AIAA Guidance, Navigation and Control Conference and Exhibit
20 - 23 August 2007, Hilton Head, South Carolina

AIAA 2007-6518

Performance Metrics for Tactical Aircraft to Aircraft Conflicts

Hollis F. Ryan*
General Dynamics Information Technology, William J. Hughes Technical Center, Atlantic City, NJ 08405

Mike M. Paglione†
Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City, NJ 08405

Shurong Liu‡
General Dynamics Information Technology, William J. Hughes Technical Center, Atlantic City, NJ 08405

An air traffic control system's main function is to separate aircraft. The computer supporting the system assists the air traffic controllers by generating a conflict alert whenever it predicts that two aircraft are about to get too close to each other. The performance of the conflict alert function is a key element to the overall functioning of the air traffic control system. A set of metrics has been designed to measure the conflict alerting performance of an aircraft traffic control system. The key factors are the missed alert rate and the false alert rate. However there are several secondary factors that are essential to measuring the performance, especially in a simulation environment. This paper describes a set of metrics that have been developed to evaluate the performance of the FAA's en route aircraft traffic control system. They have been applied to the existing system, the Host Computer System, and will be used to establish metrics for the new system now under development, the En Route Automation Modernization system. The metrics are calculated by post processing recorded data that has been produced by running a real time simulation of the air traffic system without controllers, using as input field recorded aircraft data that has been time shifted to induce aircraft-to-aircraft conflicts.

I. Introduction

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[1]. The accuracy of ERAM in generating tactical conflict alerts for the controllers will be measured as part of the FAA test program for ERAM. This paper describes the metrics developed to measure the accuracy of both the current system (HCS) and the new system (ERAM) in predicting tactical aircraft-to-aircraft conflicts. A conflict is the loss of the minimum required aircraft separation. In en route airspace the minimum allowed separation is either five nautical miles horizontally or 1000 feet vertically. A conflict alert is given to the responsible controller when a conflict is predicted to occur within the next 135 seconds. It is required that the controller be given the alert at least 75 seconds before the start of the predicted conflict. The metrics have been applied to the current HCS and will be applied in the future to the new ERAM system when it becomes available. The metrics are similar to those previously used in the FAA acceptance and Lockheed Martin regression testing of the User Request Evaluation (URET) system[2,3].

A. Background

A complete evaluation of a tactical conflict probe has two complementary aspects: qualitative and quantitative. A qualitative evaluation generally involves real-time testing of conflict probe features and user interface through

* Team Member, Conflict Probe Assessment Team, Simulation and Analysis Group.
† Lead, Conflict Probe Assessment Team, Simulation and Analysis Group, AIAA Member.
‡ Team Member, Conflict Probe Assessment Team, Simulation and Analysis Group.

Copyright © 2007 by the American Institute of Aeronautics and Astronautics, Inc.
The U.S. Government has a royalty-free license to exercise all rights under the copyright claimed herein for Governmental purposes.
All other rights are reserved by the copyright owner.

human-in-the-loop simulations and field tests; for example Refs. 4, 5, and 6 describe real-time testing of various conflict probe capabilities. A quantitative evaluation generally involves non-real-time testing directed at the conflict detection "engine" that underlies the features and user interface of a conflict probe. A comprehensive methodology for quantitative evaluation of a conflict probe is presented in Ref. 7; an application of this evaluation methodology has been reported in Ref. 8. Generic metrics for quantitative evaluation are available in Ref. 9. Conflict probe performance metrics are presented in Ref. 4, using a hybrid approach involving data collection and transformation models applied to a recorded air traffic scenario.

There are many metrics that can be applied in evaluating the performance of conflict predictions as referenced above. The metrics presented in this paper are adapted mainly from Refs. 2, 3, and 9. The categories include (1) the measurements of estimating the error events of the missed and false alerts, (2) the measurements quantifying the timeliness of the correct predictions, valid alerts, (3) the metrics related to the prediction stability of these predictions, and finally (4) prediction sensitivity measurements that quantify the overall performance of the accuracy predictions such as sharpness as defined in Ref. 9. This paper will present definitions of the missed and false alert rates and timeliness metrics.

B. Simulation

The quantitative performance of the ATC system in predicting tactical conflicts has been measured by applying a set of metrics to the results of a real time simulation. The Integration and Interoperability Facility (IIF), an en route simulation facility at the FAA's W. J. Hughes Technical Center, has been used to process a sample or scenario of aircraft flight data recorded in the field during a typical daily operation. In the simulation, the aircraft and airspace are simulated by using field recorded data. An actual HCS, running in real time, processed the data. For this study no controllers were necessary because the conflict alerts were generated automatically by the ATC system. In normal operations very few actual conflicts occur – not enough for testing. It was necessary to time shift the data to create conflicts for the system to predict[10,11]. The conflicts in the scenario are pseudo-conflicts as they did not actually occur except in the simulation. Offline processing identifies these pseudo-conflicts after the time shift preprocessing of the recorded field data[12,13].

Tactical alerts are given by the ATC system to alert controllers that a conflict is imminent. The actual time and place of the conflict is of less importance. The performance metrics described here for the tactical alerts do not include measuring the accuracy of the location of these predictions, only the timeliness.

The alerts provided by the real time simulation were recorded and then compared to the conflicts that occurred in the scenario. These conflicts were determined, as mentioned above, by offline processing of the aircraft radar track data input to the simulation. A comparison was performed by applying the set of metrics described in this paper to the data.

Approximately four hours of data was recorded at the Washington Air Route Traffic Control Center (ARTCC) or ZDC. The key data elements required to run the simulation are the flight plans and the radar track reports. The adaptation of the ATC system to the airspace being simulated (ZDC) is done offline. The tactical conflict alert function does not use flight plan route data, but this data is necessary to run the simulation. The tactical conflict alert processing does use the altitude clearances. The alerts generated during the real-time simulation run are recorded without controllers. They are compared to the conflicts in the time shifted scenario by a set of post analysis computer programs. This comparison will be explained in detail later in the paper.

C. Alerts

The basic requirement of an alert function is fairly simple. It is required that an alert be posted when there is going to be a conflict and an alert not posted when there is not going to be a conflict. However, the nature of the simulation, the characteristics of the data, and the test philosophy complicate the requirement and the definition of the metrics. There are a number of special cases created by the artifacts of the simulation. Two major artifacts are the specific start time and the specific end time of the simulation. If, in the simulation, two aircraft start in conflict, the HCS cannot be expected to predict the conflict at least 75 seconds in advance. These are artifacts that do not normally occur in the real operational system running 24 hours a day and seven days a week. The special cases result in missed conflicts being excused and false alerts being discarded. This paper describes how these factors are considered in defining the metrics (and submetrics). These details of what are basically counting rules are important because they directly influence the definitions and subsequent accuracy results of the tactical conflict alert performance being measured through simulation.

