AIRCRAFT CONFLICT PROBE SENSITIVITY TO WEATHER FORECASTS ERRORS

Mike Paglione, Federal Aviation Administration, Atlantic City, New Jersey
Hollis F. Ryan, General Dynamics Information Technology, Atlantic City, New Jersey

Abstract

This study investigated the User Request Evaluation Tool's (URET) prediction sensitivity to weather forecast error. A quantitative experiment was designed and performed by the Federal Aviation Administration's Conflict Probe Assessment Team (CPAT) to evaluate the impact of weather forecast errors on URET trajectory and conflict predictions. The experiment used approximately two hours of traffic data recorded at the Indianapolis en route center in May 1999. The flights were time shifted to generate a sufficient number of test conflicts using a genetic algorithm technique developed by CPAT. The resulting scenario was input into the URET Prototype System. To induce weather forecast error, the weather input file (Rapid Update Cycle, RUC) was altered by adding 20 or 60 knots to the wind magnitude, 45 or 90 degrees to the wind direction, and 5 or 15 degrees Kelvin to the air temperature. This produced seven URET runs for the experiment – the unaltered control run and six treatment runs. The analysis compared the control run against the treatment runs. A methodology was developed to compare the trajectory and conflict prediction accuracy of these runs. A statistical analysis provided evidence that the forecast errors in wind magnitude and direction had significant effect on the longitudinal trajectory error and a modest impact on retracted false alerts, which caused at most an increase in the false alert probability by six percent. It also showed that the air temperature runs did not have a significant effect. Based on this experiment, a controller suspecting errors in the input wind forecast should expect only a modest impact on URET predictions. The impact would mainly be a moderate increase in the number of retractions of its conflict predictions (defined in this study as a retracted false alert). If the controller notices an increase in retractions, it may be symptomatic of inaccurate wind forecasts, which should be investigated.

Introduction

Most air traffic service providers, such as the Federal Aviation Administration, have forecasted that the growing air traffic will continue to cause congestion degrading efficiency and potentially safety unless advances in ground and airborne automation are implemented. One of the most important ground based tools is a conflict detection tool or conflict probe (CP). A CP is a decision support tool (DST) that provides the air traffic controller with predictions of conflicts, or loss of minimum separation between a pair of aircraft or between an aircraft and protected airspace, for a parametric time into the future, typically 20 minutes. A CP predicts the flight path of an aircraft, continuously monitors that flight path from current aircraft position information, and probes for conflicts with other aircraft and incursions into restricted airspace. A CP makes these predictions based on air traffic control clearances, radar surveillance position reports, aircraft and airspace characteristic data, and weather forecasts. Therefore, inaccuracies in this input data is expected to cause error in the predictions the conflict probe makes.

The Conflict Probe Assessment Team (CPAT) within the Simulation and Analysis Group was tasked to examine the sensitivity of the FAA's en route CP, known as the User Request Evaluation Tool (URET), on one of its input sources, specifically the weather forecasts. To accomplish this, CPAT developed tools to induce weather forecast errors on the input weather data. The objective of the analysis was to determine URET's trajectory and conflict prediction sensitivity to degraded weather forecasts. The study was first published in the draft report [1] and now within this paper.

U.S. Government work not protected by U.S. Copyright.

