert
LATE_VA	Valid Alert	Late Valid Alert, Valid since conflict was determined a pop-up
NO_CALL_MA	Missed Alert	Missed Alert due to no call (no alert at all)
LATE_MA	Missed Alert	Late Missed Alert
NO_CALL_DISCARD	Discard Alert	Missed Alert no call discarded since out of adherence
LATE_DISCARD	Discard Alert	Late Discard since out of adherence
NO_TRK_FA_DISCARD	Discard Alert	No post processed track at predicted conflict start time so discard
NO_ADHER_FA_DISCARD	Discard Alert	Out of adherence at predicted conflict start time so discard
CLR_FA_DISCARD	Discard Alert	Retracted False Alert assigned by an ATC clearance so discard
CFL_FA_DISCARD	Discard Alert	False Alert notified beyond last conflict actual start time so discard
STD_FA	False Alert	Standard False Alert
RETRACT_FA	False Alert	Retracted False Alert, notification end time earlier than predicted conflict start time

The sets of conflict predictions generated by the legacy system run (e.g. Run A) and a new system (e.g. Run B) are first evaluated separately. The analysis results in a database table with records labeled with the reason codes defined in Table 35. Next, these evaluations are compared, which is only meaningful when both runs are provided the same input traffic scenario (this will be discussed in detail later in Section 5.2.2). The first column in Table 36 lists all combinations of intersection and union of the events from Table 35. As in Table 36 and throughout this section, the first run compared will be referred to as Run A and the second as Run B. For example, Run A may generate a missed alert that is either a missed alert or valid alert in Run B and vice versa.

To determine the various combinations of comparative events as defined in Table 36, CPAT wrote a software tool to extract them from the conflict prediction results of a pair of conflict probe runs. The program produces a database table of entries with evaluation codes for each of these events. A listing of these combinations and their corresponding codes are listed in the following Table 37.

5.2.1 Statistical Approach of Comparing Conflict Predictions

The most critical quantities to determine a statistical difference between runs are the missed alert probability and the false alert probability, as quantified in Sections 3.2.5.2 and 3.2.5.3, respectively. One approach to determine if the difference is statistically significance is to utilize a binomial distribution and perform a hypothesis test concerning the difference between population proportions [Devore, 2000]. However, this technique assumes that the respective runs are

[17] This table summarizes the total 16 detailed codes in Section 3.2 to 12 reason codes that capture the essence of the processing involved.

Table 36: Comparison of Two Run's Resulting Alert and Conflict Event Combinations

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A and B	Both predicts conflict and it occurs (V_{A1}=V_{B1}-- valid alerts both)	Both predicts conflict and it does not occur (F_{A1}=F_{B1}-- false alert both)
ALERT by A and not B	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only)	B does not predict conflict and it does not occur (NC_B -- correct no-calls/discards B only)
ALERT by A and B ALERT or non-ALERT is discarded	A predicts conflict and it occurs (V_{A3} -- valid alerts by A only)	A predicts conflict and it does not occur (** F_{A2} Continued **)
	B does not predict conflict correctly but is discarded ($Discard_B$ -- B discards only)	B does not predict conflict correctly but is discarded (** NC_B Continued **)
ALERT by B and not A	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
	A does not predict conflict and it occurs (M_{A2} -- missed alert by A only)	A does not predict conflict and it does not occur (NC_A -- correct no-calls/discards A only)
ALERT by B and A ALERT or non-ALERT is discarded	B predicts conflict and it occurs (V_{B3} -- valid alerts by B only)	B predicts conflict and it does not occur (** F_{B2} Continued **)
	A does not predict conflict correctly but is discarded ($Discard_A$ -- A discards only)	A does not predict conflict correctly but is discarded (** NC_A Continued **)
NO ALERT by both Runs A and B	Both do not predict conflict and it occurs (M_{A1}= M_{B1} -- missed alert by both)	Both do not predict conflict and it does not occur (NC -- correct no-calls by both)
Total Number of Alerts for each/both	Total Number of Conflicts *(Same for both Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both Runs!*)

Table 37: Conflict Prediction Comparison Program Evaluation Codes

Event	Evaluation Code	Description
V_{A1} or V_{B1}	SAME_VA	Both runs have valid alerts for the same conflict
M_{A1} or M_{A1}	SAME_MA	Both runs have missed alerts for the same conflict
F_{A1} or F_{B1}	SAME_FA	Both runs have false alerts for the same encounter
V_{A2} or M_{B2}	VA_MA	Run A has a valid alert and Run B has a missed alert for the same conflict
M_{A2} or V_{B2}	MA_VA	Run A has a missed alert and Run B has a valid alert for the same conflict
V_{A3} or $Discard_B$	VA_DISCARD	Run A has a valid alert while Run B discards the conflict
$Discard_A$ or V_{B3}	DISCARD_VA	Run A discards the conflict while Run B has a valid alert
F_{A2} or NC_B	FA_NC	Run A has a false alert while Run B either has no prediction or discards the alert for the same encounter
NC_A or F_{B2}	NC_FA	Run A either has no prediction or discards the alert while Run B has a false alert for the same encounter

independent. For this study, each run is not independent, since they are run with the same air traffic scenario and altered weather files.

An alternative technique is presented in [Kachigan, 1986], utilizing categorical data analysis techniques. For categorical data analysis, we examine the difference in frequencies not proportions. For this study, the frequencies directly relating the missed and false alert probabilities include the counts of these events. Paired counts that are mutually exclusive and exhaustive, which is required for this test, occur when the error event occurs in one run and the correct event occurs in the other.

For the missed alert analysis, the count of interest is the missed alert count in Run A when simultaneously getting a valid alert in Run B or vice versa for the opposite case. These include the counts V_{A2} or M_{B2} compared to the V_{B2} or M_{A2}. Therefore, the count of valid alerts in Run A and simultaneous missed alerts in Run B is statistically compared to the count of valid alerts in Run B and simultaneous missed alerts in Run A. These counts should be equally likely if the two runs are statistically equivalent. Calculating the ratio of the squared difference between the expected value of each run and the observed value can test this hypothesis. If the hypothesis is true, this ratio will follow a chi-squared distribution or χ^2 with one degree of freedom. The test statistic is as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where

O_i is the observed frequency in category i

E_i is the expected frequency in category i

k is the total number of categories

For this study, k is always two, since only paired runs are compared. For example, the observed frequencies are the extracted V_{A2}/M_{B2} and V_{B2}/M_{A2} counts for the two runs. Since the null hypothesis assumes both events are equally likely, both expected frequencies are equal and calculated from the following equation:

$$E_i = \frac{\sum_{j=1}^{k} O_j}{k} \quad (2)$$

The resulting test statistic in Equation (1) can be expressed as a probability or P-value[18] by assuming a chi-squared distribution with one degree of freedom. For example, let's say we observe a $V_{A2}/M_{B2} = 8$ and a $V_{B2}/M_{A2} = 22$. The expected frequency from Equation (2) is 15 for both values, and the resulting test statistic from Equation (1) is 6.53. Therefore for this example exercise, the P-value is 0.011. This expresses that the hypothesis that these runs have equivalent missed alerts is only about one percent likely and provides evidence to reject the null hypothesis. For this test in the study, a P-value, which is less than 0.10, is considered sufficient to reject the hypothesis.

False alert probabilities can be analyzed in an analogous way. For the false alert counts, the observed frequency of F_{A2}/NC_B and F_{B2}/NC_A are compared.

5.2.2 Issues for Different Input Scenarios

Similar to the trajectory accuracy performance, the input traffic scenario needs to be the same (or very close) to compare the two systems. Figure 32 illustrates the relationships in context of trajectory predictions. The relationships illustrated for trajectory predictions can be applied to compare conflict predictions as well. Assume for visualization purposes that the area below the two ovals in Figure 32 is the input test conflicts for the respective system's conflict prediction testing. If for example, the exact same scenario was input into both systems, the ground truth conflicts both systems need to predict would be the equivalent. These conflicts are represented by area B from Figure 32, since A and C would be zero. However, if there were differences in the two input scenarios, A and C would have some quantity. These differences could confound the analysis described previously.

For RFR runs between URET and ERAM described in this paper, the traffic scenario for URET was completely generated by the HCS and only time shifted by CPAT. These positions are not

[18] In [Devore, 2000], the P-value is defined as the "smallest level of significance at which the null hypothesis would be rejected when a specified test procedure is used on a given data set." Thus, the P-value is the probability of the null hypothesis has occurred, so a small P-value (less than 0.10) would indicate the null hypothesis unlikely and should be rejected and if large should be assumed correct.

just the input into running URET but are used as the basis of determining the conflicts used to evaluate its conflict predictions. In a similar manner, ERAM is input with a traffic scenario and time-shifted with the same values. However, for ERAM CPAT time shifts raw radar reports and ERAM tracks these positions. The positions are input internally to ERAM's subsystem, called the Conflict Probe Tool (CPT), where conflict predictions are calculated. The track position reports are used to evaluate its conflict predictions as well. The differences in tracking between the two systems can cause differences in the timing and even determination of conflicts between the two systems.

For the RFR run, URET had 248 conflicts, while ERAM had generated 229 conflicts[19]. Besides the differences in processing, problems in the ERAM surveillance data processing in the simulation environment for the test caused flights to be dropped, explaining the slightly lower conflict counts. These conflicts were then matched by aircraft identification strings and conflict start time (within a threshold of five minutes). There were 202 matched conflicts. Referring back to Figure 32, this would be equivalent to the area referred to as area B. There were an additional 46 conflicts supplied in the URET scenario that did not match in the ERAM version, represented as area A from Figure 32. For ERAM, there were 27 conflicts not matching to URET, represented by area C. Thus, if we applied the techniques described above in Sections 5.2.1 and labeled in Table 37, the comparison would not only be comparing the prediction differences but the differences in the traffic scenarios. The traffic scenario differences are outside the scope of conflict prediction accuracy and needs to be blocked from the analysis.

The only events that need to be blocked from the analysis are the mismatched conflict events of Figure 32's area A and C. From the statistical methods developed in Section 5.2.1, the relationship between VA_MA to MA_VA and NC_FA to FA_NC are the focus of the study. Thus, any of these events caused by mismatches in the conflict events (area A and C again) should be discarded. The events that remain should be applied to the tests.

5.2.3 URET-to-ERAM and ERAM-to-URET Comparison of Conflict Predictions

Ideally the software tool that identified the events in Table 37 would apply the blocking discussed above in Section 5.2.2, but at the time of the report this upgrade had not been completed. Therefore, an alternative approach was applied. First, both sets of scenarios, route files, and alert files were altered to set all computer identification strings (CIDs) to "000". Since all aircraft identification strings were unique for this study (CID changes were removed due to simulation issues in ERAM), the resulting scenarios could be applied to both system's analyses. Thus, the URET alert file was input into CPAT's analysis tools. Next, using the same traffic scenario, the ERAM alert file was input as well. The resulting pair of analysis results was compared. This was repeated in the other direction as well (i.e. ERAM run with ERAM traffic scenario and URET run with ERAM traffic scenario).

Table 38 lists the results of the four runs, coded as follows:

- U2U = URET run using HCS scenario and URET output files
- E2U = ERAM run using HCS scenario and ERAM output files
- E2E = ERAM run using ERAM scenario and output files
- U2E = URET run using ERAM scenario and URET output files

[19] The quantities of conflicts are slightly different (by one conflict) than those reported in Section 3.2. This was caused by minor upgrades in the software tools that were rerun here. They had eliminated some unreasonable track reports.