II. Basic Processing

The basic processing analyzes the input data to the HCS and the output data from the HCS to see if the system correctly provides tactical conflict alerts. The ATC posts alerts, updates them, and then deletes them. An alert specifies the call signs or aircraft identifications (ACIDs) and computer identification codes (CIDs) for a pair of aircraft and a time and place of the start of a predicted conflict. The first alert posted for a predicted conflict and the last alert posted for the same conflict are combined to form a Notification Set (NS) for a specific conflict that may or may not exist. The set of all Notification Sets for the scenario defines what the ATC system does to predict conflicts. Offline processing of the radar track data[12,13] finds the conflicts in the scenario that the ATC should predict. An NS successfully predicts a conflict if it is posted between 75 and 135 seconds before the start of the conflict and remains posted when the conflict actually begins. The predicted time and place of the start of the conflict is not considered in this study. Such a match between an NS and a conflict is called a matched NS or a valid alert (VA). If there is no conflict to match, an unmatched NS has been found. This case is a false alert (FA) or possibly may be discarded due to special circumstances. If a conflict is found that has no matching NS, an unmatched conflict has been found. This is a missed alert (MA). The following table – a confusion matrix - illustrates these three cases and the remaining fourth case (No Call, NC) when there is neither an NS nor a conflict for the given non-conflict encounter event.

Table 1: Basic Cases of Conflict/Notification Set Events

	Notification Set (NS) Posted	**Notification Set (NS) Not Posted**
Conflict Occurs	**CASE 1**: there is a conflict and there is a NS posted against it – valid alert (**VA**)	**CASE 2**: there is a conflict and no NS is posted for it – a missed alert (**MA**)
Conflict Does Not Occur	**CASE 3**: there is no conflict but there is a NS posted – a false or nuisance alert (**FA**)	**CASE 4**: There is no conflict and there is no NS - correct no-call (**NC**)

In order to quantify the false alerts, the offline processing, in addition to finding the conflicts in the scenario, also finds the encounters. An encounter is the close approach of two aircraft that may or may not be close enough to qualify as a conflict. In this study an encounter was defined to be an aircraft-to-aircraft separation of simultaneously less than 40 nautical miles and less than 8000 feet. These numbers are identical multiples of the 5 nautical mile and 1000 foot separation defining a conflict. For this study, there were no alerts generated for separations greater than these values.

The ground truth conflicts and encounters are determined using the radar track data after interpolating then to 10 second increments in post processing. A conflict must have a minimum of two successive data samples in conflict to be considered a conflict. Single conflicting points are ignored. Two adjacent conflicts are merged into one if the time gap separating them is 40 seconds or less.

III. Post Processing of Recorded Simulation Data - Overview

An overview of the processing required to extract the metrics from the data generated by a real time simulation run is given in this section and illustrated in Figure 1 and Figure 2. This overview describes the identification of the four cases given in Table 1 above. The following sections describe how these four cases are further subdivided into a total of 16 sub-cases.

The process begins with the simulation, which is run in real-time. Next, analysis of the aircraft surveillance position reports and output tactical alert predictions are post processed. The position reports composed of latitude, longitude, altitude, and time fields of the aircraft are processed to determine all conflict events, producing a conflict list. At a minimum, a tactical conflict alert message includes two call signs, two computer ID's, and a posting time. These alerts are assembled into a Notification Set list. Next, these two lists are matched by the aircraft identification call signs as illustrated in Figure 1. The process of matching the NSs to the conflicts produces matched NSs or unmatched conflicts, resulting in three outcomes: valid alerts, missed alerts, and discards. Valid alerts occur when the conflict has a matching Notification Set and is presented in a timely manner. This happens when the Notification Set was started (first presented) with a warning time greater than or equal to a minimum warning time requirement (MWTR) for the given aircraft-to-aircraft conflict.

If no matching NS is found and the conflict is unmatched, a search is performed for any NS started at or after the actual conflict start time (ACST). If none are present, the event is labeled as a missed alert event. If a NS is

found and the both the conflict and the NS starts near the start of track or near a gap in track data, the event is discarded (recorded as a discard). A conflict that starts near the start of track or near a gap in surveillance track reports of one of the aircraft is referred to as a pop-up.

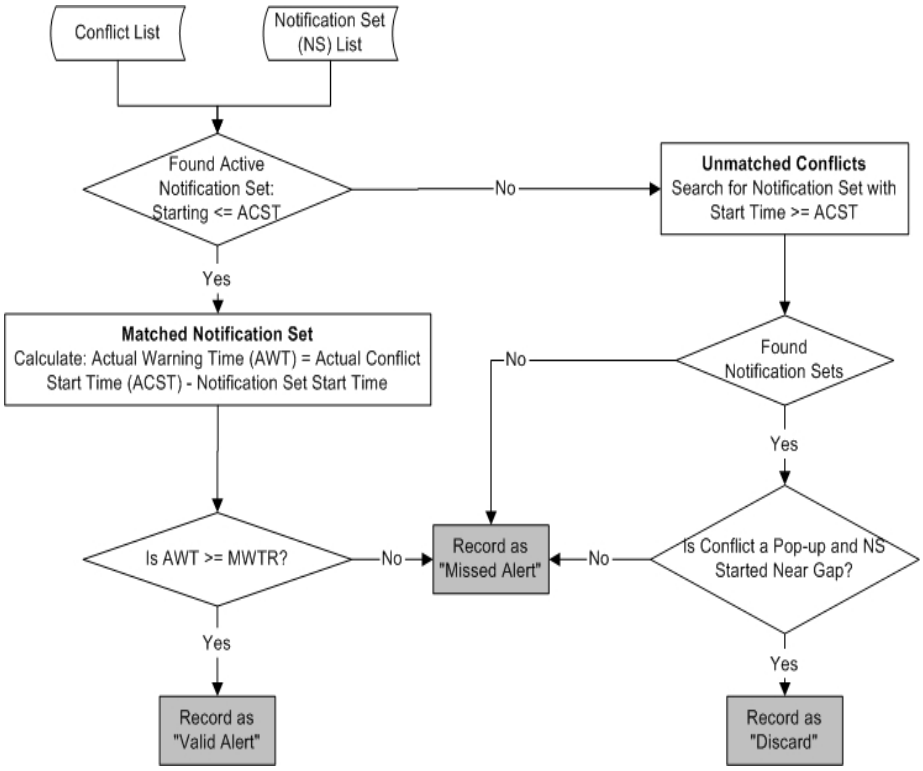

Figure 1. Processing of Valid and Missed Alert Events

For the remaining NSs that were not matched, these events are potentially false alerts (a.k.a. nuisance alerts). As illustrated in Figure 2, four basic checks are performed to verify that the NS is either a false alert or should be discarded. First, the track reports must exist for both aircraft at the NS start or the event is labeled a discard. Second, if an actual conflict exists and the NS's start time is within the duration of this conflict, the NS is discarded. Third, at the NS start time the flights are linearly extrapolated forward in time using past track position reports to predict whether a conflict will occur. If no conflict is predicted, the event is immediately labeled a false alert. Finally, if a conflict is predicted, more linear extrapolations are performed within the NS time interval (between it starts and ends). This is iterated until either no conflict is predicted or the end of track occurs for one of the aircraft. If this ending time is near the end of the NS, the NS event is labeled discard and if not it is labeled as a false alert.