Previous Research

Several organizations and researchers have investigated the impact of weather forecast errors on DSTs. However, no one has performed an experiment and applied a comprehensive analysis to determine both URET's trajectory and conflict prediction's sensitivity to weather forecast error. The original developers of URET, MITRE Corporation's Center for Advanced Aviation System Development, came the closest in [2]. The sensitivity of the URET Prototype to weather data was measured by running the Algorithmic Evaluation Capability (AEC) version of URET (i.e. a simulation version of URET) using a five-hour air traffic scenario with and without its input weather forecast files. The conflict alerts generated and their predicted warning times (the time intervals between the posting of the alerts and their predicted conflict start times) for the two runs were compared. The trajectory accuracy and reconformance[1] rates were also compared. It was found that the lack of weather data increased the longitudinal track-to-trajectory deviations at large look-ahead times and the lateral and vertical deviations were relatively unchanged. The predicted warning times for the alerts common to both runs increased slightly. Alerts were generated by the no-wind run, which were not generated by the baseline run and vice versa. The trajectory reconformance rate went up slightly. It was concluded that URET can provide valuable conflict alert information in the absence of weather forecast data and the major effect was a modest increase in the number of marginal conflict alerts. The study did not examine these conflict predictions in terms of their accuracy degradation in the absence of the weather forecasts but reported that the quantity of predictions increased as a result and inferred they were caused by increases in the longitudinal track-to-trajectory deviations.

In [3], a collaborative effort of researchers from Massachusetts Institute of Technology Lincoln Laboratory (MIT/LL), National Aeronautics Space Administration (NASA) Ames Research Center, and National Oceanographic and Atmospheric Administration (NOAA) Forecast Systems Laboratory (FSL) reported on a year-long weather study. The data was collected over the Denver Air Route Traffic Control Center (ARTCC) airspace. The study was conducted to better understand wind prediction errors, to establish metrics for quantifying large wind prediction errors, and to validate two approaches to improve wind prediction accuracy. Besides an exhaustive analysis of 13 months of wind prediction data from the Rapid Update Cycle (RUC) forecasts and the Aircraft Communication Addressing and Reporting System (ACARS), a series of aircraft flight tests were also performed. The on-average wind prediction accuracy was reported to be sufficient, but the analysis revealed that occasionally large errors existed over large regions of airspace. It was concluded that these large errors were present sufficiently to degrade the operational acceptance of DST predictions. One key result of the flight tests reported that the wind prediction error caused the greatest impact to trajectory prediction error at a look-ahead time of 20 minutes. Furthermore, two approaches were presented that improved the original RUC wind predictions and greatly reduced the occurrence of these large wind prediction errors. Therefore, the research in [3] provided insight into the wind prediction errors and guidance on realistic error levels to investigate for the CPAT's study on URET. It also supplied further evidence on the impact wind error has on DST predictions.

In another study, documented in [4], MITRE CAASD evaluated the use of aircraft speed and wind reports to reduce trajectory prediction errors in URET algorithms. The reports were obtained from the aircraft in flight via ACARS and added to the trajectory modeling process. It was found that the aircraft reports improved the trajectory longitudinal prediction error by an average of 10% to 15%. The number of trajectory reconformances was also reduced. The ACARS reported data was used to create a statistical model of the airspeed and wind variations. Therefore, the research in [4] provided insight into the wind errors themselves as well as the impact on URET's trajectory predictions. However, it had provided no analysis on the impact on URET's conflict prediction

---

[1] Trajectory reconformance is defined as URET's method of monitoring and rebuilding its aircraft trajectory predictions when the current reported track position is outside the trajectory's conformance bounds. The conformance bounds are regions of uncertainty built around the trajectory centerline. The more URET rebuilds or reconforms a trajectory indicates its uncertainty in its trajectory prediction.

accuracy, which is a major emphasis in this CPAT study.

In [5], the researchers from NOAA compared the accuracy of the RUC-1 and the newer RUC-2 weather forecasts. RUC-2 has higher resolution, a one hour assimilation cycle rather than a three hour assimilation cycle, more input data, and better physical models. The actual winds aloft were obtained from aircraft in flight via the ACARS data link. The differences between the observed winds aloft (from ACARS) and the predicted winds aloft (from RUC) were used to calculate along path distance prediction errors and errors in the predicted times of arrival. An analysis of 140,000 flights, collected over a 13-month period, found that 15 minute en route segments accumulated time of arrival errors of 15 seconds and that 15 minute ascent/descent route segments accumulated time of arrival errors of eight seconds. The focus of this report is on the quality of the weather data, which once again provides insight into the underlying accuracy of these weather forecasts. This research does provide some analysis on the effects of the weather forecast errors on trajectory predictions and none on the sensitivity to a DST's aircraft conflict predictions.