Therefore, U2U and E2U have the same count of conflicts and encounters because they are both using the HCS scenario – the source of these test events. The alert records and notification sets are different because they are produced by each system independently. The same is true for the ERAM supplied scenario for E2E and U2E. From the results, it is clear under both traffic scenarios, URET is outperforming ERAM during the RFR. Table 38 lists 5 to 29 and 14 to 4 missed alerts for URET to ERAM and ERAM to URET, respectively. False alert errors are similar. ERAM consistently has more false alerts and larger error rates compared to URET.

Table 38: Run Results for Comparison of ERAM and URET

Description	Run Code			
	U2U	E2U	E2E	U2E
Actual Conflicts	248	248	229	229
Actual Encounters	20616	20616	19529	19529
Alert Records	25364	30817	30817	25364
Notification Sets	2607	3611	3591	2613
Missed Alerts	5	29	14	4
False Alerts	653	1321	1026	682
Valid Alerts	234	196	203	214
Discards	1715	2065	2348	1713
Verifiable Conflicts	239	225	217	218
Verifiable Alerts	887	1517	1229	896
Missed Alert Prob, P(M|C)	0.021	0.129	0.065	0.018
False Alert Prob, P(F|A)	0.736	0.871	0.835	0.761
Valid Alert Prob, P(V|A)	0.264	0.129	0.165	0.239

Table 39 provides the detailed reason code counts for the four runs. There are two types of missed alerts: those that are caused by delayed notifications, LATE_MA, and those caused by not generating an alert at all, NO_CALL_MA, for a particular conflict. For ERAM and URET, both produce similar quantities of late missed alerts (LATE_MA). However, ERAM is generating significantly more no call missed alerts (NO_CALL_MA). There are several reasons for this. First, for the E2U run where ERAM is analyzed using URET's HCS generated traffic scenario, there are several flights that the ERAM tracker did not process. As a result, ERAM's conflict probe functions would not generate conflict predictions for these flights. For the E2E run, ERAM is analyzed using its track generated traffic scenario. For these conflicts, ERAM generated four times the number of no call missed alerts than URET (12 to 3). These events are directly related to several trajectory generation problems reported at the RFR post test meeting.

Table 39 shows that the quantity of false alerts, both retracted (RETRACT_FA) and standard (STD_FA) types, are consistently higher for ERAM versus URET for the RFR runs. Table 40 produces matched counts for the various reason codes. For false to correct no call (FA_NC) or correct no call to false alert, the URET to ERAM (i.e. U2U versus E2U runs) had 268 to 936 events. In other words, 268 of the URET false alerts were correct no calls for ERAM using the HCS scenario for both analysis runs. In contrast, URET had 936 correct no calls while ERAM had false alerts for these same events. Table 41 applied the statistical hypothesis test presented in Section 5.2.1. It shows that the difference between 268 and 936 was very much statistically significant. Similar results are reported in comparing the ERAM to URET analysis (i.e. E2E versus U2E runs). Table 40 reports that ERAM has 642 false alerts while URET correctly did not predict alerts, while ERAM generated 298 no calls for URET's false alert events. The contrast was tested in Table 41 as well and also shown to be statistically significant.

Table 39: Detailed Reason Codes for Comparison Runs of ERAM and URET

Reason Code	Run Code			
	U2U	E2U	E2E	U2E
CFL_FA_DISCARD	35	74	55	34
CLR_FA_DISCARD_A	376	486	430	299
LATE_DISCARD	3	1	3	3
LATE_MA	4	8	2	1
LATE_VA	184	154	162	163
NO_ADHER_FA_DISCARD	481	748	741	390
NO_CALL_DISCARD	6	22	9	8
NO_CALL_MA	1	21	12	3
NO_TRK_FA_DISCARD_A	0	0	50	65
NO_TRK_FA_DISCARD_B	253	220	275	305
NO_TRK_FA_DISCARD_C	555	499	515	539
NO_TRK_FA_DISCARD_D	6	15	270	70
RETRACT_FA	276	702	527	314
STD_FA	377	619	499	368
STD_VA	50	42	41	51

Table 40: Comparative Counts Between ERAM and URET Comparison Runs

Comparative Counts		Comparison Runs: Run A Versus Run B	
Event	Evaluation Code	URET Versus ERAM	ERAM Versus URET
V_{A1} or V_{B1}	SAME_VA	189	194
M_{A1} or M_{B1}	SAME_MA	2	1
F_{A1} or F_{B1}	SAME_FA	385	384
V_{A2} or M_{B2}	VA_MA	27	3
M_{A2} or V_{B2}	MA_VA	2	12
V_{A3} or $Discard_B$	VA_DISCARD	17	6
$Discard_A$ or V_{B3}	DISCARD_VA	5	8
F_{A2} or NC_B	FA_NC	268	642
NC_A or F_{B2}	NC_FA	936	298
Comparative Statistics		URET Versus ERAM	ERAM Versus URET
$P(M_A\|C_A)$		0.021	0.065
C_A		239	217
$P(M_B\|C_B)$		0.129	0.018
C_B		225	218
$P(M_A\|C_A) - P(M_B\|C_B)$		-0.108	0.047
$P(F_A\|A_A)$		0.736	0.835
A_A		887	1229
$P(F_B\|A_B)$		0.871	0.761
A_B		1517	896
$P(F_A\|A_A) - P(F_B\|A_B)$		-0.135	0.074

Table 40 reports the combination events of missed to valid alerts and valid to missed. Analogous to the false alert analysis described above, Table 42 reports both these combinations are statistically significant as well. The difference in the rate of missed alert ranged from 11 to 5 percent. For missed alert rates, these are very large differences between the two systems, in this case ERAM has consistently more errors than URET at the time of the RFR.

Table 41: Statistical Results for False Alert Comparison of ERAM and URET Runs

Statistics	Comparison Runs: Run A Versus Run B URET Versus ERAM	ERAM Versus URET
FA_NC	268	642
NC_FA	936	298
Total	1204	940
Expected	602	470
X^2	370.6	125.9
P-value	0.0000	0.0000

Table 42: Statistical Results for Missed Alert Comparison of ERAM and URET Runs

Statistics	Comparison Runs: Run A Versus Run B URET Versus ERAM	ERAM Versus URET
VA_MA	27	3
MA_VA	2	12
Total	29	15
Expected	14.5	7.5
X^2	21.6	5.4
P-value	0.0000	0.0201

6 Conclusions

The results of the study confirm that ERAM failed, as reported, many of the accuracy requirements for the formal Factory Acceptance Test Run-For-Record (FAT/RFR) runs. The procedures developed for this study will be applied in the future to ERAM test scenarios as the ERAM software problems are addressed in the regression testing phases of the program.

The first of the failed RFR requirements was the vertical trajectory accuracy requirement, listed in Table 1. Two examples were included in this report, illustrating that ERAM performed much worse than URET. In the first example, Section 3.1.4.1, URET generates 12 trajectories with an average altitude error of 2022 feet while ERAM generates 18 trajectories with an average altitude error of 6307. The numbers for the second example, Section 3.1.4.2, are 13 and 1789 for URET and 20 and 5840 for ERAM. Correctly predicting descent profiles is particularly challenging for all trajectory predictors and indicates here that several problems still existed in ERAM at the time of the RFR.

Table 1 listed the ERAM Run 4 and Run 5 pass/fail results for aircraft-to-aircraft and aircraft-to-airspace conflict probe requirements, based upon the use of the off-line Lockheed Martin OFPD and EOPD test tools. Table 20 was generated based upon CPAT tool generated warning time, missed alert, and false alert statistics, and utilized the same z-statistic and t-statistic analysis to quantify individual requirement pass/fail status.

ERAM passed the aircraft-to-aircraft conflict prediction warning time requirements as listed in Table 20. ERAM warning times were comparable to the URET baseline requirements. However, ERAM failed the aircraft-to-aircraft missed alert and false alert requirements as listed in Table 20. Missed and false alert rates for ERAM were generally at least twice as high as the corresponding rates for URET.

ERAM failed the aircraft-to-airspace missed alert and false alert rate requirements as listed in Table 20. Again the missed alert and false alert rates for ERAM were generally at least twice as high as the corresponding URET rates. ERAM did pass two of the five range “bins” of the aircraft-to-airspace false alert requirement, but a passing grade in all five “bins” is required to pass the overall false alert requirement.

In order to verify the conflict prediction performance results for URET and ERAM generated by LM’s OFPD and EOPD support tools, respectively, analyses were performed on the RFR data. These analyses involved comparing aircraft-to-aircraft (AC-AC) and aircraft-to-airspace (AC-AS) conflict predictions between LM’s OFPD and EOPD support tools and the FAA’s CPAT tools. As a result of these comparisons, we found that most of the conflict predictions results matched and the remaining discrepancies were identified. These discrepancies were detailed in this report, and documented in the list of PRs found in Appendix F – List of Problem Reports (PRs) in OFPD, EOPD, and CPAT tools. All CPAT tools PRs were fixed immediately, and rerun on the RFR dataset. Discrepancies in OFPD and EOPD were reported to LM and PRs were generated. Most of the discrepancies were in the ERAM conflict prediction performance results reported by the EOPD support tool. It is recommended that all PRs be addressed prior to future OFPD/EOPD runs to ensure validity in the conflict prediction performance results.

Finally, the third objective of this report produced a methodology and set of accuracy test metrics that covered the full range of trajectory and conflict prediction performance. Presented in Section 5 and in particular Figure 32, the methodology begins with a description of how two systems

input with the same traffic may not even model the same aircraft. Thus, a set of context statistics were defined and applied to the RFR runs. The results are presented in Table 33, showing there was a difference in about one percent in modeled aircraft flights between URET and ERAM with the later being lesser of the two.

For the flights modeled in common, a full range of trajectory accuracy metrics were defined, including both horizontal and vertical dimensions. This improved on the trajectory related requirements presented in Section 2, which focused solely on the vertical dimension. All three sets of metrics presented were statistically significant through application of a paired data test. However, the most practical difference was found in the along track dimension at 1.2 nautical miles on average. This indicates improvement is needed from the Drop 1 ERAM release used for the FAT/RFR.

An analogous approach was applied for conflict prediction accuracy. Utilizing a categorical statistical approach, the matching events of missed-to-valid and valid-to-missed were compared. Similarly the false-to-correct-no-call and correct-no-call-to-false alert were also compared. Results for the RFR runs were presented in Table 40. In summary, the both the missed and false alert comparisons were statistically significant, indicating several problems yet to be addressed in ERAM at the time of the RFR runs.

In conclusion and at the time of the FAT/RFR runs, ERAM's accuracy in aircraft trajectory and strategic conflict notification functions were significantly degraded as compared to the legacy URET system. This was illustrated in both the formal requirements and in the expanded test metrics presented in this paper. Besides the ERAM performance, the paper reviewed the performance of the ERAM measurement tool, EOPD. As a result of both ERAM's performance and EOPD, many problem reports have been generated and many addressed, since time frame of the RFR runs. The methodology and metrics presented here continue to be used to help determine if these upgrades of ERAM and EOPD have indeed addressed the problems. This will be subject of future reports.