IV. Special Cases of Valid and Missed Alerts

There are several special cases of alerts and NSs. False alerts in particular have a number of special circumstances. The basic cases 1, 2, and 3 in Table 1 are further subdivided into additional sub-cases. Each of these special cases has its own set of processing rules and labeled by a "Reason Code", which denotes the particular rule applied . An occurrence of one of these cases is called an "event". The post simulation run analysis counts all the different events that result from the conflicts in the scenario and the NSs generated by the ATC system.

A. Edge Effects

The scenario is bounded in both time and space. It has a fixed starting time and a fixed ending time. In this study only the alerts and conflicts above Flight Level 180 (FL 180) are of interest. Individual flights may suddenly appear at the start of the scenario, when they are handed off to ZDC from a neighboring Center, or when they climb through FL 180. Similarly they may suddenly disappear when the scenario ends, when they are handed off to a neighboring

Center, or when they descend through FL 180. In addition, the radar track data that the HCS receives in the field and in the simulation occasionally contains gaps. The radar data for an aircraft flight may have dropped out for a period of time. This missing data causes aircraft to disappear and re-appear at random times. These characteristics of the scenario and data dictate that conflicts and NSs that occur below FL 180, after the end of the scenario, or in the middle of a gap in track data be discarded.

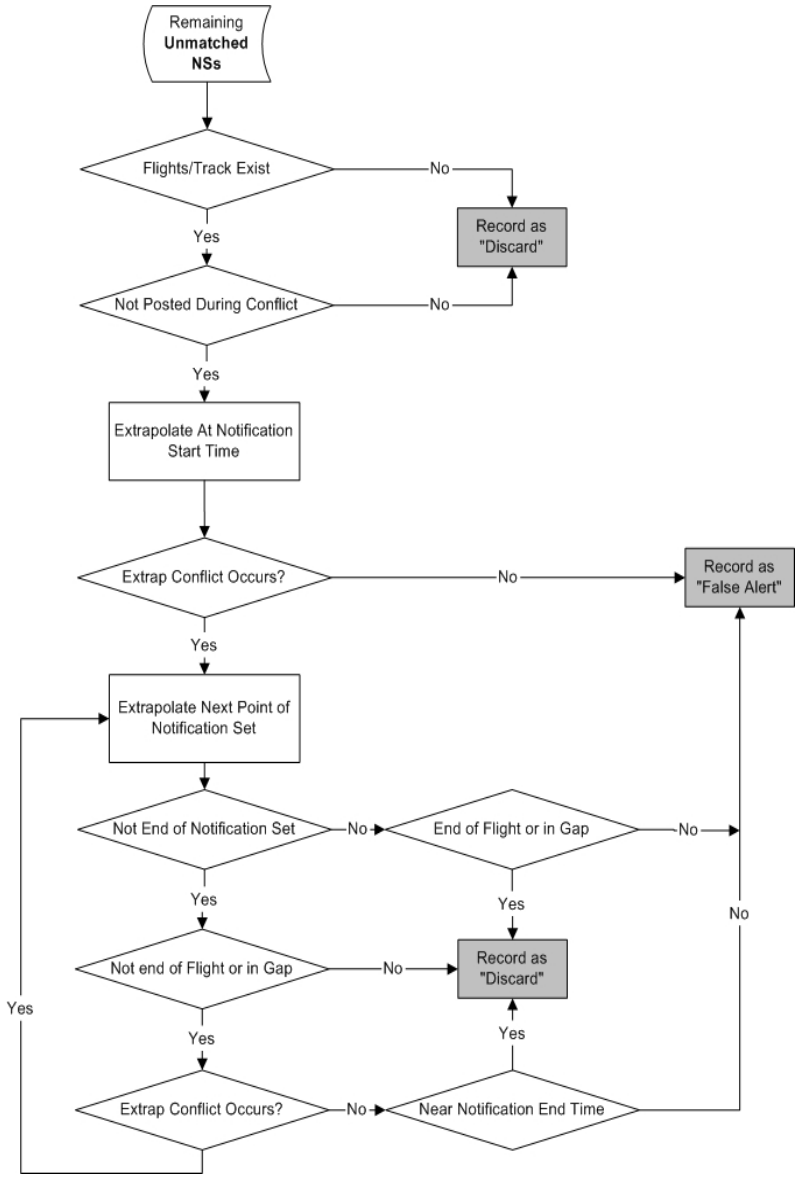

Figure 2. Processing of False Alerts and Discards

The tactical conflict alert is required to provide the controller with at least 75 seconds of warning time. If one or both aircraft suddenly appear and are in conflict or are in conflict shortly after they appear, the ATC system is unable to provide the 75 second warning time. The 75 second requirement does not apply in such cases and is relaxed to a smaller value. The ATC system is given 40 seconds to see an existing conflict after the appearance of track data for both aircraft. The 75 seconds is also relaxed if a maneuver is detected – described later. Another edge effect is that an alert may be posted shortly after the start of the conflict. This error is forgiven and the alert and conflict discarded if the conflict starts immediately after the (re-)appearance of track data and the alert occurs very soon (within 40 seconds) after the start of the conflict.

B. Maneuver Detection – Relaxation of Warning Time

In the test scenario with the induced conflicts an aircraft may maneuver in such a way to create a conflict and the conflict may start less than 75 seconds after the maneuver that created it. In this case the ATC system is not expected to be able to predict the conflict with a 75 second warning time. The required warning time is reduced. The following method is used to detect the maneuver and determine a new, smaller value for the required warning time.

The position, ground speed, heading, and rate of climb or descent are determined for each aircraft in the conflict. The future positions are estimated using a straight line extrapolation of the flight paths of the two aircraft to see if a conflict is predicted. A search is performed to see how many seconds before the start of the conflict a linear prediction would have predicted the conflict. This calculation is done starting at the time of the start of the conflict. If a conflict is predicted by the linear extrapolation at this point in time, the extrapolation is repeated, starting 10 seconds earlier. If a conflict is predicted by linear extrapolation from this point, the calculation is repeated again, starting another 10 seconds earlier. This search procedure is repeated until no conflict is predicted or until the starting point for the extrapolation is more than 75 seconds before the starting point of the actual conflict. If a point in time is found where a conflict is no longer seen by the extrapolation, the time difference between the following point where a conflict was seen and the starting time of the conflict becomes the new warning time requirement for the conflict. If this procedure leads to a relaxation of the warning time requirement of 75 seconds, it is assumed that an unpredictable maneuver caused the conflict.