In yet another study, [6], MITRE CAASD reported on the accuracy lost in forecasted winds aloft when the data is provided in a resolution below what is available with either RUC-1 or RUC-2. A MITRE tool known as Winds Aloft Require Evaluation System (WARES) was presented and used to filter bad data from RUC, ACARS, and Meteorological Data Collection and Reporting System (MDCRS) data sources. The experiment paired aircraft wind reports (along and cross-track wind vector components) with the forecasted reports, filtered erroneous observations and statistically compared the difference. The study presented the specifics of twenty independent experiments that corresponded to the combinations of data resolution and forecast intervals available with RUC-1 and RUC-2. The statistical analysis used the Root Mean Square Wind Vector Error (RMSWVE) which is the standard RMS statistic employed by WARES. The study concluded coarser wind models like RUC-1 relative to RUC-2 can reduce the random noise in the wind aloft forecasts and consequently offset any loss of accuracy due to the decreased resolution. Thus, the study deduced that resolution based requirements for gridded-forecast weather data (like RUC) do not necessarily provide the best available accuracy regarding winds aloft prediction. This research documented in [6] provides a thorough background into URET's input weather forecast files (RUC files) and thus this CPAT study. However, it does not provide any analysis of the impact for URET's predictions and suggests this as a future research area.

In conclusion, these references provide an extensive foundation from which the FAA CPAT weather sensitivity study is applied. The references present detailed descriptions and performance data on the existing weather forecast products and in some cases offer improved solutions for the future. Since the weather products are their primary focus, they only indirectly examined the impact on DSTs like URET. The study documented in [2] was the exception. It directly examined URET sensitivity to the absence of timely weather forecasts, but the impact focused mainly on URET trajectory accuracy and only partially on URET conflict prediction accuracy. Therefore, this section's review of related literature provided further justification of performing a comprehensive analysis on the impact on both URET's trajectory and conflict prediction accuracy. In the more recent MITRE CAASD study documented in [6], it was concluded: "Future research will include performing sensitivity analysis of Air Traffic Management (ATM) automation to winds aloft error." This is precisely the objective of the CPAT weather study documented in this paper.

Design of Experiment

The focus of this study is to investigate URET's prediction sensitivity to weather forecast error. To examine these errors, a quantitative experiment was developed. The objective of the experiment was to evaluate what impact weather forecast errors have on URET trajectory and conflict predictions, if any, and determine whether or not the impact is statistically significant. To understand this phenomenon, wind and air temperature forecast errors were induced by altering URET's input weather forecast files.

The experiment consisted of extracting traffic data from Indianapolis ARTCC field recordings made on May 26, 1999. Two hours of traffic data was extracted and time-shifted to generate a scenario with a total of 211 aircraft-to-aircraft conflicts. The method used to generate this time-shifted traffic sample of conflicts is documented in [7] and [8]. The experiment used the same traffic scenario throughout and only altered the weather forecast files.

A set of software tools were developed that altered the weather forecast files that were input into URET. In brief, the forecasted weather data was obtained from the National Weather Service (NWS) for same day in May 1999. This forecast data was formatted as Rapid Update Cycle 2 (RUC-2) gridded-binary files. As the main input source for the experiment, these files were modified throughout in wind magnitude, wind direction, and air temperature. The control run had no RUC file modifications, while all treatment runs had modified RUC files. The URET Prototype was run with the same air traffic scenario and these modified weather files in single center operation.