List of Acronyms/Abbreviations

ACB-330	*Legacy* Simulation and Analysis Group, WJHTC, FAA
ACB-550	*Legacy* ERAM & ECG Group, WJHTC, FAA
ACET	Actual Conflict End Time
ACP	Azimuth Change Pulse
AC1	Aircraft Number 1
AC	Aircraft
AC2	Aircraft Number 2
ACID	Aircraft Identifier (Call Sign)
ACST	Actual Conflict Start Time
AIAA	American Institute of Aeronautics and Astronautics
AIR	Fictitious airline name
AMTWG	Automation Metrics Test Working Group
APDIA	Automated Problem Detection Inhibited Area
ARTCC	Air Route Traffic Control Center
AS	Airspace
ASCII	American Standard Code for Information Interchange
ATC	Air Traffic Control
ATCA	Air Traffic Control Association
ATM	Air Traffic Management
AWT	Actual Warning Time
BOC	Bottom of Climb
BWI	Baltimore International Airport
C	Conflict
CID	Computer Identifier
CMS	Common Message Set
COI	Critical Operational Issue
CPA	Closest Point of Approach
CPAT	Conflict Probe Assessment Team
CPT	Conflict Probe Tool
CSV	Comma Separated Version
DASC	Digital Avionics Systems Conference
DEL	Delete
DP	Departure Message
DS	Display System
ECG	En route Communications Gateway
EOPD	ERAM Offline Problem Determination
ERD	En Route Domain
ERAM	En Route Automation Modernization
ERIT	En Route Radar Intelligent Tool
FAA	Federal Aviation Administration
FDP	Flight Data Processing
FL	Flight Level
FPC	Falsely Predicted Conflict
GFP	Government Furnished Equipment
GH	General Information
GPS	Global Positioning Satellite System

GSGT	Graphical Simulation Generation Tool
GWH	LM driver for testing URET
HADDS	Host Air Traffic Management Data Distribution System
HCS	Host Computer System
HDG	Heading
Host	ARTCC main frame computer
IAD	Dulles International Airport
IBST	Interval Based Sampling Technique
IIF	Integration and Interoperability Facility
IRD	Interface Requirements Document
JMP	Statistical software from SAS
JVN	JVN Communications, Inc.
JSA	Joseph Sheairs Associates Inc.
LM	Lockheed Martin Corporation
MA	Missed Alert
MA_LATE	Missed Alert – Late
MA_STD	Missed Alert – Standard
MWT	Minimum Warning Time
MWTR	Minimum Warning Time Requirement
Nm	Nautical Miles
NS	Notification Set
OFPD	URET Offline Problem Determination
ORR	Online Radar Recording
PC	Personal Computer
PCET	Predicted Conflict End Time
PCST	Predicted Conflict Start Time
PROOF	Post processing simulation animation software
RFR	Run-For-Record
RVSM	Reduced Vertical Separation Minima
SAA	Special Activity Airspace
SAR	System Analysis Recording
SAS	Statistics Analysis System
SDP	Surveillance Data Processing
SDRR	Looped Simulation Drive System Replacement
SEP	Separation
SQL	Structured Query Language
SRTQC	Search Real Time Quality Control
SUA	Special Use Airspace
TOD	Top of Descent
TP	Trajectory Predictor
TPM	Technical Performance Measure
URET	User Request Evaluation Tool
UTC	Coordinated Universal Time (see www.time.gov/about.html)
VA	Valid Alert
VA_LATE	Valid Alert – Late
VA_STD	Valid Alert – Standard
WJHTC	William J. Hughes Technical Center
ZDC	Washington ARTCC

List of Selected Airport Identifiers

ACY	Atlantic City International Airport
BWI	Baltimore Washington International Thurgood Marshall Airport
DOV	Dover Air Force Base
IAD	Washington Dulles International Airport
ILM	Wilmington International Airport
ORF	Norfolk International Airport
RDU	Raleigh Durham International Airport
RIC	Richmond International Airport
ROA	Roanoke Regional Airport

References

1. Box, G. E. P., Hunter, J. S., Hunter, W. G., (2005), "Statistics for Experimenters", 2nd Edition, John Wiley & Sons.

2. Cale, M. L., Paglione, M. M., Ryan, H. F., Timoteo, D., Oaks, R. D., (1998) "User Request Evaluation Tool (URET) Conflict Prediction Accuracy Report," DOT/FAA/CT-TN98/8, WJHTC, Atlantic City: FAA, April. (See http://acy.tc.faa.gov/cpat/docs/ucpa_fin.pdf)

3. Cale, M. L., Liu, Shurong, Oaks, R. D., Paglione, M. M., Ryan, H. F., (2001), "A Generic Sampling Technique for Measuring Aircraft Trajectory Prediction Accuracy", 4th USA/EUROPE Air Traffic Management R&D Seminar, Santa Fe, NM, December.

4. FAA, (2002) "Blueprint for NAS Modernization 2002 Update," October (See http://www.faa.gov/nasarchitecture/Blueprint2002.htm).

5. Lockheed Martin Air Traffic Management, (2002) "User Request Evaluation Tool (URET) Core Capability Limited Deployment (CCLD) Test Report for the Formal Accuracy Test, Volume III," Lockheed Martin Corporation, Rockville Maryland.

6. Mondoloni, S., Swierstra, S., Paglione, M. M., (2005), "Assessing Trajectory Prediction Performance – Metrics Definition", 24th Digital Avionics Systems Conference (DASC), Washington, D.C., October.

7. Paglione, M. M., Ryan, H. F., Kazunas, S., Cale, M. L., (1997), "Generic Metrics for Conflict Probe Tools Developed for Free Flight", Proceedings of the Air Traffic Control Conference (ATCA), September.

8. Paglione, M. M., Ryan, H. F., Oaks, R. D., Summerill, J. S., Cale, M. L. (1999a), "Trajectory Prediction Accuracy Report: User Request Evaluation Tool (URET) / Center-TRACON Automation System (CTAS)", DOT/FAA/CT-TN99/10, WJH Technical Center, Atlantic City International Airport, NJ, FAA, May.

9. Paglione, M. M., Cale, M., Ryan, H. F., (1999b) "Generic Metrics for the Estimation of the Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool," Air Traffic Control Quarterly Journal, Air traffic Control Association (ATCA), Volume 7(3) 147-165. (See http://acy.tc.faa.gov/cpat/docs/atcq_051899.pdf)

10. Paglione, M. M., (2001) "Multi-Site URET CCLD Accuracy Refresh Parameters, Revision 2," Memorandum, FAA WJHTC/ACT-250, October 22.

11. Paglione, M. M., Oaks, R. D., Summerill, J. S., (2003a) "Time Shifting Air Traffic Data for Quantitative Evalution of a Conflict Probe," Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference, Austin TX, August. (See http://acy.tc.faa.gov/cpat/docs/2003_5343.pdf)

12. Paglione, M. M., Oaks, R. D., Bilimoria, K. D., (2003b) "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Technical Forum, November. (see http://acy.tc.faa.gov/cpat/docs/2003_6724.pdf)

13. Paglione, M. M., Baldwin, W. C., Putney, S., (2005a) "Host Radar Tracking Simulation and Performance Analysis," DOT/FAA/CT-TN05/31, WJHTC, Atlantic City: FAA, August. (See http://acy.tc.faa.gov/cpat/docs/SDPMetricsReportDrop2t.pdf)

14. Paglione, M. M., Garcia-Avello, C., Swierstra, S., Vivona, R., Green, S., (2005b), "Collaborative Approach to Trajectory Modeling Validation", 24th Digital Avionics Systems Conference (DASC), Washington, D.C., October.

15. Paglione, M. M., Oaks, R. D., (2006a) "Determination of Horizontal and Vertical Phase of Flight in Recorded Air Traffic Data", Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference, Keystone, CO, August.

16. Paglione, M. M., Ryan, H. F., Liu, Shurong, (2006b) "Evaluation of En Route Host Computer System's Tactical Alert Processing: Description of Methodology", DOT/FAA/CT-TN07/14, WJH Technical Center, Atlantic City International Airport, FAA, November.

17. Paglione, M. M., (2007) "Results and Recommendations from ERAM FDP/CPT Accuracy Tests (Run 4&5)", Memorandum, FAA WJHTC, October 9.

18. Ryan, H. F., Paglione, M. M., Green, S., (2004) "Review of Trajectory Accuracy Methodology and Comparison of Error Measurement Metrics", Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation, and Control Conference, Providence, RI, August.

19. Ryan, H. F., Paglione, M. M., (2005) "Comparison of Host Radar Tracks to Aircraft Positions from the Global Positioning Satellite System," DOT/FAA/CT-TN05/30, WJH Technical Center, Atlantic City: FAA, August. (See http://acy.tc.faa.gov/cpat/docs/SDPMetricsReportDrop1t.pdf)

20. SAS Institute Inc., (2003) *JMP Statistics and Graphics Guide.* Cary, NC: SAS Institute Inc.

21. WJHTC/ACB-500, (2003) "Test and Evaluation Master Plan En Route Automation Modernization (TEMP)," Atlantic City: FAA, October 24.

22. WJHTC/ACB-550 and ACB-330, (2004) "Progress Report of the Automation Metrics Test Working Group (AMTWG)," Atlantic City: FAA, June 16.

23. WJHTC/AOS-300, (2004) "ARTCC HOST Computer System Interface Requirements Document (IRD) for Air Traffic Applications", NAS-IR-8217-0001, Atlantic City: FAA, October 4.

24. WJHTC/ACB-330, (2005) "En Route Automation Modernization Automation Metrics and Preliminary Test Implementation Plan Automation Metrics Test Working Group," Version 2.7, Atlantic City: FAA, June.

Appendix A – Processing to Analyze Conflict Predictions

In this appendix we detail CPAT tools' procedures on measuring the accuracy of conflict predictions. We provide illustrations of these processes in Figures 30 to 32. These illustrations represent the processes for measuring AC-AC conflict prediction accuracy, however the processes for measuring AC-AS conflict predictions are slightly different, and when encountered in Sections 0, 0, and 0 these differences are clarified.

Process A – Load Conflicts and Build Notification Sets

The first step in CPAT tools' measurement of conflict prediction accuracy is to load the conflicts and their alerts from the database. The conflicts are stored as a collection in RAM and the alerts are built into various notification sets then stored into RAM as well.
For every conflict stored in the database, a Conflict object is stored in a Collection as shown in the initial flow of Figure 34. Next, in order to define the start of a notification set, AC-AC conflict alerts are filtered based on whether the starting alert was posted within a threshold amount of time (1 hour for RFR) prior to the predicted conflict start time, whether the predicted minimum horizontal separation is less than a parametric distance (8.0 nm for RFR), and whether to use muted alerts or not (only unmated alerts were used for RFR). For AC-AS conflict alerts no filtering is performed. Valid starting and ending (deleting) conflict alerts are then linked together to form notification sets. AC-AC notification sets can end with a message that is a Delete, has minimum horizontal separation less than parametric distance, or is muted. AC-AS notification sets end only with Delete messages. This produces a Collection of all notification sets, which CPAT tools uses in Process B and C to compute the conflict prediction accuracy results.

Process B – Valid and Missed Alert Processing

In Process B (see Figure 35), conflicts are evaluated in order of actual conflict start time and matched against eligible notification sets. To even be eligible for matching to a specific conflict, a notification set must have a posting time at or prior to the start of the actual conflict, must have an end or delete time at or after the start of the actual conflict[20], and must have not been previously matched with another conflict. The result is listing of Valid Alerts and Missed Alerts associated to all the input conflicts provided by the scenario.

There are two types of Valid Alerts: STD_VAs and LATE_VAs. To be a STD_VA, the actual conflict must have matched with an active notification set and the conflict prediction must have been presented to the controller at least a threshold number of minutes prior to the actual conflict start time (ACST), known as actual warning time (AWT). This threshold is called minimum warning time (MWT), and was 5 minutes during the RFR processing (see number 1 callout in Figure 35). If the conflict prediction fails the timeliness requirement it is considered to be a LATE_VA if the conflict is a pop-up. CPAT tools classify a conflict as a pop-up if one of the following cases is true for one (or both) of the aircrafts in AC-AC conflicts (only the one aircraft in AC-AS conflicts):

- The ACST is within MWT of the start of either aircraft's track.
- The ACST is less than MWT later than the receipt of an ATC clearance for either aircraft.
- The ACST is less than MWT later than the end of a gap in the track of either aircraft.

[20] Notification Start Time ≤ ACST ≤ Notification End Time

- The track report for either aircraft at the ACST is within an adapted altitude (300 feet for RFR) from a cleared interim or hold altitude.