This calculation is done for every conflict in the scenario. The values obtained are compared with any relaxed values found after the (re-)appearance of track data. The minimum value obtained is selected as the adjusted minimum warning time required for each conflict.

C. Reason Codes for Valid and Missed Alerts

The processing rules previously described are used to subdivide *Valid Alerts* and *Missed Alerts* cases into finer sub-cases. A regular *Valid Alert* is labeled *Valid Alert – Standard*. An alert that is valid except that its warning time is not greater than 75 seconds, and the warning time has not been relaxed to a smaller value is labeled *Missed Alert – Late*. An alert that is valid except that its warning time is less than 75 seconds, and the required warning time has been relaxed to a value less than the value obtained, is labeled *Valid Alert – Late.* If the otherwise valid alert's warning time is still less than the relaxed value, it is still labeled *Missed Alert – Late*.

A conflict that has no matching NS, and does not follow a gap in track data is labeled as *Missed Alert – Standard A*. If the conflict follows a gap in the track data, and there is no NS for the aircraft pair the conflict is labeled *Missed Alert Standard – B*. If the conflict follows a gap in the track data, and there is a NS that closely follows in time the start of the conflict, the conflict and the NS are discarded and the event is labeled *Missed Alert – Discard*. If the conflict follows a gap and there is an NS, but it is does not closely follow the start of the conflict, the event is still labeled *Missed Alert Standard – B*.

V. Special Cases of False Alerts

The initial processing divides the resulting events into three cases – (1) Matched Notification Sets, (2) Unmatched conflicts, and (3) Unmatched Notification Sets. The disposition of the first two cases has just been described above in the processing of missed and valid alerts. The third case, the subject of this section, involves the processing of the unmatched NSs into either false alert or discard events. If there is no track data available for either one of the aircraft at the time of the posting of the NS, the NS is discarded. If the NS is posted during a conflict, it is discarded. This error is counted elsewhere as well, usually as a missed alert event.

An unmatched NS can occur even when there is a conflict between the two aircraft. If the NS is posted before the start of the conflict and then withdrawn before the start of the conflict, it will not be matched to the conflict. Similarly if the NS is posted some time after the start of the conflict it will not be matched. A valid alert may be posted before the start of a conflict and then be deleted because the controller maneuvered the aircraft out of harm's way. This result is not an error and should not be counted as such. Linear extrapolation is used to detect such a maneuver.

A. Maneuver Detection – Unmatched Notification Set Discard

The same linear extrapolation of the aircraft flight paths is used for classifying the potential false alerts as is used to relax the minimum warning time requirement. However, for unmatched NS events, the extrapolation is applied to the aircraft flight paths at the posting time of the alert. If a conflict is predicted, the posting can be assumed correct. The starting point for the extrapolation is then advanced to the next track point and the extrapolation is repeated. If the conflict is still predicted by the linear extrapolation, the process is repeated, advancing one point and

recalculating. This process is continued until the conflict is no longer predicted or the end of the NS is reached or there is no track data. If the alert is withdrawn at about the same time that the extrapolation stops predicting a conflict, the NS is valid and is discarded; that is it is not counted as a false alert. The assumption is that the NS was valid and was withdrawn because the aircraft maneuvered to avoid the conflict. If the extrapolation runs out of track data, a determination of what happened cannot be made and the NS is discarded. If the withdrawal of the alert does not coincide with a loss of an extrapolated conflict, the event is labeled as a false alert.

The track data is sampled every 10 seconds, synchronized to the hour, and the time of posting of an alert is sampled to the nearest second. Therefore the extrapolation processing is started at the track points immediately proceeding the posting time of the alert. The reason codes for the event have been chosen to specify where in the processing the extrapolation was terminated.

B. Reason Codes for False Alerts

If there is no track data at the time of the posting of the alert the event is labeled *False Alert – No Track – Discard 1* (no track data at all) or *False Alert – No Track – Discard 2* (no track data specifically at the posting time). An NS posted during an actual conflict is labeled *False Alert – Actual Conflict Start Time – Discard*. If no conflict is seen by the extrapolation at the first test point the event is labeled *False Alert – Standard 1*. If a conflict is seen at the first point but not at the second and the alert withdrawal is not near in time to the time of the second point, the event is labeled *False Alert – Standard 2A*. If the withdrawal is close in time to the second point, the event is labeled *False Alert – Event – Discard A*.

If the conflict is no longer seen at a later point in the tracks, the event is labeled as *False Alert – Standard 2B* if the withdrawal is not near the time the conflict is no longer predicted by the extrapolation, and is labeled *False Alert – Event – Discard B* if it is close. If the extrapolation sees the conflict to the end of the NS, and there were no gaps in the track data, the event is labeled *False Alert – Standard 3*. If there was a gap in the track data, the event is labeled *False Alert – No Track – Discard 3*.

The detailed reason codes just described are mapped in the following Table 2 to the four basic alert types: valid, missed, false, and discard.

Table 2: Detailed Reason Codes for Conflict/Notification Set Events

INDEX	ALERT TYPE	REASON CODE (ABBREVIATED)
1	VALID	VA_STD
2	VALID	VA_LATE
3	MISSED	MA_STD_A
4	MISSED	MA_STD_B
5	MISSED	MA_LATE
6	FALSE	FA_STD1
7	FALSE	FA_STD2_A
8	FALSE	FA_STD2_B
9	FALSE	FA_STD3
10	DISCARD	MA_DISCARD
11	DISCARD	FA_NOTRK_DISCARD1
12	DISCARD	FA_NOTRK_DISCARD2
13	DISCARD	FA_NOTRK_DISCARD3
14	DISCARD	FA_ACST_DISCARD
15	DISCARD	FA_EVENT_DISCARD_A
16	DISCARD	FA_EVENT_DISCARD_B

VI. Example of Processing Results

A pair of flights has been chosen to illustrate the results of the processing described above. The drawing in Figure 3 is a time line plot that shows the aircraft tracks, the conflicts between the two aircraft, and the NSs or alerts generated by the ATC system, the HCS in this example. There are no gaps in the track data in this example. There are four conflicts, labeled C1, C2, C3, and C4. Their minimum warning times have been adjusted from 75 seconds to 70, 40, 10, and 30 by the linear extrapolation maneuver detection processing. There are six NSs. Two NSs are posted after the end of the track data in the scenario and are discarded as *FA_NOTRK_DISCARD2* and are not

shown on the figure. The other four NSs are labeled NS1, NS2, NS3, and NS4. They are associated with the four conflicts. The placement of the conflicts and NSs on the drawing gives the time sequence of the conflicts between the aircraft and the alerts posted by the ATC system. The post processing analysis of the data has labeled seven events. NS1 and NS2 are alerts for C1 and C2 respectively but are late. Their warning times are 55 seconds and 17 seconds. These times are respectively less than the values of 70 and 40 seconds required for these two conflicts and their reason codes are *MA_LATE*. NS3 is an alert for C3 but occurs 15 seconds after the start of C3 and its reason code is *FA_ACST_DISCARD*, a discard. The conflict C3 is missed and is labeled a MA_STD_A. NS4 is an alert for conflict C4 with a warning time of 30 seconds that is greater than or equal to the adjusted minimum warning time requirement of 30 seconds and therefore it is labeled a *VA_LATE*.