The three weather factors were altered individually at two different levels. The selection of these factors and levels were chosen based on research presented in [3] and [4] and an internal empirical study on the control run RUC file. For example, in [3] a wind magnitude error of up to 60 knots was observed in a year-long study over Denver Air Route Traffic Control Center (ARTCC) airspace. In CPAT's analysis of the RUC file from May of 1999, a software tool was developed and implemented that extracted the wind magnitude, wind direction, and air temperature for each flight's Host Computer System (HCS) reported positions. The resulting weather forecasts were then extracted and summarized. The level one or low level was selected to cover approximately 50 percent of the data range. The level two is a higher value selected to cover most of the data range. Hence, wind magnitude was modified by adding 20 knots or 60 knots to all the forecasted winds. Similarly, wind direction was modified by adding 45 degrees or 90 degrees. Air temperature was modified by adding 5 degree Kelvin or 15 degree Kelvin to all the temperature forecast grid-points. This resulted in a total of seven URET runs; the one control run and six treatment runs. Table 1 lists these seven runs and their assigned run codes. These run codes are used throughout this paper to refer to the associated URET run. The analysis compares each treatment run against the control run and in some cases the other treatment run in its category. For example, for the wind magnitude factor the control Run 000 is compared to the wind magnitude run with 20 knots added, Run 100, and the Run 200 with 60 knots added. For this example, the comparisons would be listed as 000-100 and 000-200. In some cases, the 100-200 will also be explored.

Table1. Experiment Control and Treatment Combinations

Factor	Level	Run Code
Control Run	No change to RUC file	000
Wind Magnitude	Add 20 knots	100
	Add 60 knots	200
Wind Direction	Add 45 degrees	010
	Add 90 degrees	020
Air Temperature	Add 5 degrees Kelvin	001
	Add 15 degrees Kelvin	002

Results

A designed experiment is a statistical method used to identify which system input factors and which levels within the factors significantly affect the output response. The methodology is to first identify factors and levels that cover some meaningful range of input variables, next make multiple runs of the system at the identified factors and levels and finally statistically analyze the system response. In this study, the levels believed to cover a valid range for each weather forecast factor (i.e. the wind magnitude, wind direction, and air temperature as described previously) were determined and multiple runs of the URET system were conducted using RUC files where one or more of the factors were altered. Next the results (measured trajectory deviation error) from the multiple runs were compared to determine which weather forecast factors had an impact on trajectory accuracy. With this approach it is also possible to alter multiple factors, run the experiment and get information on possible interactions between factors and factor levels. The study initially envisioned two blocks of runs to evaluate both single and combinations of multiple factors.

The focus of this paper was on the main weather forecast factors only (wind magnitude and direction, air temperature). This was accomplished by systematically altering one factor while holding the remaining two factors at their nominal level and running the experiment. The statistical analysis then compares the difference in trajectory error between a run using the altered file and a baseline run using the nominal file or between two altered files where the factors are at different predetermined levels. This type of analysis where only a single factor is altered provides information on which levels within that factor cause a statistically valid change (if any) in trajectory deviation error. Thus, the main effect of the weather forecast factors would be determined.

Three primary analyses were executed for this study. First, a trajectory accuracy analysis was performed. This analysis examined the trajectory prediction accuracy differences between the baseline and treatment runs using metrics in the horizontal, vertical, lateral, and longitudinal dimensions. Next, the trajectory stability was measured, indicating how often trajectories are rebuilt after being determined to be out of conformance. This occurs when an aircraft's track position is outside the region of uncertainty (conformance region) centered at the trajectory centerline. Finally, the conflict prediction accuracy was investigated. Categorical statistics were employed to determine if the missed and false alert predictions were equivalent between the baseline and treatment runs. The following subsections will present these results, respectively.

Trajectory Prediction Results

The trajectory accuracy sample distribution is determined to be non-normal in [1], so a nonparametric approach to statistical testing was utilized. Another technique considered was to transform the data to achieve normality and do an analysis on the new data set. This idea was discarded as the study involves the management of multiple data sets resulting from the numerous combinations of look-ahead times and trajectory error types. Also, correcting for normality may unequally affect sample variance in different data sets and equality of variance is a requirement for many of the traditional statistical tests.