Under these conditions a conflict probe would not be expected to predict the conflict beyond the MWT threshold. However, regardless whether the conflict is a pop-up, a Valid Alert needs to be posted prior to the ACST. If the conflict was not a pop-up, then it is considered either a Missed Alert or an excusable Discarded Alert. If the shorter of the times at which the aircrafts are adhering to their route is greater than the adherence age parameter time (see number 2 callout of Figure 35), then the conflict is considered to be a LATE_MA, otherwise it is considered to be a LATE_DISCARD, because the aircraft(s) are not laterally adhering to their route.

Conflicts that have no matching active notification set are considered to be unpredicted conflicts. Conflicts not predicted are considered to be NO_CALL_MAs, unless excused by adherence age, wherein then labeled as NO_CALL_DISCARDs.

Process C – False Alert Processing

The remaining notification sets not matched as Valid or Missed Alerts are potentially False Alerts. In Process C, (see Figure 36), the unmatched notification sets are evaluated to determine which are truly False Alerts and which can be discarded. There are several reasons for discarding False Alerts. In order to discard a False Alert one of the following cases must be true for both aircrafts in the predicted AC-AC conflict (only one aircraft for predicted AC-AS conflicts). Also, order matters with the inverse of each case, i.e. Cases 1 to (N-1) must be false for Case N to be considered.

1. If no surveillance data exist for either aircraft then the notification set is discarded as a NO_TRK_FA_DISCARD_A.
2. If the predicted conflict start time (PCST) is before the start of either aircrafts' track data then the notification set is discarded as a NO_TRK_FA_DISCARD_B.
3. If the predicted conflict start time (PCST) is after the end of either aircrafts' track data then the notification set is discarded as a NO_TRK_FA_DISCARD_C.
4. If a gap exists in either aircraft's track data at the PCST then the notification set is discarded as a NO_TRK_FA_DISCARD_D.
5. If the adherence age check is applicable and the minimum adherence age of both aircrafts at the PCST is less than or equal to the False Alert Planning Horizon Time[21] then the notification set is discarded as a NO_ADHER_FA_DISCARD.
6. If a primary ATC clearance of either aircraft exists between a top and bottom threshold (i.e. 5 seconds before and 1 second after) around the notification set end time then the notification set is discarded as a CLR_FA_DISCARD_A.
7. If a secondary ATC clearance of either aircraft exists between a top and bottom threshold (i.e. 5 seconds before and 1 second after) around the notification set end time then the notification set is discarded as a CLR_FA_DISCARD_B (only for interfacility mode).
8. If the notification set start time is before the ACST of the latest conflict for this aircraft pair then the notification set is discarded as a CFL_FA_DISCARD.
9. If the notification set was not discarded for any of the above cases, then it is considered to be a False Alert and if retracted[22] classified as a RETRACT_FA, otherwise is a STD_FA.

[21] False Alert Horizon Planning Time is equal to the difference between the PCST and notification set start time.
[22] A retracted notification set is when the notification set end time is before the PCST.

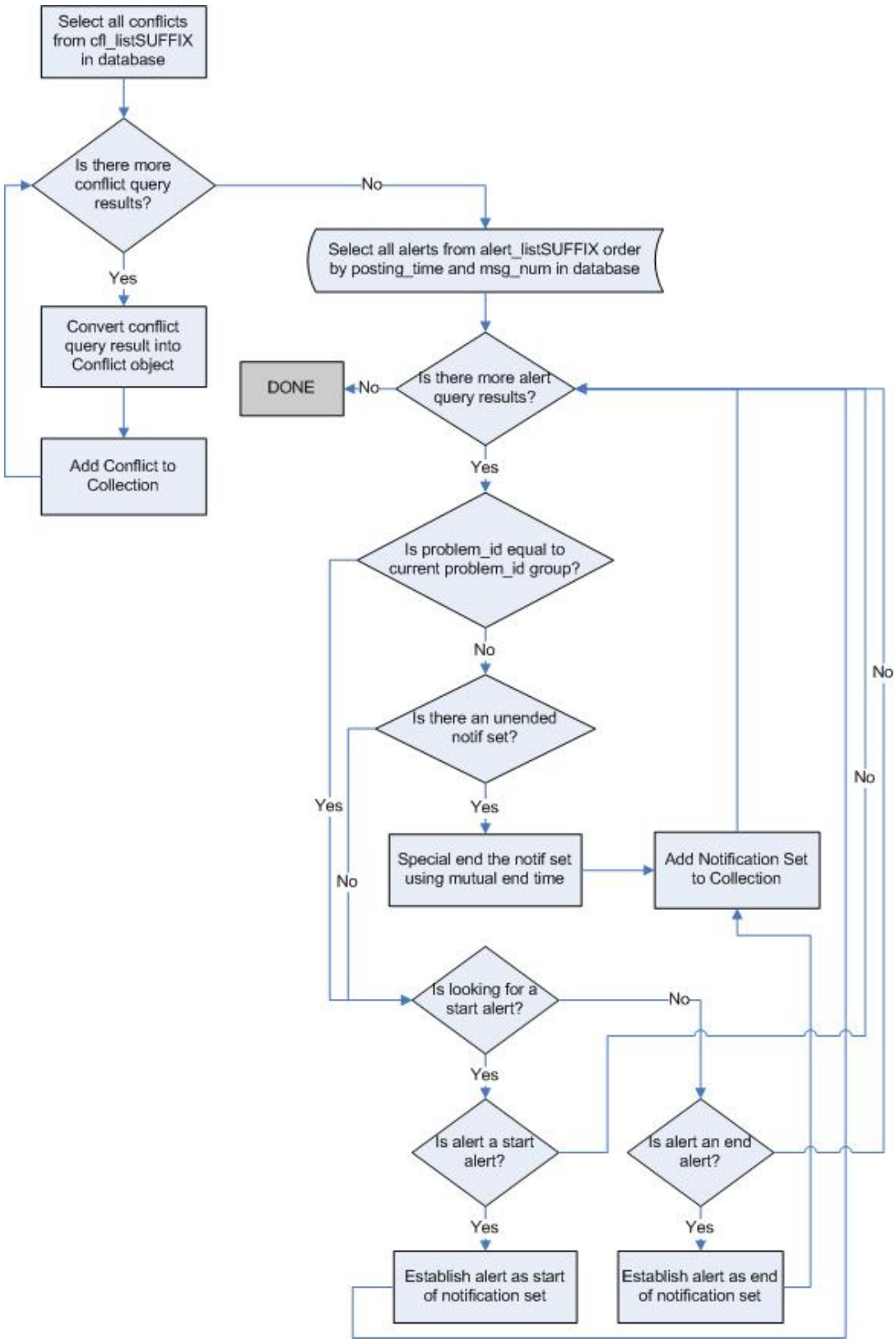

Figure 34: Process A - Loads conflict and build notification sets for processing in Process B and C.

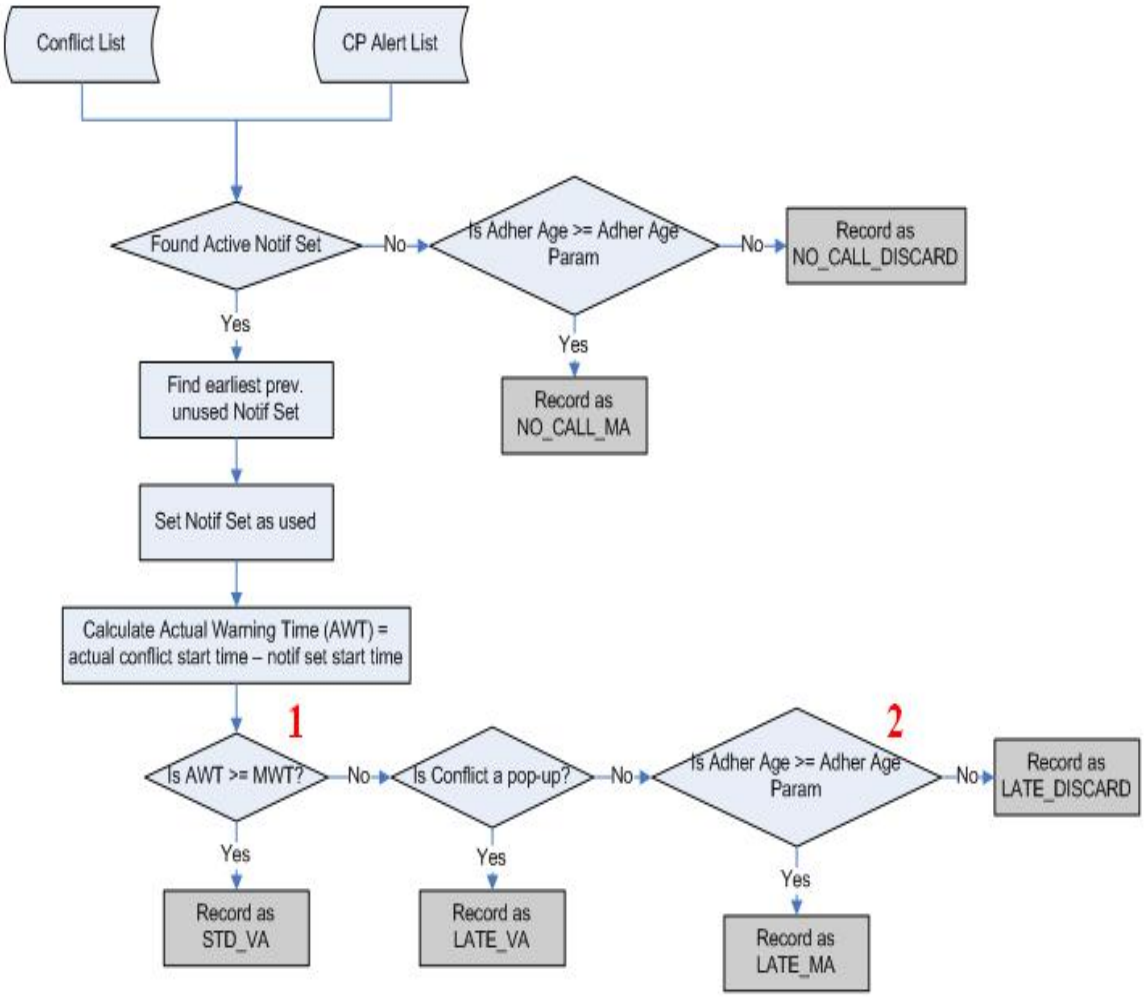

Figure 35: Process B - Evaluate all conflicts and determine all Valid and Missed Alerts.