VII. Conflict Prediction Error Rates

The counts of the various events listed above in Table 2 are obtained by recording air traffic in the field using the HCS recording facilities, time shifting the individual flights to produce aircraft-to-aircraft conflicts, and running the scenario thus constructed through the WJHTC en route simulation facility, the IIF. The post processing of the conflicts in the scenario data and the tactical conflict alerts generated by the ATC system provide the counts for all of the different cases. These numbers are turned into error rates by the following calculations.

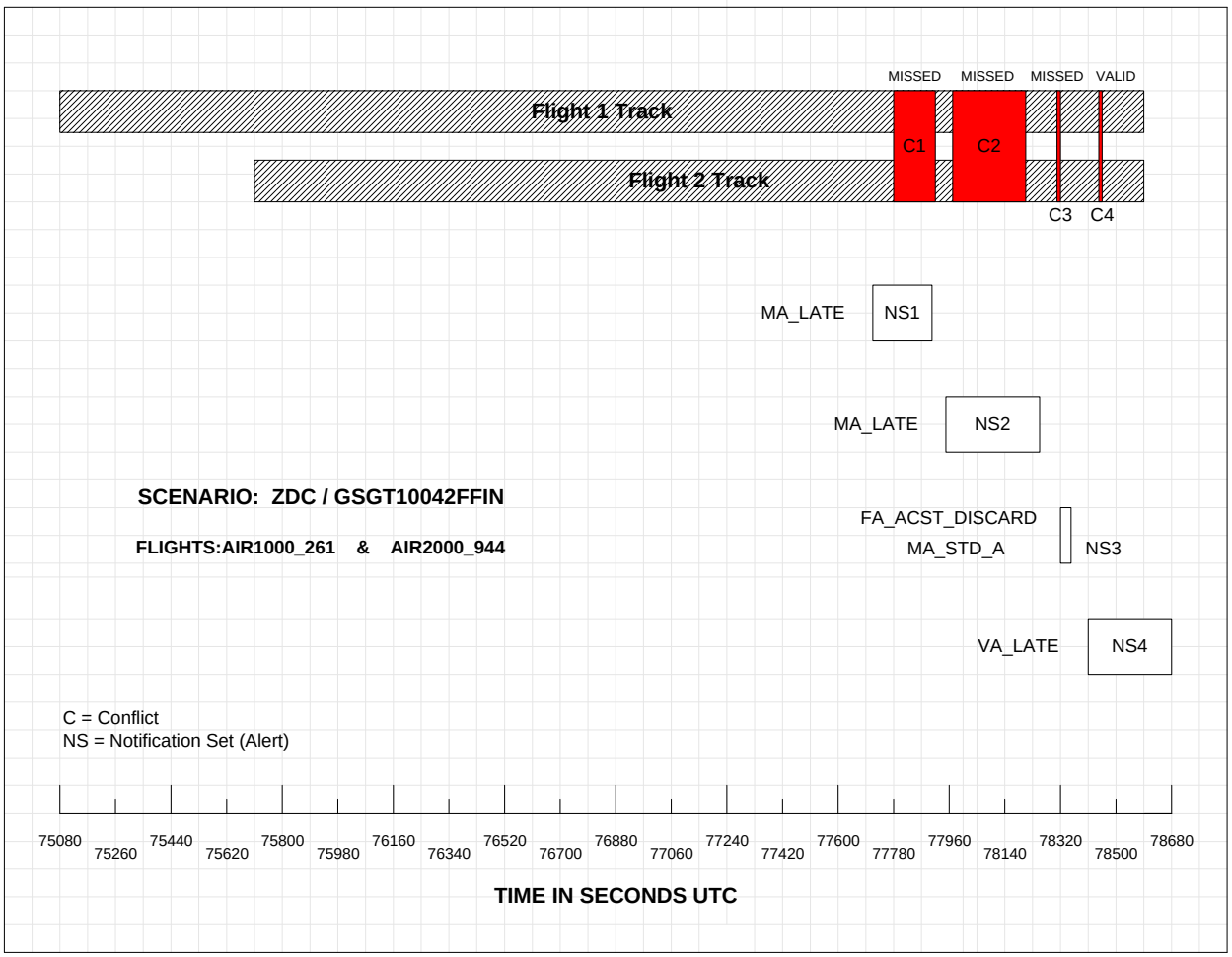

Figure 3: Example of Tactical Conflict Alert Data

A. Missed Alert Rate

The missed alert rate is the number of missed alerts divided by the number of conflicts and expressed as:

$$R_{MA} = MA / C \tag{1}$$

where R_{MA} is the rate, MA is the count of missed alerts and C is the number of conflicts. In to Table 2, the number of missed alerts MA is the sum of the counts for *MA_STD_A, MA_STD_B, and MA_LATE*. The number of conflicts C is the sum of the counts for *VA_STD, VA_LATE, MA_LATE, MA_STD_A,* and *MA_STD_B*. The number of discarded popup conflicts, *MA_DISCARD*, is not included.

B. False Alert Rate – Overall

The overall false alert rate is the number of false alert events over the total number of alerts. This intuitive metric provides the proportion of alerts that are falsely presented to the ATC system. It is expressed in the following equation:

$$R_{FA} = FA / A \tag{2}$$

where R_{FA} is the false alert rate, FA is the number of false alerts, and A is the number of alerts. Again referring to Table 2, the number of false alerts FA is the sum of the counts for *FA_STD1, FA_STD2_A, FA_STD2*_B, and *FA_STD3*. The number of alerts A is the sum of the counts for *FA, VA_STD, VA_LATE,* and *MA_LATE*.