Nonparametric data analysis is an alternative that makes few assumptions regarding the characteristics of the parent population. These methods are based on data ranks and use the corresponding sampling distributions to determine significance. This field of statistics is considered classical in the sense that the methods are well investigated, understood and accepted. The nonparametric test appropriate for this study is a pair wise comparison method called the Wilcoxon signed-rank test [9]. A detailed description of this non-parametric technique applied to a trajectory prediction accuracy problem was presented in [10].

Before presenting the statistical test results, the median horizontal error and wind magnitude levels as a function trajectory prediction look-ahead time are illustrated in Figure 1 (in units of nautical miles versus seconds). Look-ahead time is the prediction time horizon. As expected, the error difference increases as the look-ahead time increases.

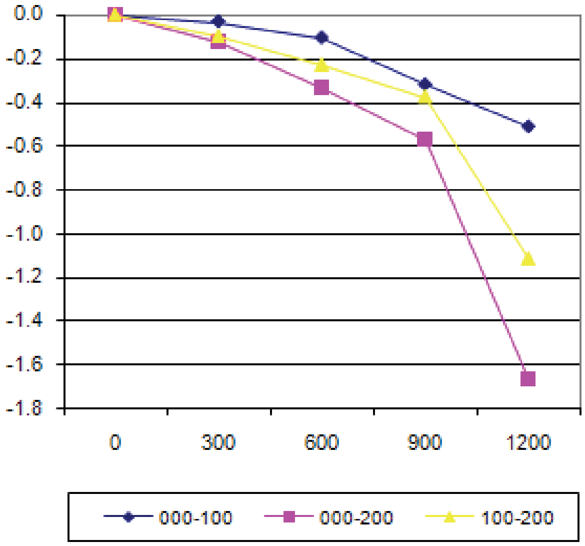

Figure 1. Plot of Median Difference in Horizontal Error by Look-Ahead Time for Wind Magnitude during Level Flight

The horizontal error is the time coincident straight line error distance between the trajectory-predicted point and the surveillance track point. The vertical error is the time coincident altitude distance between the predicted trajectory prediction and the reported altitude within the surveillance track point. The lateral error is the side-to-side time coincident error, and the longitudinal error is the along trajectory error. Again all these measures were applied with look-ahead times every five minutes from 0 to 20 minutes.

The nonparametric Wilcoxon signed-rank test was then applied. For this paper, only the straight and level results are presented, while [1] presented results for transitioning positions as well. Table 2 provides a summary of application of the Wilcoxon signed-rank test at a significance level 0.05.

Table 2. Trajectory Accuracy Statistical Test Results

Error Type	Run Codes	Look-Ahead Time				
		0	300	600	900	1200
Wind Mag. Horizontal Error	000-100	Yes	Yes	Yes	Yes	Yes
	000-200	Yes	Yes	Yes	Yes	Yes
	100-200	Yes	Yes	Yes	Yes	Yes
Wind Mag. Longitudinal Error	000-100	Yes	Yes	Yes	Yes	Yes
	000-200	Yes	Yes	Yes	Yes	Yes
	100-200	Yes	Yes	Yes	Yes	Yes
Wind Dir. Horizontal Error	000-010	No	Yes	Yes	Yes	Yes
	000-020	Yes	Yes	Yes	Yes	Yes
	010-020	Yes	Yes	Yes	Yes	Yes
Wind Dir. Longitudinal Error	000-010	No	Yes	Yes	Yes	Yes
	000-020	No	Yes	Yes	Yes	Yes
	010-020	Yes	Yes	Yes	Yes	Yes
Wind Dir/Mag Vertical and Lateral Error	000-010	No	No	No	No	No
	000-020	No	No	No	No	No
	010-020	No	No	No	No	No
Temperature All Errors	000-010	No	No	No	No	No
	000-020	No	No	No	No	No
	010-020	No	No	No	No	No

Adding 20 or 60 knots to the wind magnitude factor was determined to affect horizontal error and longitudinal error (one of the two orthogonal components to the horizontal error) considered in the study. The results for lateral error showed extensive zero observations which in a control minus treatment experiment logically support a conclusion of no difference between runs. Vertical error also showed extensive zeros which supports a conclusion of no difference.