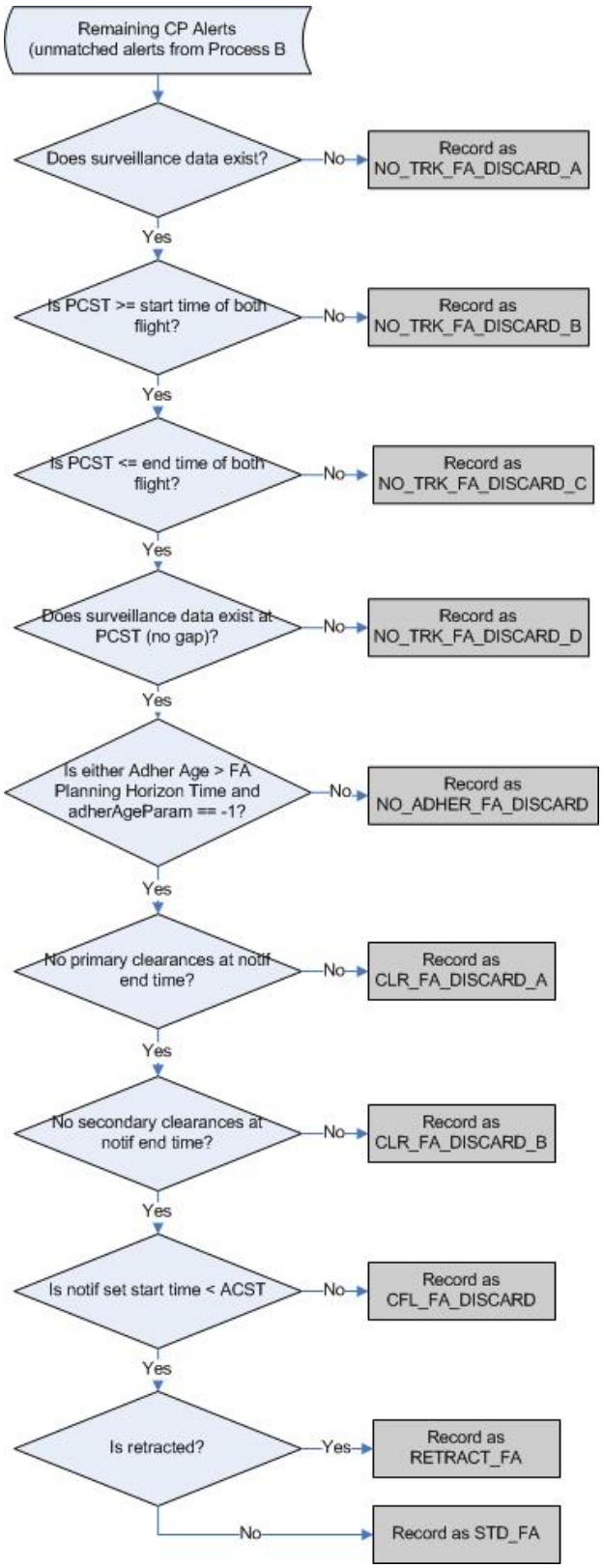

Figure 36: Process C - Evaluate all unmatched notification sets and determine False and Discard Alerts.

Appendix B – Trajectory Examples

Three examples have been selected from the test scenario to illustrate and compare ERAM and URET altitude modeling errors. Many trajectories are synthesized for each flight as the aircraft continues along its route. Two trajectories are given for each of the three flights.

Flight 1

The altitude prediction errors in Flight 1, the first trajectory, (Figure 37), are caused by unanticipated changes in the cruise altitudes and later on by mis-predicting the Top of Descent (TOD). The altitude errors in Flight 1, second trajectory (Figure 39), are caused by a TOD error. ERAM and URET make the same errors.

Flight 2

The first trajectory for Flight 2 makes a fairly good prediction of the aircraft's descending flight path in both ERAM and URET data. Flight 2 has a glitch in the radar track. The track jumps 21000 feet from 13000 feet to 34000 feet and back in 50 seconds. This glitch or spike occurs only in the track data processed by URET because, in the ERAM simulation, the track data is terminated at the point of handoff. The spike in the track data causes a trajectory to be built which starts at the 34000 foot altitude when the aircraft is actually at 13000 feet. An artifact in the simulation prevents ERAM from having the opportunity to make the same mistake. The cleanup processing of the bad track data by URET in inadequate.

Flight 3

The large altitude errors in the trajectories shown in Figure 45and Figure 46 are caused by predicting the TOD to be much later than actually occurred. The URET and ERAM trajectory predictions are the same. The altitude step at 27000 feet predicted in the climb out is in error as the actual step was at 14000 feet. The final cruise altitude is in error because the altitude was changed from 40000 feet to 43000 after the trajectories were built. In Figure 47 and Figure 48, the large errors are caused by the trajectories being built just before the interim altitude clearance to 14000 feet was removed.

The following two tables, Table 43 and Table 44, characterize the trajectories and their altitude errors in these examples. In this appendix two URET trajectories and two ERAM trajectories are shown for each of the three example flights. The trajectory errors are measured only when the trajectory is the active trajectory. When a new trajectory is built, the old trajectory is discarded. The count in the tables is the number of measurements made on each trajectory. Only one measurement was made on trajectory number 27037 because, in the scenario, the track was terminated immediately after the trajectory was built. In the ERAM scenario, the radar tracks were terminated when the aircraft were handed off to the next Center. The **START TIME** is the time of the first error measurement on the trajectory and the **END TIME** is the time of the last measurement.

Table 43: Altitude Errors for the URET Trajectories

FLIGHT	ACID	CID	TRAJECTORY BUILD TIME	TRAJECTORY NUMBER	COUNT	AVG ERROR	MAX ERROR	MIN ERROR	START TIME	END TIME
1	AIR0017	51P	70000	12832	51	1360	2000	0	70040	71540
1	AIR0017	51P	72448	12841	31	8189	19576	0	72470	73370
2	AIR0029	627	75619	23116	38	185	1817	0	75660	76950
2	AIE0029	627	78439	23132	13	22688	23795	20089	78450	78810
3	GEN0023	182	73781	18481	29	4468	8261	725	73820	75020
3	GEN0023	182	73917	18483	40	16475	29000	0	73940	75260