C. False Alert Rate – by ρ_m Value

The false alert rate is the number of false alerts divided by the number of encounters where an encounter is a near miss. Since some alerts occur for large separations of two aircraft, the term "near" is used rather loosely in this study. Any aircraft-to-aircraft separation that is simultaneously less than 40 nautical miles and 8000 feet is considered to be an encounter. A false alert is most likely when a pair of aircraft comes close to violating the minimum separation standards but does not quite come close enough to cause a conflict. As the minimum separation becomes larger, a false alert becomes less likely. The rate of decrease of the false alert rate with increasing minimum separations is a good measure of the effectiveness of the tactical conflict alert processing. This rate is obtained by calculating a histogram of false alert rates for increasing ranges of aircraft-to-aircraft minimum separation. The separation is defined by using a normalized metric, ρ_m. This metric is explained in Ref. 9. It is calculated as follows.

First, the horizontal separation of two aircraft at any point in time is normalized by dividing their horizontal separation in nautical miles by 5, the minimum horizontal separation standard. Second, the vertical separation in feet, at the same point in time, is divided by the vertical separation standard of 1000 feet. Third the maximum value of these two parameters is selected. This number is the normalized separation of the two aircraft at the chosen point in time. Fourthly, the minimum value of this selected value is found for all points in time for which there is track position data for both aircraft in the scenario. This number is the min-max ratio ρ_m for the aircraft pair in the scenario.

The ranges of ρ_m to be used in this study to partition the false alerts are from 1 to 2, 2 to 3, 3 to 4, up to 7 to 8. The false alert rate for a given bin is the number of false alerts for the aircraft pairs whose ρ_m value is contained in the range of values for the bin, divided by the number of encounters whose ρ_m values fall into range of values of ρ_m contained in the bin. The false alert rate will drop off with increasing values of ρ_m, and should any instances of false alerts with ρ_m values greater than 8 occur, the last bin boundary of 8 is increased to include that value.

Nominally there should be no false alerts for aircraft pairs which have a ρ_m value of less than one, since this value corresponds to a true conflict. However, false alerts may occur when the aircraft separation has a ρ_m value of less than one. If an alert is posted for a real conflict but is withdrawn before the conflict starts, the alert is labeled false. In addition the ρ_m may have a value of less than one when there is no conflict because the loss of separation did not last long enough to be called a conflict (6 seconds is required).

D. Timeliness Metrics

Timeliness metrics for conflict predictions serve to estimate the amount of prediction time provided for valid predictions. The warning time given by an alert is the time interval between the time of the alert and the time the conflict starts. Warning times have been shown graphically in Figure 3. Warning times are calculated only for the set of matched NSs evaluated as valid alerts. The calculation is listed below in Eq. 3.

$$AWT = ACST - NS_{t_0} \tag{3}$$

where *AWT* is the Actual Warning Time and *ACST* is the Actual Conflict Start Time, and NS_{t_0} is the start time of the NS evaluated as a valid alert.

Standard descriptive statistics on the *AWT*, such as average, median, standard deviation, and percentiles are calculated for the population of valid alerts. Histograms and other data visualization methods can all be employed to analyze these time intervals as well.

VIII. Flight Example

An effective method in describing the detailed processing rules presented in the previous sections is presentation of an actual flight example. This section presents such an example, which includes both a conflict and matched alert. However, the alert is too late to meet the required 75 seconds of warning time. Linear extrapolation of the flight paths is performed to see if the warning time requirement of 75 seconds can be relaxed. It is determined that the warning time requirement cannot be relaxed because the extrapolations consistently predict the conflict. Thus, the alert is labeled a *MISSED_LATE*.

In this example, Flight TEST1 is a Boeing MD80 series aircraft flying from Palm Beach, Florida to LaGuardia airport in New York City, with intermediate fixes at PERMT, ILM, TYI, HPW, PXT, and KORRY3. Flight TEST2 is an Airbus A300 series aircraft flying from Orlando, Florida to Boston, Massachusetts, with intermediate fixes at CHS, JFK, and ORW3. During the time frame of this example, both aircraft were assigned to and were flying at FL350**.** Figure 4 below depicts the flight paths of these two aircraft immediately before and after the conflict that occurred.

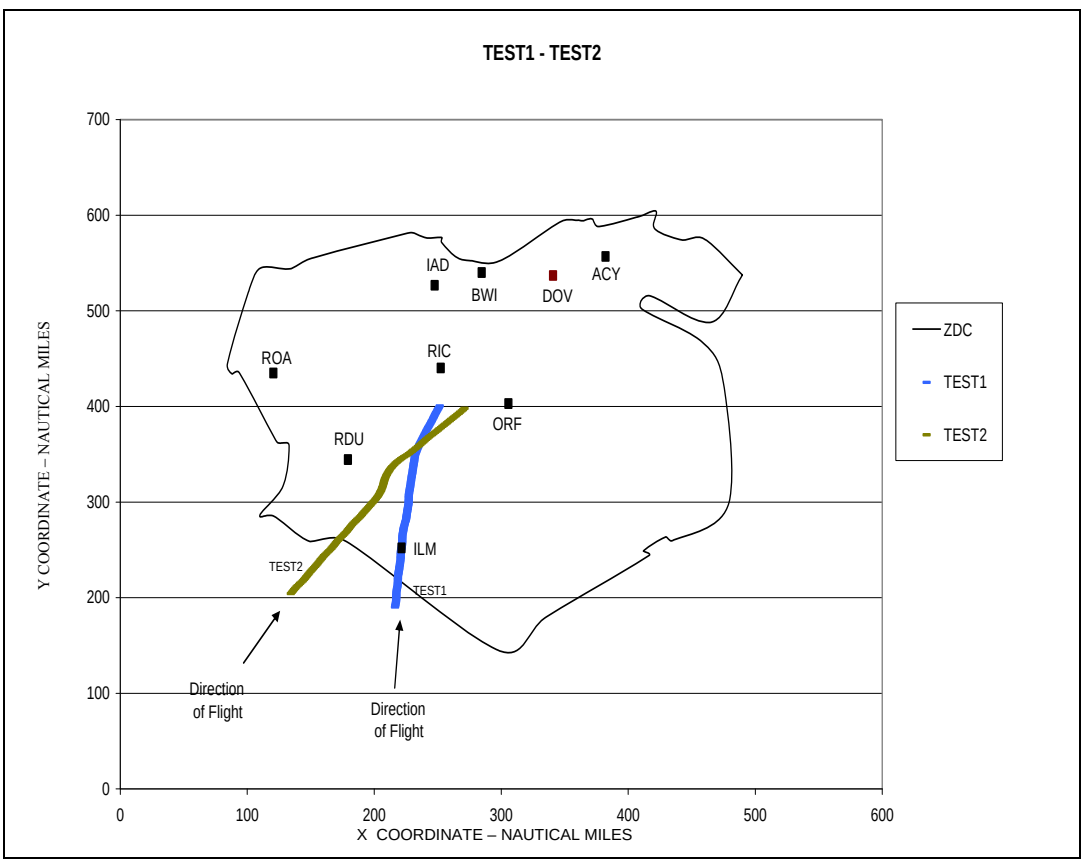

Figure 4: Flight Paths of Flights TEST1 and TEST2

A. Conflict Geometry

Track data for these two aircraft within Washington Center airspace started at 82130 seconds (UTC) into the scenario. At this point TEST1 and TEST2 were each flying at FL350, were separated by 22.5 nm, and were flying headings of 19 and 3 degrees respectively, resulting in an encounter angle of 16 degrees. Over the next five minutes, both aircraft gradually turned right, the TEST2 aircraft more so, such that at the time of conflict start, 82430 seconds into the scenario, the encounter angle had increased to 36 degrees. During the conflict, the TEST1 flight continued a gradual turn to the right, towards the TEST2 aircraft, until at the time of conflict end at 82560 seconds into the scenario the encounter angle had decreased to 23 degrees. During the conflict, the point of closest approach for the two aircraft was 2.1 nm at 82480 seconds into the scenario.