Adding 45 or 90 degrees to the wind direction factor was determined to almost uniformly affect the horizontal and longitudinal trajectory deviation error for all look-ahead times. The single exception was at the lower factor level (000-010 run) for the zero look-ahead time. Additionally there was a conclusion of not significantly different from zero for the 000-020 zero look-ahead time group but no corresponding test result for horizontal error. Both vertical and lateral error were determined to be not significantly different from zero based on the extensive zero observations present on the data.

Adding 5 or 15 degrees to the air temperature factor was determined to have no statistically significant effect on trajectory deviation error. Test results for horizontal error, the orthogonal components lateral and longitudinal error, and for vertical error were consistently not significant for all other groups and were further validated by extensive zero observations in the data.

Trajectory Stability Results

As defined in [11], trajectory stability indicates how often trajectories are rebuilt after being determined to be out of conformance. This occurs when an aircraft's track position is outside the region of uncertainty (conformance region) centered at the trajectory centerline. In this study, the total number of trajectories for each run is listed in Table 3. From this table, the greatest number of trajectories was generated in Run 200 and was followed by Run 020. This is consistent with the treatment runs with the highest trajectory errors.

Table 3. Trajectory Counts Per Run

Run Code	Total Trajectory Count
000	5156
100	5268
200	5622
010	5271
020	5501
001	5136
002	5181

Conflict Prediction Results

When URET predicts that a future conflict will occur between two aircraft, it posts an alert to the air traffic controller's display. The alert remains posted until the conflict is past or is no longer predicted. Usually the controller will redirect one of the aircraft so that the conflict will not occur. URET automatically reads this change in flight path and deletes the alert. The alert may be updated (in time or space), while it is posted to the controller's display. The initial posting of the alert and its final deletion form a notification set which can be matched to an actual conflict.

A CP, like URET, is not perfect and does make mistakes in its conflict predictions. To quantify these errors, the conflict prediction accuracy metrics describe two fundamental events: a conflict and an alert. These events, which are not mutually exclusive, have four possible outcomes (see Table 4). The conflict accuracy metrics quantify the two fundamental error outcomes: missed alert and false alert. CPAT first defined these errors and rules to measure them in [12], but others have applied similar techniques in [13] and [14].

The most critical quantities to determine a statistical difference between runs are the quantity of missed alerts and false alerts as defined in Table 4. When comparing a pair of runs, in this study the baseline run versus the treatment run, the difference between these values measures the effect of the treatment level. To determine the statistical significance of this effect, one approach is to utilize a binomial distribution and perform a hypothesis test concerning the difference between population proportions [15]. However, this technique assumes that the respective runs are independent. For this study, each run is not independent, since they are run with the same air traffic scenario and altered weather files.

Table 4. CP Alert and Conflict Event Combinations [18, 12]

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT	CP predicts conflict and it occurs (VA -- valid alerts)	CP predicts conflict and it does not occur (FA -- false alert)
NO ALERT	CP does not predict conflict and it occurs (MA -- missed alert)	CP does not predict conflict and it does not occur (NC -- correct no-calls)
Total Number of Alerts	Total Number of Conflicts	Total Number of Non-Conflicts (Encounters that did not have conflicts)

An alternative technique is presented in [16], utilizing categorical data analysis techniques. For categorical data analysis, we examine the difference in the paired counts of missed and false alerts, which are mutually exclusive and exhaustive and a requirement for this test. They 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 (e.g. baseline run) when simultaneously getting a valid alert in Run B (e.g. treatment run) or vice versa for the opposite case. 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. This test is described in detail in [1] and later in [17].