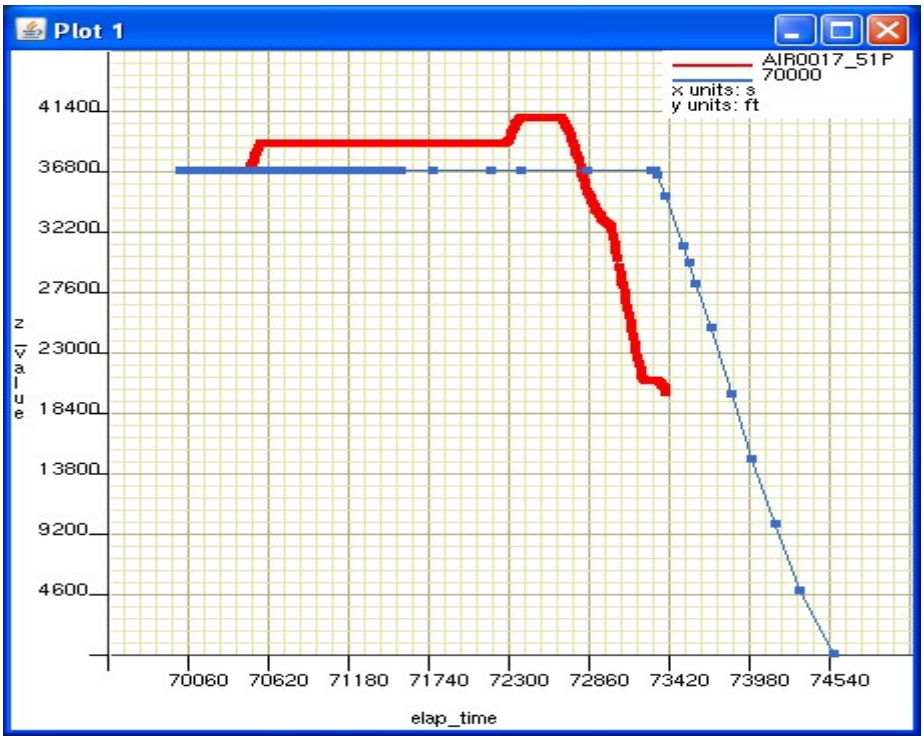

Figure 37: URET - Altitude profile and initial trajectory for Flight 1

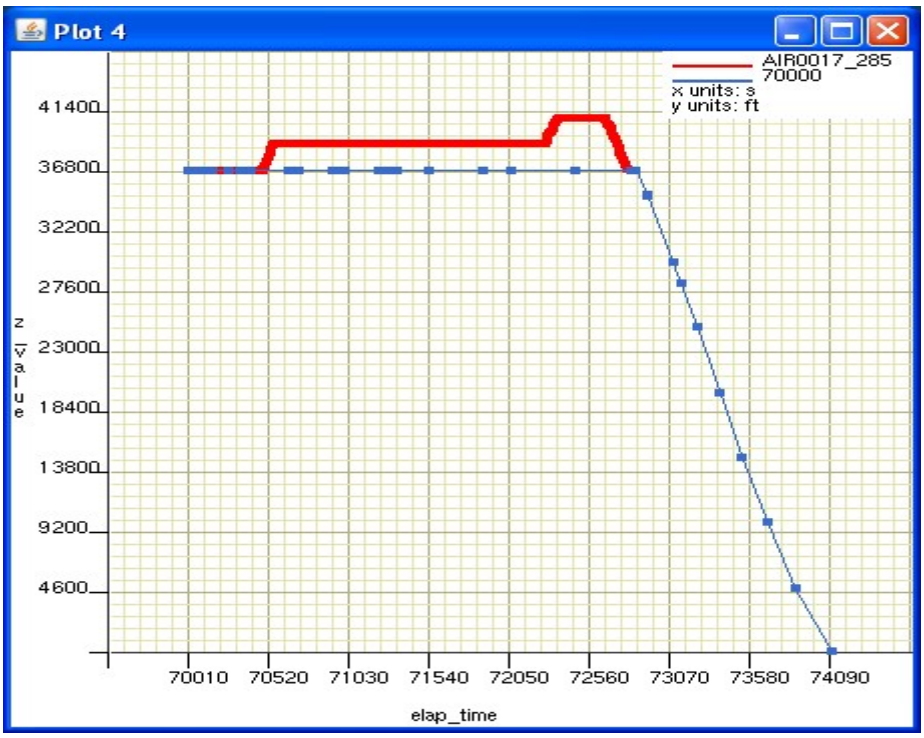

Figure 38: ERAM - Altitude profile and initial trajectory for Flight 1

Figure 39: URET - Altitude profile and a later trajectory for Flight 1

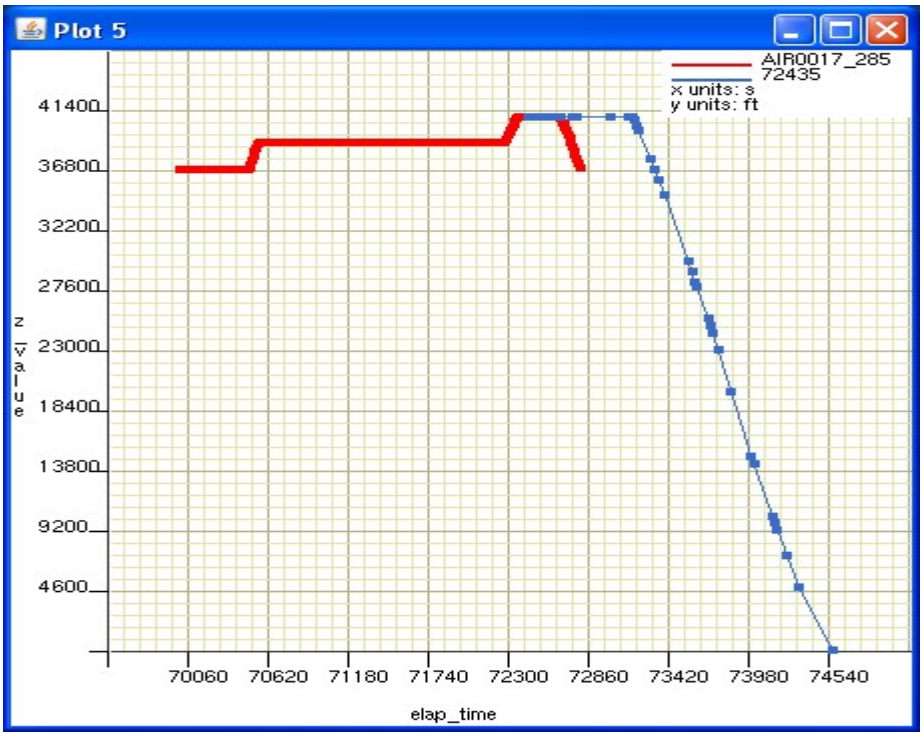

Figure 40: ERAM - Altitude profile and a later trajectory for Flight 1

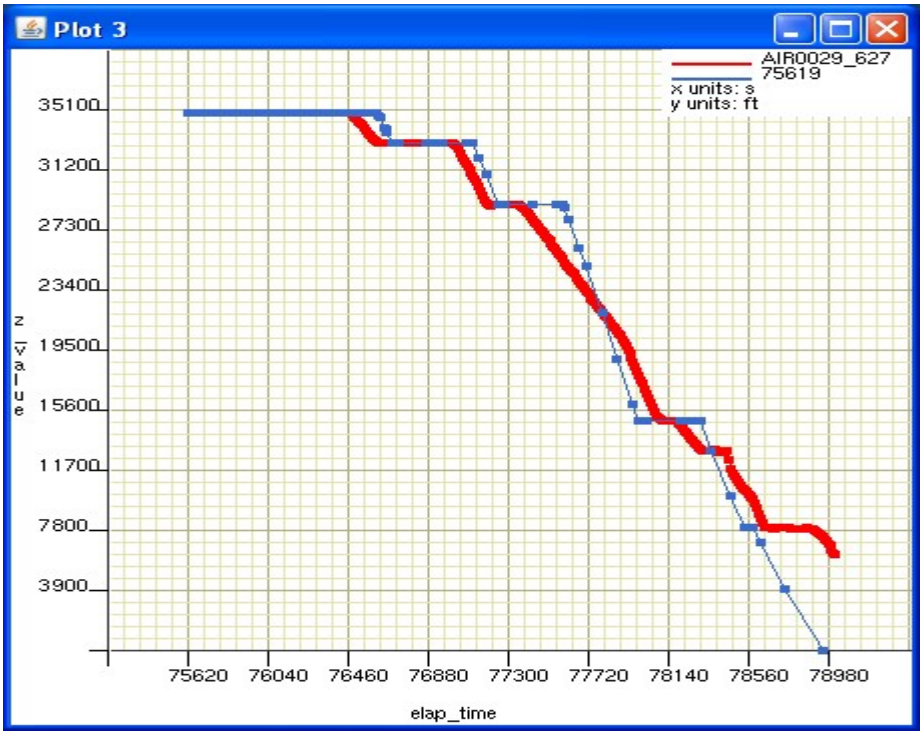

Figure 41: URET - Altitude profile and initial trajectory for Flight 2

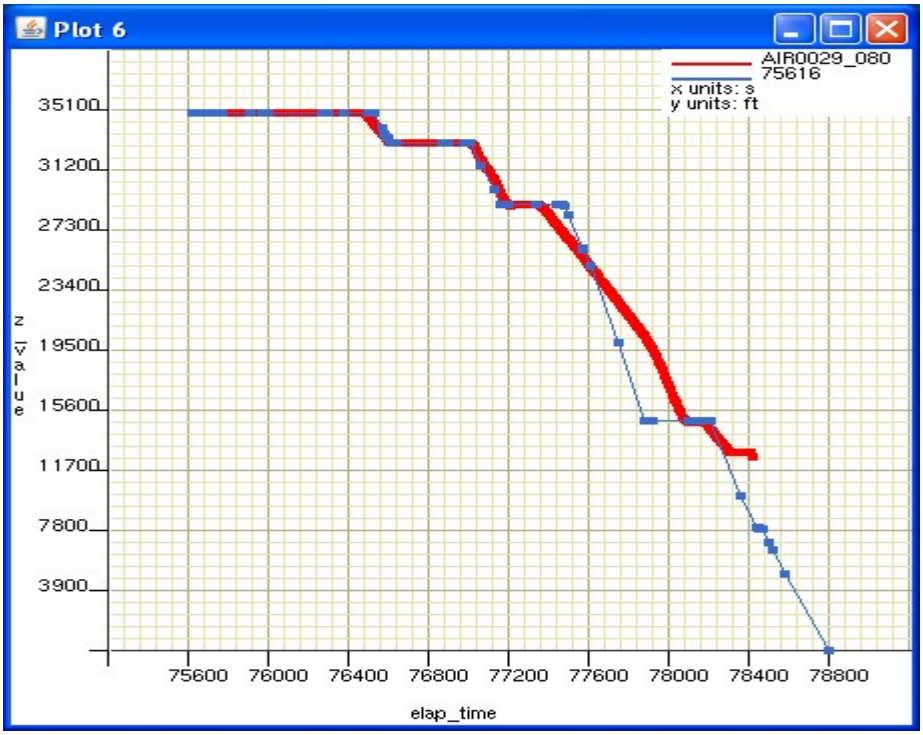

Figure 42: ERAM - Altitude profile and initial trajectory for Flight 2

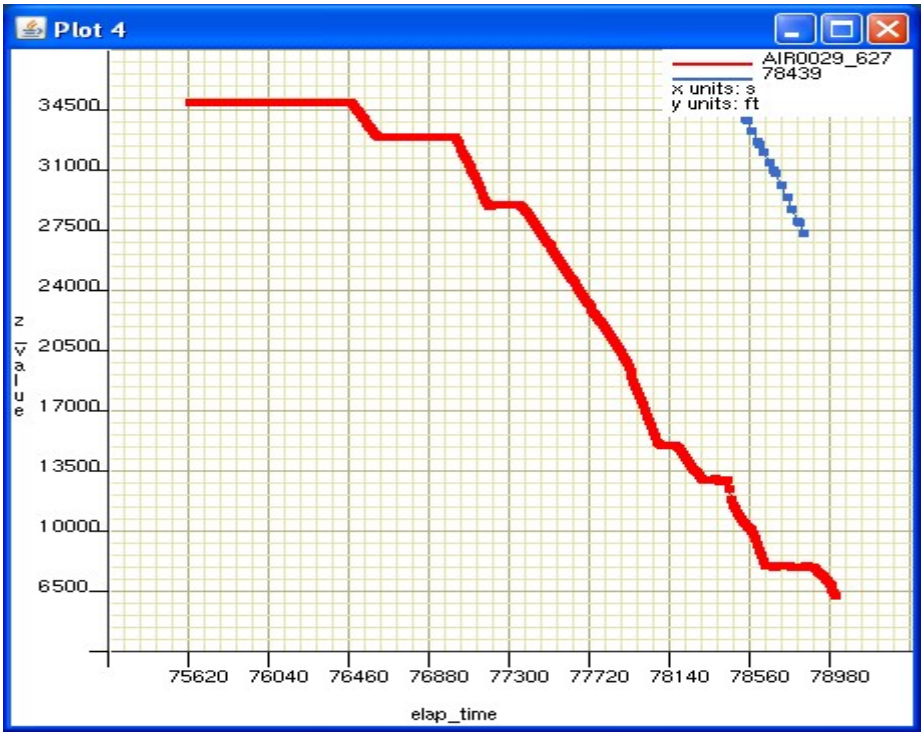

Figure 43: URET - Altitude profile and a later trajectory for Flight 2

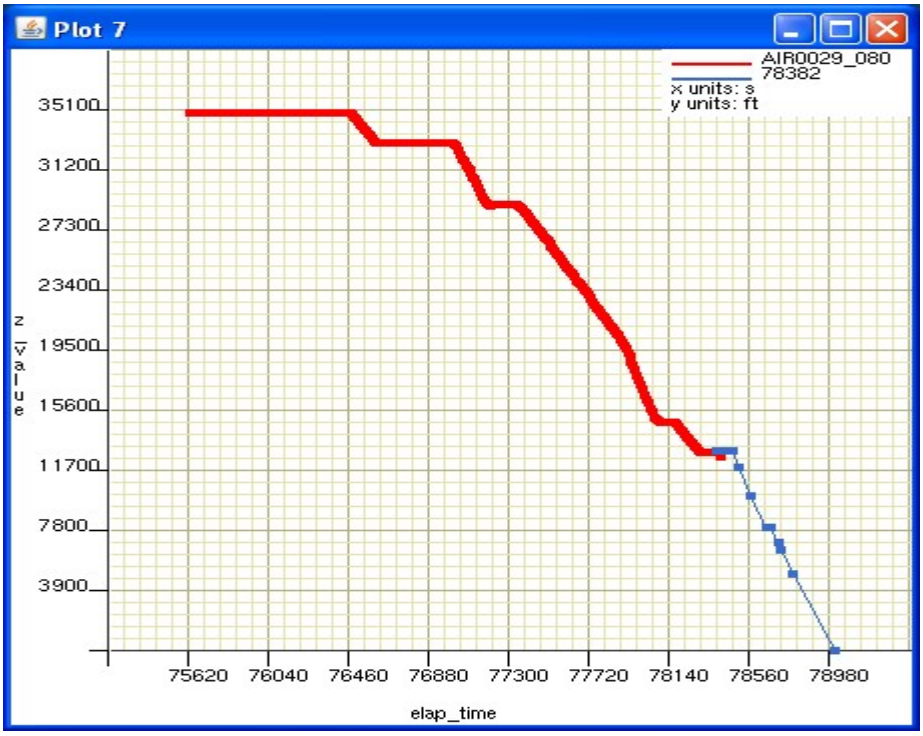

Figure 44: ERAM - Altitude profile and a later trajectory for Flight 2

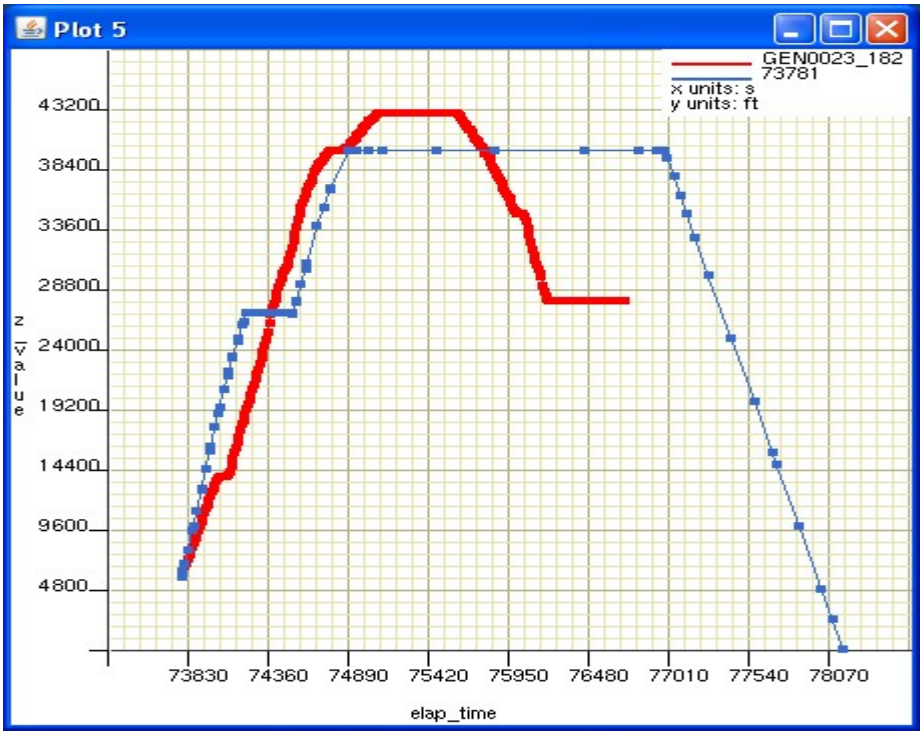

Figure 45: URET - Altitude profile and initial trajectory for Flight 3

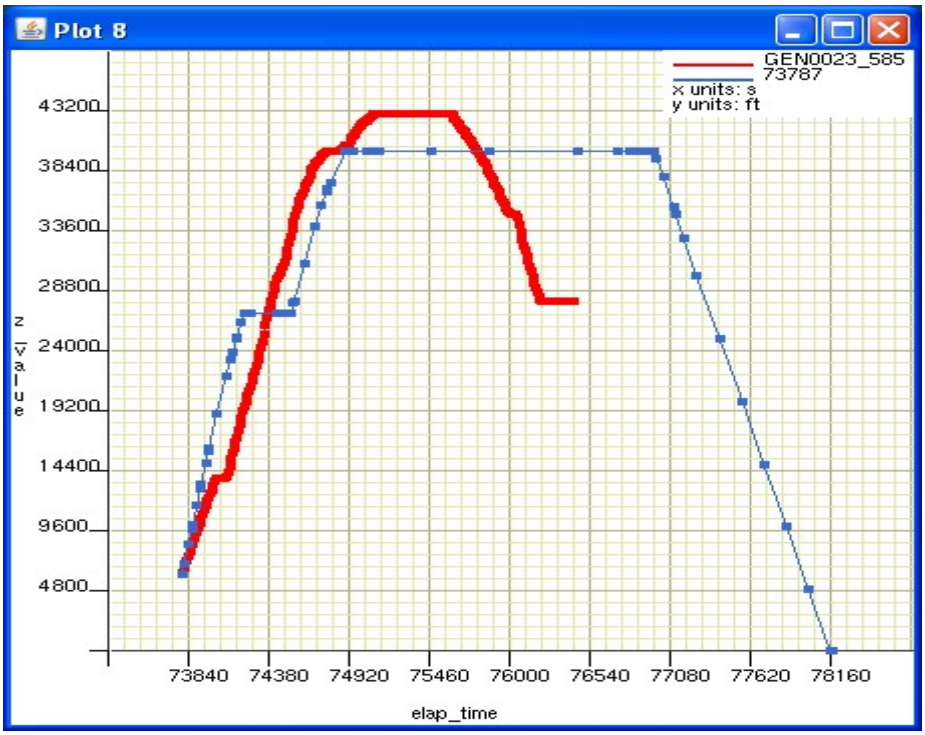

Figure 46: ERAM - Altitude profile and initial trajectory for Flight 3

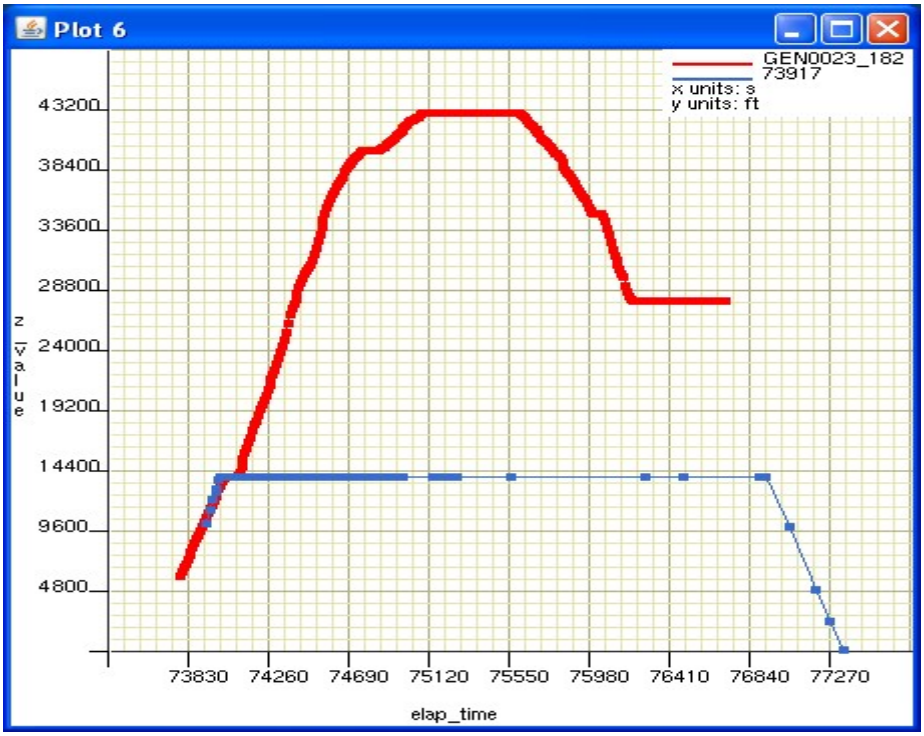

Figure 47: URET - Altitude profile and a later trajectory for Flight 3

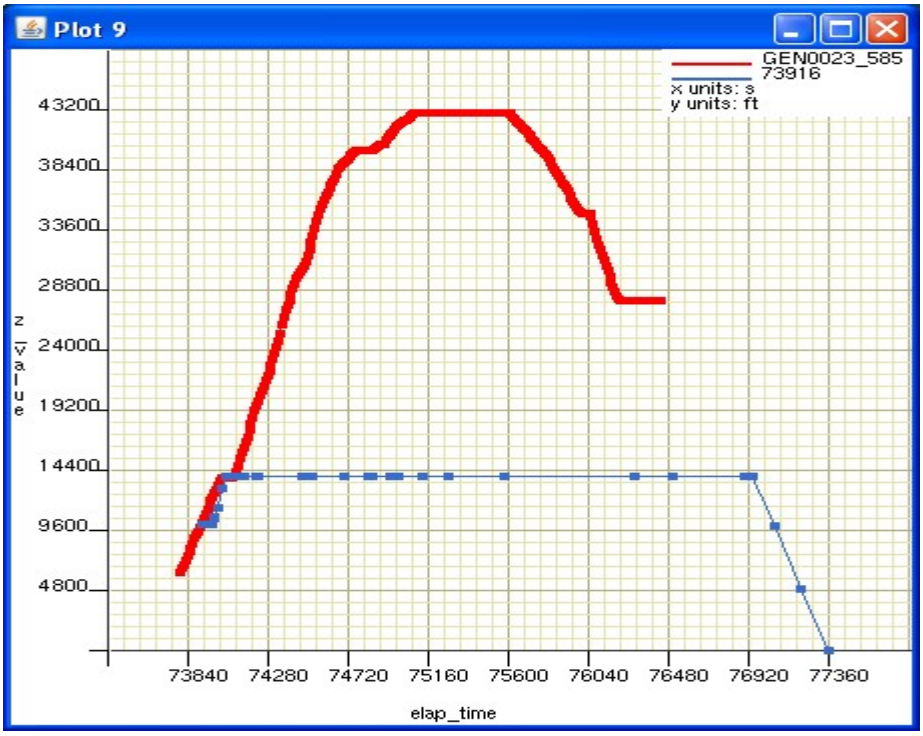

Figure 48: ERAM - Altitude profile and a later trajectory for Flight 3

Table 44: Altitude Errors for the ERAM Trajectories

FLIGHT	ACID	CID	TRAJECTORY BUILD TIME	TRAJECTORY NUMBER	COUNT	AVG	MAX	MIN	START TIME	END TIME
1	AIR0017	285	70000	15304	35	1076	2001	0	70040	71300
1	AIR0017	285	72435	15327	12	626	3267	0	72440	72770
2	AIR0029	080	75616	27015	32	21	419	0	75660	76890
2	AIR0029	080	78382	27037	1	100	100	100	78390	78390
3	GEN0023	585	73787	21889	35	4318	8361	266	73830	75090
3	GEN0023	585	73916	21890	29	11732	29000	45	73920	75120

The data listed here is from the NTS40C8W data set for the URET output data and the RFRNTSCB data set for the ERAM output data.

APPENDIX C – Statistical Test for Altitude and Conflict Data

Normal Distribution

ERAM is required to accurately predict aircraft altitudes. An altitude metric was defined in Section 3.1.2. Each window constructed predicts an aircraft altitude which is either accurate to within 500 feet for level flight (or to within 1500 feet for aircraft in transition) or is not. The value of the metric is the number of altitude windows in error divided by the number of windows sampled. This metric is a sampled random variable which has a binomial distribution. The binomial distribution can be approximated by the normal distribution where

μ = the mean of the normal distribution = np

σ = the standard deviation of the normal distribution = $\sqrt{npq}$

where

n = the number of samples,
p = the probability of an error
q = 1 – p

The normal distribution is converted into a standard normal distribution with a mean of zero and a standard deviation of 1 by defining the random variable Z[23]

where

$$Z = \frac{x - np + 0.5}{\sqrt{npq}} = f(x), \text{ a function of the error count x.}$$

The adjustment factor 0.5 is a continuity correction that compensates for the discreteness of the underlying binomial distribution and makes the normal distribution above more closely conform to the binomial distribution.

Similarly the missed alert rate and the false alert rate are also random binomial variables.

Application of the Normal Distribution