Scenario track data related to the conflict was processed graphically using an animation program. Figure 5 is a screen capture of this animation at the time of conflict start, and Figure 6 is a screen capture at the time of conflict end.

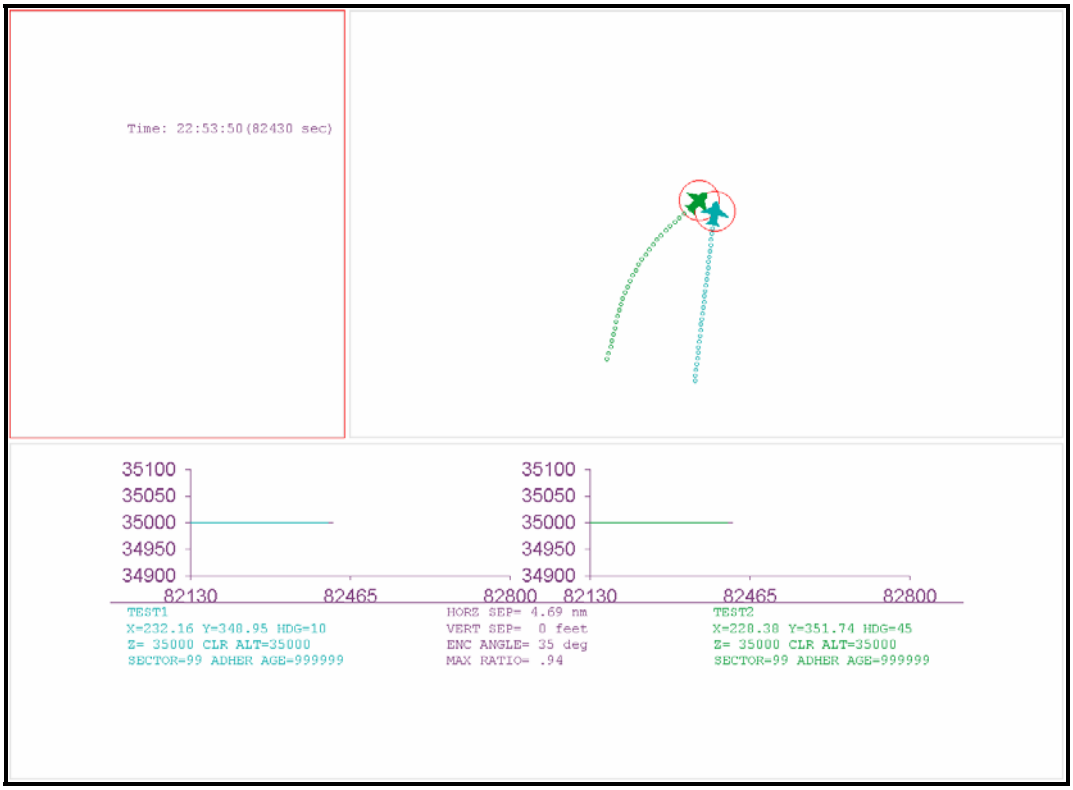

Figure 5: Start of Conflict

B. Analysis

As a result of this conflict, the HCS generated a NS. The first entry in this NS, the ADD alert, was generated at 82357 seconds into the scenario, and represents the earliest notification by the HCS of the pending conflict. The actual start of the conflict was at 82430 seconds into the scenario, so the actual warning time provided was 73 seconds, or 2 seconds less than the specified minimum warning time of 75 seconds.

The track data for the two aircraft involved was examined to determine if the late warning time might be excused. First, it was verified that good track data was present for both aircraft during the time immediately preceding the conflict. Second, an extrapolation test was performed on the track data to check for the presence of an unexpected maneuver that might impact warning time. Starting at a point one sample interval prior to the actual start of the conflict, and working back one sample interval at a time to a point 80 seconds prior to *ACST*, a straight-line estimate was made of each aircraft's track based upon current speed, altitude, and rate of climb. As shown in Table 3 below, all of the extrapolations predicted a conflict in that minimum-max-ratio was less than 1.0. Given the absence of any warning time reduction factors, this conflict was placed in the *MISSED LATE* category.

IX. Conclusions

Performance metrics have been defined for determining how well an ATC system predicts near term (tactical) conflicts. The metrics are suitable for use with a simulation run of a sample of recorded and time shifted field air traffic data. Since the results are a direct result of the detailed rules defined and other evaluations could have defined the rules differently, they are more suited to comparing ATC system performances than for absolute measurements. Since the acceptance of the new ERAM ATC system by the government requires a comparison to the legacy HCS, these metrics are suitable testing. The calculation of the metrics is complex due to the nuances introduced by both the characteristics of the data and the limitations of simulation. The complex details of the computation required are of critical importance because they directly affect the results obtained.