The following Tables 5 and 6 provide the results of comparing all the alerts from the baseline run to each treatment run. None of the missed to valid alert comparisons in Table 5 exhibited a statistically significant effect. However, for the false alert to no-call events presented in Table 6, all the wind magnitude and direction treatments were statistically significant (i.e. the p-value < 0.05). All levels of the temperature treatments were not statistically significant.

Table 5. Missed and Valid Alert Comparison

Statistics	Comparison Runs: Run A Versus Run B					
	000-100	000-200	000-010	000-020	000-001	000-002
VA_MA	6	7	8	6	4	2
MA_VA	7	10	3	6	2	4
Total	13	17	11	12	6	6
Expected	6.5	8.5	5.5	6	3	3
X^2	0.077	0.529	2.273	0.000	0.667	0.667
P-value	0.782	0.467	0.132	1.000	0.414	0.414

Table 6. False and No-Call Alert Comparison

Statistics	Comparison Runs: Run A Versus Run B					
	000-100	000-200	000-010	000-020	000-001	000-002
FA_NC	38	52	41	50	32	50
NC_FA	61	84	59	92	27	45
Total	99	136	100	142	59	95
Expected	49.5	68	50	71	29.5	47.5
X^2	5.343	7.529	3.240	12.423	0.424	0.263
P-value	0.021	0.006	0.072	0.000	0.515	0.608

Conclusion

In summary, the experiment used approximately two hours of traffic data recorded at the ZID ARTCC in May 1999. The flights were time shifted to generate a sufficient number of test conflicts using a genetic algorithm technique developed by CPAT [7]. This time-shifted scenario was used as input to the URET Prototype. To induce weather forecast error, the weather input file (RUC) was altered by adding 20 or 60 knots to the wind magnitude, 45 or 90 degrees to the wind direction, and 5 or 15 degrees Kelvin to the air temperature. This produced seven URET runs for the experiment – the unaltered control run and six treatment runs (see Table 1 for listing of runs). The analysis consisted of comparing the treatment runs against this control run.

URET's trajectory predictions were analyzed for statistically significant effects. For both wind magnitude levels (20 and 60 knots), horizontal trajectory error and its along path component, longitudinal trajectory error, were statistically significant for all look-ahead times (i.e. 0 to 20 minutes). Similar results occurred for the wind direction runs. The air temperature runs did not differ statistically from the control run. As illustrated in Table 3, the errors in trajectory predictions cause URET to produce more trajectories per flight because it reconforms to correct for the longitudinal error. This is consistent with the trajectory error results, since it is only demonstrated in the wind treatment runs.

Similarly, the missed and false alert errors were evaluated for each run and then comparisons were performed. The complete comparison results are presented in the Tables 5 and 6. There was no evidence that the missed alert events differed between the control and any of the treatment runs. However, there was a difference detected for the false alert events in some of the treatment runs.

The air temperature treatment runs had no evidence of a difference for either missed or false alert error. For the wind magnitude and direction runs, the false alert frequency was statistically different, but the differences were not very high with an increase in false alert rate of at most six percent. Furthermore, the difference was dominated by the number of retracted false alerts. Retracted false alerts are alerts that are determined to be false because they are not associated with a matching actual conflict event but are labeled retracted because they are removed before the predicted conflict start time. In other words, they reflect that the CP changed its mind and withdrew the conflict prediction. This is consistent with the trajectory prediction accuracy results.

Operationally, weather forecasts may be inaccurate due to the presence of highly dynamic weather or outages in the interfaces to the NWS. This study showed that induced errors, as high as 60 knots in wind magnitude and 90 degrees in wind direction, had a modest effect on URET predictions. Therefore, a controller suspecting errors in the input wind forecast should expect only a modest impact on URET predictions. The impact would mainly be a moderate increase in the number of retracted false alerts, yet no overall affect on missed alert error. This is consistent with [11], which reported URET predictions still have utility under degraded weather forecast errors. If a controller notices an increase in retractions, it may be symptomatic of inaccurate wind forecasts, which should be investigated.