The customary way to apply a statistical test is to hypothesize that the data satisfies the accuracy requirement and then check the data to see if it is in agreement with the hypothesis. If the data does not support the hypothesis, that is the hypothesis is rejected, the test is failed. If the data does not reject the hypothesis, the test is passed.

The accuracy requirement has been established by sampling the URET performance data to obtain an estimate of the average accuracy value. The sample URET average is a random variable and as such has an associated confidence interval. The normal distribution approximation to the binomial distribution can be used to define the confidence interval. Then

[23] The Box (2005) reference, [1], pp. 52-55, gives the normal approximation with the Yates continuity correction of 0.5.

the sampled average of the system under test, ERAM, is checked to see if it is in this confidence interval. If the ERAM value is in the confidence interval the hypothesis is not rejected and test is passed. This is the method used by Lockheed Martin to test the altitude prediction accuracy error rate and the missed and false alert rates. A similar method, described in Appendix D, using the t distribution, is applied to the warning time data.

Since there is only a maximum error rate and no minimum error rate, a single tailed test is used. Only the upper bound on the confidence interval is relevant. Imposing a confidence interval on the URET derived specification data gives an upper bound to the specification which is higher than the actual measured value.

The upper bound can be calculated in two ways. An upper bound on the allowable error rate can be directly calculated and compared to the observed error rate in the test data or a Z statistic can be calculated from the test data and compared to the required Z statistic. The first method is used to check the altitude prediction data; the second method is used to check the missed and false alert data;

The upper bound (UB) on the requirement, the maximum allowable error rate, is, approximating the binomial distribution by the normal distribution

$$UB = Z_{1-\alpha}\sqrt{np_o(1-p_o)} + np_o - 0.5$$

Where

$Z_{1-\alpha}$ = standard normal random variable (single-tailed test)
n = the number of samples in the test
p_o = the URET error rate (average of the sampled data)
α = level of significance

The following example illustrates how this test method is applied to the ERAM altitude prediction data.

Example – Upper Bound on the URET Specification Data – Altitude Prediction Accuracy

The upper bound on the observed error rate of the altitude prediction accuracy test for aircraft in transition flight is calculated as follows. A 10% level of significance is chosen, that is 10% of the distribution is in the upper tail, and $\alpha = 0.1$.

n = the number of ERAM samples in the test = 80,638[24]
p_o = the URET measured error rate = 0.1431
α = level of significance = 0.1
$Z_{1-\alpha}$ = the standard normal random variable = 1.2816

$$UB = 1.2816\sqrt{(80638)(0.1431)(0.8569)} + (80638)(0.1431) + 0.5$$

$$= 11{,}666$$

[24] n has a large value making the normal distribution a good approximation to the binomial distribution. Both $np_o = 11539$ and $n(1-p_o) = 68971$ are large.

Since the observed error rate was 13,354 the test is failed.