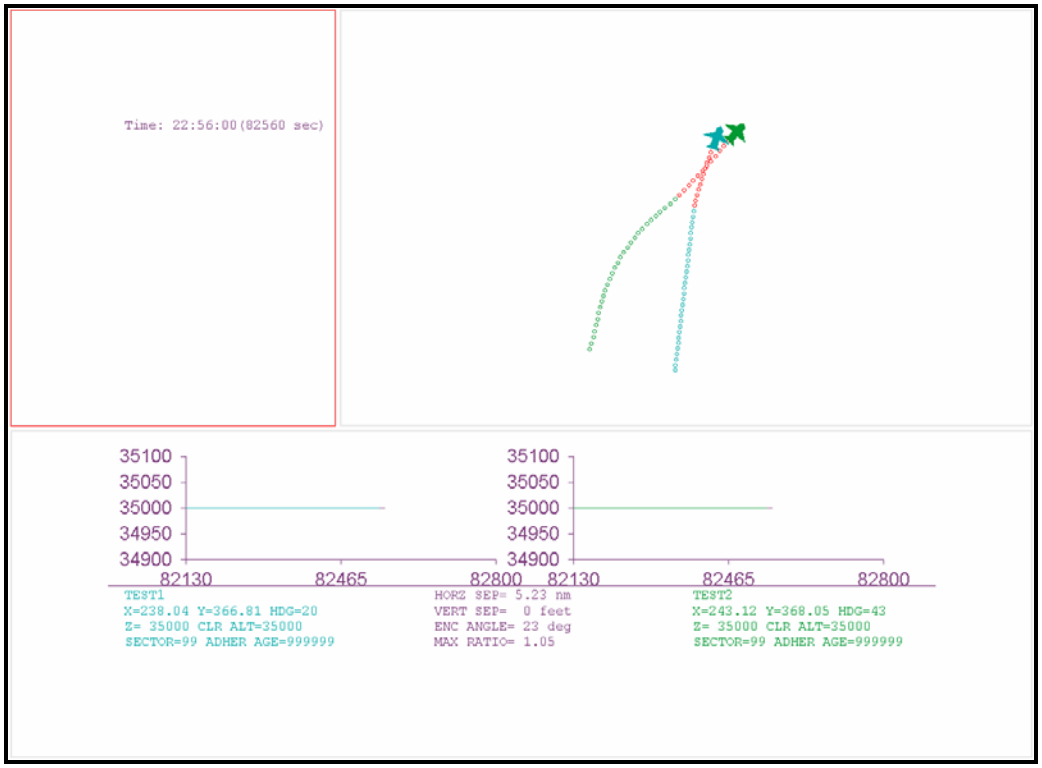

Figure 6: End of Conflict

Table 3: Conflict Prediction Results for Flights TEST1 and TEST2

Extrapolation Time (sec)	Min-Max-Ratio	Minimum Horizontal Separation (nm)	Minimum Vertical Separation (ft)	Predicted Conflict Start Time (sec)
82350	0.135	1.838	0	82470
82360	0.002	0.245	0	82460
82370	0.002	0.229	0	82450
82380	0.004	0.332	0	82450
82390	0.005	0.364	0	82440
82400	0.001	0.168	0	82440
82410	0.001	0.178	0	82440
82420	0.009	0.481	0	82430

Acknowledgments

The authors of this paper would like to acknowledge the FAA's William J. Hughes Technical Center's Thomas Choyce of the ERAM Test Group. His guidance and input on the methodology of implementing the metrics documented was not only invaluable but critical to its development. A special thanks is due to George Chandler from the FAA's Simulation and Analysis Group, who was responsible for documenting the flight example in this report. The staff of the William J. Hughes Technical Center's Integration & Interoperability Facility (IIF), most notably Engineering Lead William Monsour the IIF Lab Manager Steve Souder performed the simulation work. Finally, we would like to thank the FAA's ERAM Program Office for funding this study via the ERAM Test Group.

Acronyms

ACID	Aircraft Identifier (Call Sign)
ACST	Actual Conflict Start Time
ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
AWT	Actual Warning Time
C	Conflict
CID	Computer Identifier
CPAT	Conflict Probe Assessment Team
ERAM	En Route Automation Modernization
FA	False alert (a.k.a. nuisance alert)
FAA	Federal Aviation Administration
FL	Flight Level
HCS	Host Computer System
Host	ARTCC main frame computer
IIF	Integration and Interoperability Facility
MA	Missed alert
MA_LATE	Missed alert – Late
MA_STD	Missed alert – Standard
MWTR	Minimum Warning Time Requirement
NC	No call
NS	Notification Set
$\mathbf{R_{FA}}$	False alert rate
$\mathbf{R_{MA}}$	Missed alert rate
$\boldsymbol{\rho_m}$	Min-max ratio (separation distance parameter)
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

References

[1]Lockheed Martin Transportation and Security Solutions, "En Route Automation Modernization (ERAM), System/Segment Specification (SSS), Volume II, System Architecture Design Document (SADD)", Rockville, MD, June 2003.

[2]Cale, M. L., Paglione, M. M., Ryan, H. F., Timoteo, D., Oaks, R. D., "User Request Evaluation Tool (URET) Conflict Prediction Accuracy Report," DOT/FAA/CT-TN98/8, FAA Wm. J. Hughes Technical Center, Atlantic City, NJ, April 1998.

[3]Paglione, M. M., R. D. Oaks, and H. F. Ryan, "Methodology for Evaluating and Regression Testing a Conflict Probe", Proceedings of the Digital Avionics Systems Conference (DASC), Washington, DC, October 2004.

[4]Brudnicki, D.J. and A.L. McFarland, "User Request Evaluation Tool (URET) Conflict Probe Performance and Benefits Assessment," 1st USA/Europe Air Traffic Management R&D Seminar, FAA, Washington, DC, 1997.

[5]McNally, B.D., R. Bach, and W. Chan, "Field Test of the Center-TRACON Automation System Conflict Prediction and Trial Planning Tool," Paper No. 98–4480, *AIAA Guidance, Navigation, and Control Conference*, Washington, DC, 1998.

[6]Van Doorn, B., B. Bakker, and C. Meckiff, "Evaluation of Advanced Conflict Modelling in the Highly Interactive Problem Solver," 4th USA/Europe Air Traffic Management R&D Seminar, FAA, Washington, DC, 2001.

[7]Bilimoria, K.D., "Methodology for the Performance Evaluation of a Conflict Probe," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 3, AIAA, Washington, DC, 2001, pp. 444-451.

[8]Loureiro, T., K. Blin, E. Hoffman, and K. Zeghal, (2001) "Development of a Tool for Comparing Conflict Detection Algorithms for Air Traffic Management," Paper No. 2001-4053, *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Washington, DC, 2001.

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

[10]Paglione, M., Oaks, R. D., Summerill, J. S., "Time Shifting Air Traffic Data for Quantitative Evaluation of a Conflict Probe," *Proceedings of the American Institute of Aeronautics and Astronautics Guidance, Navigation, and Control Conference*, AIAA 2003-5343, Washington, DC, August 2003.

[11]Paglione, M. M., Oaks, R. D., Bilimoria, K. D., "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," *Proceedings of the American Institute of Aeronautics and Astronautics Aviation Technology, Integration, and Operations Technical Forum*, AIAA 2003-6724, Washington, DC, November 2003.

[12]Ryan, H. F. and M. M. Paglione, "Heuristic Methods to Post Process Aircraft Radar Track Data", *Proceedings of the American Institute of Aeronautics and Astronautics Guidance, Navigation, and Control Conference*, AIAA 2003-5575, Washington, DC, August 2003.

[13]Paglione, M. M., H. F. Ryan, and S. Liu, "Evaluation of En Route Host Computer System's Tactical Alert Processing: Description of Methodology", Technical Note, DOT/FAA/TC-TN07/13, FAA Wm. J. Hughes Technical Center, Atlantic City, NJ, March 2007.