References

[1] Paglione, M., L. Charles, R. D. Oaks, S. Liu, H. F. Ryan, J. S. Summerill, October 2002, "User Request Evaluation Tool (URET) Conflict Probe Sensitivity to Weather Forecast Errors," Report Draft, FAA WJHTC/ACB-330.

[2] Lindsay, Kenneth S., June 1997, "Xeval Users Manual," MITRE/CAASD Technical Report, MTR 97W0000030.

[3] Cole, Rodney E., Green, Steve, Jardin, Matt, Schwartz, Barry E., Benjamin, Stanley G., June 13-16, 2000, "Wind Prediction Accuracy for Air Traffic Management Decision Support Tools," 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italy.

[4] Wanke, Craig, June 17-20, 1997, "Using Air-Ground Data Link to Improve Air Traffic Management Decision Support System Performance," MITRE Corporation, 1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France.

[5] Schwartz, Barry E., Benjamin, Stanley G., September 1998, "A Study of the Accuracy of

RUC-1 and RUC-2 Wind and Aircraft Trajectory Forecasts Using ACARS Observations," NOAA Technical Memorandum ERL FSL-22.

[6] Sherry, Joseph E., July 1999, "Rapid Update Cycle (RUC) Winds Aloft Forecast Accuracy and Recommended ATM Requirements (Revision 1)," MITRE/CAASD Technical Report, MTR 98W0000098R1.

[7] Paglione, M., R. D. Oaks, K. D. Bilimoria, November 2003, "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," AIAA Aviation Technology, Integration, and Operations (ATIO) Technical Forum.

[8] Oaks, Robert D., August 2002, "A Study on the Feasibility of Using a Genetic Algorithm to Generate Realistic Air Traffic Scenarios Based on Recorded Field Data," AIAA Guidance, Navigation, and Control Conference, Monterey, CA.

[9] Hollander, M., D. A. Wolfe, 1999, Nonparametric Statistical Methods, Second Edition, John Wiley & Sons, Inc.

[10] Paglione, Mike M., July 26-29, 2004, "Non-Parametric Approach to a Paired Data Hypothesis Test in Trajectory Accuracy Analysis," International Conference on Computational & Experimental Engineering and Sciences (ICCES), Madeira, Portugal.

[11] Lindsay, Kenneth S., April 21, 1997, "Sensitivity of AEC/URET Performance to Wind Data", MITRE/CAASD Technical Memorandum, F042-M-087.

[12] Cale, M., M. Paglione, H. Ryan, D. Timoteo, R. Oaks, April 1998, "URET Conflict Prediction Accuracy Report," DOT/FAA/CT-TN98/8, WJHTC/ACT-250.

[13] Brudnicki, D., W. Arthur, K. Lindsay, April 1998, "URET Scenario-based Functional Performance Requirements Document," MITRE/CAASD Technical Report, MTR98W0000044.

[14] Bilimoria, Karl, May-June 2001, "A Methodology for the Performance Evaluation of a Conflict Probe," Journal of Guidance, Control, and Dynamics, Vol. 24 (3).

[15] Devore, Jay L., 2000, Probability and Statistics for Engineering and the Sciences, Fifth Edition.

[16] Kachigan, Sam Kash, 1986, Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods.

[17] Paglione, M. M., R. D. Oaks, H. Ryan, October 24-28, 2004, "Methodology for Evaluating and Regression Testing a Conflict Probe," 23rd Digital Avionics Systems Conference (DASC), Salt Lake City, Utah.

[18] Paglione, M., M. Cale, H. Ryan, Fall 1999, "Generic Metrics for the Estimation of the Conflict Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool," Air Traffic Control Quarterly, Vol. 7 (3).

26th Digital Avionics Systems Conference
October 21, 2007