In terms of probabilities, the upper bound of the allowable observed error rate is

UB = 11666 / 80638 = 0.14468

and the observed error rate for ERAM is

= 13354 / 80638 = 0.1656

Example – Z Statistic on the Missed Alert Rate

The Z statistic is calculated for the missed alert rate using the formula above for converting a binomial distribution into a normal distribution.

n = the number of ERAM samples in the test = 222
x = the number of ERAM samples in error – missed conflict predictions = 7
p_o = the URET missed alert rate = 7 out of 240 = 0.02083

$$Z = \frac{x + 0.5 - np}{\sqrt{npq}} = \frac{7 - (222)(0.02083) + 0.5}{\sqrt{(222)(0.02083)(1 - 0.02083)}} = 1.35145$$

The threshold value of the z statistic for a single tailed confidence interval of 95% is 1.64485. The observed value of the ERAM missed alert rate is contained within the 95% confidence interval and therefore the hypothesis is not rejected and the test is passed.

The P value (the cumulative probability in the upper tail) corresponding to a Z value of 1.35145 is 0.088349.

Example – Z Statistic on the False Alert Rate

The Z statistic is calculated for the false alert rate for aircraft to aircraft encounters having a minimum separation of less than 10 nautical miles. The formula above for converting a binomial distribution into a normal distribution is used.

n = the number of ERAM samples in the test = 3074
x = the number of ERAM samples in error – false alerts = 626
p_o = the URET missed alert rate = 593 out of 3270 = 0.18135

$$Z = \frac{x + 0.5 - np}{\sqrt{npq}} = \frac{626 - (3074)(0.18135) - 0.5}{\sqrt{(3074)(0.18135)(1 - 0.18135)}} = 3.23131$$

The threshold value of the z statistic for a single tailed confidence interval of 95% is 1.64485. The observed value of the ERAM missed alert rate is not contained within the 95% confidence interval and therefore the hypothesis is rejected and the test is failed.

The P value (the cumulative probability in the upper tail) corresponding to a Z value of 3.1852 is 0.00072327

APPENDIX D – Statistical Test for Warning time Data

A sample of warning times calculated by ERAM is obtained from the scenario test run and it is compared to a sample of warning times obtained from a URET scenario run. The means of the two samples are calculated and compared. The accuracy requirement is that the average warning time for ERAM (μ) is as large as or larger than the average warning time from URET (μ_r). It is hypothesized that the average warning time is the same for both ERAM and URET. The unpaired t test is used for the comparison. If the hypothesis is not rejected by the data, ERAM passes the test.

It is assumed that the warning times for ERAM and URET are both normally distributed and that their variances are approximately equal. The distribution of the difference between the sample means is a t distribution where the number of degrees of freedom is a function of the number of samples taken from the two systems – that is $n_1 + n_2 - 2$ where n_1 is the number of samples taken for URET and n_2 is the number of samples taken from ERAM. The value of the t statistic is calculated from the following equations. The probability of the occurrence of this value, assuming the two systems have equal warning times, when it is sufficiently high, states that the hypothesis is not rejected.

Calculation of the t statistic

1. Calculate the difference between the sample means $\Delta\mu$

   $\Delta\mu = \mu_1 - \mu_2$

   where μ_1 = average warning time of the URET sampled data
   μ_2 = average warning time of the ERAM sampled data

2. Calculate the pooled standard deviation s_p

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

   where n_1 = the number of URET warning time samples
   n_2 = the number of ERAM warning time samples
   s_1 = the standard deviation of the URET sample
   s_2 = the standard deviation of the ERAM sample

3. Calculate the standard error of the difference between the means s_d

$$s_d = s_p \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

4. Calculate the t statistic t

$$t = \frac{\Delta\mu}{s_d}$$

Example – Warning Times

The hypothesis that the averages of the warning times are the same for ERAM and URET for warning times greater than 10 minutes will now be checked using this t test.

URET data	$n_1 = 11$	ERAM data	$n_2 = 12$
	$\mu_1 = 1085.5$		$\mu_2 = 1000.2$
	s1 = 288.9		$s_2 = 391.46$

Pooled standard deviation

$$s_p = \text{SQRT}\ (\ ((11-1)*288.9**2 + (12-1)*391.46**2)\ /\ (11 + 12 - 2)\ ) = 346.43$$

Standard error

$$s_d = 346.43 * \text{SQRT}\ (\ (11+12)/(11*12)\ ) = 144.61$$

t statistic

$$t = (1085.5 - 1000.2)\ /\ 144.61 = 0.5899$$

The number of degrees of freedom of the t distribution is n1 + n2 – 2 = 21. The inverse t distribution, at the 5% level of significance, gives a value of the t statistic of

1.7207

Therefore, since 0.5892 < 1.7207, the hypothesis is not rejected and the warning time accuracy test is passed.

Appendix E – Overview of Conflict Prediction Performance Comparison

In this appendix we detail illustrations detailing the eight phases the entire conflict prediction performance comparison between CPAT tools and LM's OFPD and EOPD. Each phase of the comparison is detailed in a specific section. The results of the comparison between CPAT tools and OFPD are illustrated in Figure 49. The results of the comparison between CPAT tools and EOPD are illustrated in Figure 50.

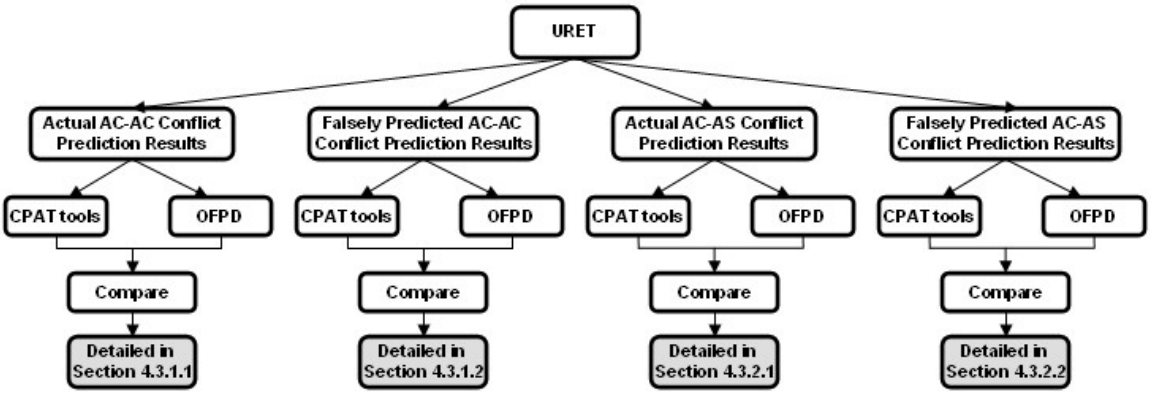

Figure 49: This flowchart illustrates how AC-AC and AC-AS conflict predictions results for actually occurring conflicts and falsely predicted conflicts were compared between CPAT tools and OFPD.

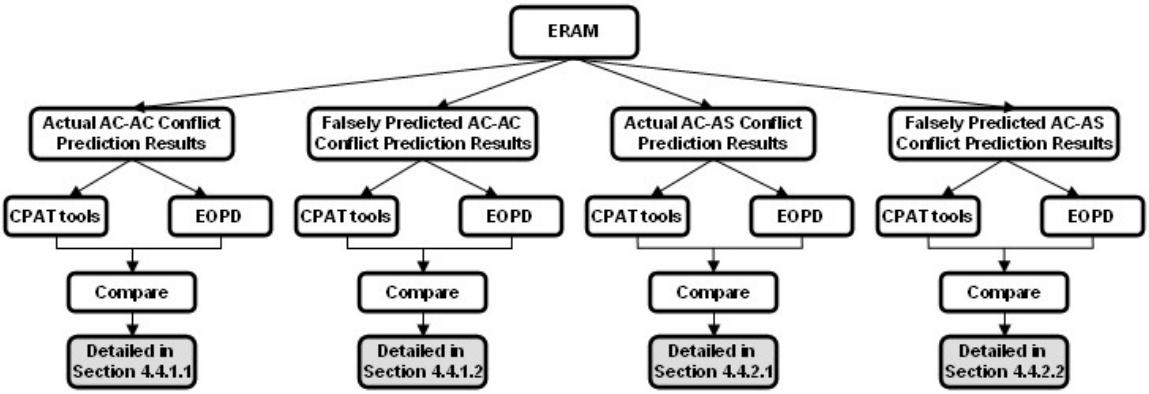

Figure 50: This flowchart illustrates how AC-AC and AC-AS conflict predictions results for actually occurring conflicts and falsely predicted conflicts were compared between CPAT tools and EOPD.

Appendix F – List of Problem Reports (PRs) in OFPD, EOPD, and CPAT tools.

In this appendix we provide the list of problem reports (PRs) within LM's OFPD and EOPD support tools, and the FAA CPAT tools. For each PR in Table 45 we list the tool in which the PR resides, the description of the PR, its impact level on the results, and its status.

Table 45: List and status of problem reports (PRs) in OFPD, EOPD, and CPAT tools.

Tool	PR #	Description	Impact	Status
OFPD	N/A	Test SAAs do not have 3nm and 500ft buffer as in EOPD.	Medium	Open
OFPD	N/A	OFPD labels AC-AS conflict as Conflict_Starts_At_Start_Of_Track when the beginning of the track starts in an APDIA and upon departing APDIA, this first track is the beginning of the conflict, but this point is not the actual start of track, hence it is not a popup conflict. OFPD considers it to be a Strategic Missed Conflict (LATE_MA) correctly, but labeling the start of conflict as the start of track is not exactly correct.	Low	Open
OFPD	N/A	OFPD continues conflict determination logic after center handoff times. See conflict COA1953_218 and BTA2147_82J starting at 72380 (20:06:20) from RFR URET data. Handoff is at 72390 (20:06:30) for flight COA1953_218.	High	Open
OFPD	N/A	OFPD considers track in APDIA reason to discard FA based on no track. Similar problem in EOPD.	High	Open
EOPD	PR_28728	EOPD wrongly discards conflicts when should be considered a miss; setting first track to ADH_AGE = 1	High	Open
ERAM	PR_28096	Trajectory build fails with ALGORITHMIC_FAILURE, causing conflicts to be missed deriving from PR_28728.	High	Open
EOPD	PR_30869	EOPD incorrectly discards AC-AC and AC-AS FAs based on clearance change.	High	Open
EOPD	PR_30191	When dealing with two conflicts; EOPD determines the first conflict to end upon entering the APDIA and it correctly determines the end of the second conflict. However, it leaves the start of the second conflict at the point of entering the	High	Open

		APDIA region instead of at its departure.		
EOPD	PR_30868	EOPD does not consider AC-AS conflicts to be popups correctly.	High	Open
EOPD	N/A	EOPD continues conflict determination logic after center handoff times. See conflict JBU376_001 and OPT729_155 starting at 76870 (21:21:10) in the RFR data. Handoff for flight JBU376_001 is 76790 (21:19:50).	High	Open
EOPD	N/A	EOPD considers track in APDIA reason to discard FA based on no track. Similar problem in OFPD.	High	Open
EOPD	N/A	Track for CHQ6289_174 begins at 65961 (18:19:21), and EOPD produces first PROTRK report at 65960 (18:19:20). Therefore, EOPD incorrectly creates track point prior to the real start of track.	High	Open
EOPD	N/A	EOPD erroneously drops an interim altitude, making the flight appear to be out of adherence, and therefore incorrectly discards FAs based on adherence.	High	Open
CPAT Tools	N/A	Bug in CPAT tools track reasonableness processor (FlightPathValidatorLite); altitudes spikes are not removed.	High	Closed
CPAT Tools	N/A	Discrepancy in logic compared to EOPD. EOPD allows for a conflict to be matched with a notification set even if the ACST is equal to the notification set end time, CPAT tools originally did not, but now allows this.	Medium	Closed
CPAT Tools	N/A	Bug in CPAT tools where for a FA to be discarded based on change in clearance at notification set end time, the FA also had to be retracted, which is incorrect. The check for retracting FA was removed in this situation.	High	Closed