FUNCTIONAL PERFORMANCE ASSESSMENT OF THE USER REQUEST EVALUATION TOOL (URET)

Nicholas E. Rozen*
The MITRE Corporation, Center for Advanced Aviation Systems Development, McLean, VA 22102

Abstract

URET is designated as the conflict probe for the Federal Aviation Administration's Free Flight Program. Objective analyses of the conflict detection and notification functions of URET are known as *functional performance assessments*. This paper introduces metrics and techniques that are applied to perform these assessments on data recorded from the URET systems at the Memphis and Indianapolis Air Route Traffic Control Centers. Feedback from the operational testing provides us with a context within which to interpret the results of the functional performance analysis.

Background

The User Request Evaluation Tool (URET) is a prototype decision support system that aids Air Route Traffic Control Center (ARTCC) controllers in predicting and resolving conflicts, and managing flight data. The support that URET provides is intended to help relax some of the restrictions that are imposed in today's air traffic environment. Specifically, URET provides controllers with early warnings of aircraft-aircraft and aircraft-airspace conflicts. The early warnings can help the controllers manage traffic in a less structured environment and reduce and/or distribute their work over time. The URET prototype was developed by The MITRE Corporation's Center for Advanced Aviation System Development (CAASD) and has been deployed to and is currently operational in the Indianapolis (ZID) and Memphis (ZME) ARTCCs.

URET has evolved since the initial 1996 deployment to ZID. Each major delivery has undergone a functional and operational performance evaluation. Quantitative functional performance evaluations assess the accuracy of the core URET algorithms (trajectory modeling, track management and conflict detection). Results of recent URET functional performance evaluations are documented in (Lindsay, Rozen 2000). Qualitative operational performance evaluations, performed in the field, evaluate the operational acceptability and suitability of URET based on controller input.

At both ZID and ZME, URET is scheduled for operation 22 hours a day seven days a week. As of May 2001, the URET prototype had accumulated almost one million sector hours of daily use, providing extensive feedback via operational use and evaluation. The consensus that has emerged from the operational performance evaluations is that URET alerts are reliable and that URET is useful and beneficial as an en route decision-support tool. This has resulted in the use of the prototype to establish conflict probe requirements for the limited deployment of URET at other ARTCCs as part of the FAA's Free Flight Phase 1 (FFP1) program†, to develop operational procedures and training, and to develop methods to collect and analyze benefits-related data.

Scope

We will refer to quantitative assessments of the modeling and detection functions of URET as *Functional Performance Assessment*. Since the initial URET field development activities, Functional Performance Assessments have added value to both requirements specification and software/system quality assurance activities. The measures used to describe functional performance can be thought of as a hierarchy comprising three distinct levels.

* Senior Information Systems Engineer; MS W174;
Email: nrozen@mitre.org

† Free Flight Phase 1 is introducing new technologies and procedures at selected locations in the National Airspace System (NAS) through 2002; URET is one of five tools to be deployed in this time-frame.

Level 1
This is the Accuracy of Trajectory Modeler. The fundamental measurements include: Longitudinal, Lateral and Vertical Position Error (PE) (described in the second section of the paper); and, Vertical Prediction Accuracy.

Level 2
The Accuracy of Conflict Detector. This includes measurements of the alerting accuracy and timeliness, such as: Alert Rates; Conflict Warning Time Distribution; and, Alert Stability (not discussed in this paper).

Level 3
This is the Utility to Controller. The top level of the hierarchy, it is qualitatively assessed by: Specific evaluations by controller teams and feedback from Daily-Use operations; and, Case Studies about the relationship between alerts and controller actions.

This concept is illustrated in Figure 1. In the URET Functional Performance Assessment, the measures used to describe performance are called Technical Performance Metrics (TPMs) and belong to the first and second levels.

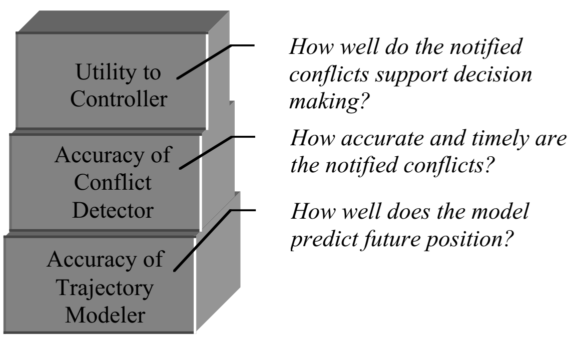

Fig. 1 "Hierarchy" of Performance Measures

This paper will present a set of results and techniques for the functional performance of URET. The first section of the paper describes URET's functionality, the selection of input data, and the TPMs. The following sections describe PE modeling and certain probability calculations that are used to generate the metrics. Their application to specific empirical data is explained. The paper concludes by addressing the interpretation of the results and the operational context.

Overview of URET's Conflict Probe Functions

URET's requirements include providing graphical views of the trajectories of multiple aircraft within a controller's assigned sector and advanced notification (i.e., conflict notification) when aircraft-to-aircraft or aircraft-airspace separation is predicted to fall below or near NAS-specified minima. URET maintains a complete set of current trajectories—for those aircraft with flight plans in the system—conforming to operational constraints such as altitude restrictions, Special Use Areas, and published Standard Instrument Departures (SIDS) and Standard Terminal Arrival Routes (STARS). URET uses Notification Logic, based on the likelihood of a predicted conflict, to decide when to alert the controller(s) at a selected sector. By the frequent off-line updates of the center's airspace adaptation, URET ensures that the modeled trajectories conform to real, valid operational constraints.

Automated Problem Detection (APD) algorithms detect both aircraft-to-aircraft and aircraft-to-airspace problems predicted to occur within a parameter "look-ahead" time interval in the future and provides the location, time and the geometry of the problem to the user interface.[5] Separation criteria are defined in FAA Order 7110.65. APD detects problems in a time frame that can be characterized as strategic when compared to current radar-based control systems, providing notification of up to 20 minutes for aircraft-aircraft problems, and longer for aircraft-airspace problems. A URET Trial Plan, created directly by the controller, uses APD to test modified flight profiles for problems.

URET relies on NAS Host Computer System track messages to maintain a trajectory that represents the planned route of each aircraft. When track reports no longer agree with the predicted trajectory, URET applies *Reconformance* logic in an attempt to rationally re-model trajectories in order to improve overall trajectory accuracy. Each aircraft's predicted positions along its trajectory are enclosed in box-shaped conformance regions defined by three parameters, one for each dimension. These parameter values allow for adjustments for route turns, altitude changes and navigation equipage specific to each aircraft. Reconformance logic is signaled by track management algorithms monitoring aircraft movement relative to their conformance regions.

In order to facilitate strategic ATC processes, URET tries to reconcile the goals of early detection of problems and keeping "nuisance" alerts to a minimum. Conflict detection and notification by URET is not applied within all airspaces. Around busy airports, in particular, there may be areas where control remains short term or tactical, and APD would not be operationally beneficial. The system provides the capability to identify areas where APD should not operate. The areas are known as APD Inhibited Areas (APDIAs). It is useful, therefore, to distinguish the conflict detection part of URET from a tracker-based warning system that operates on a short time-frame.

URET notifies controllers of conflicts based on the loss of separation within a maximum warning time parameter, nominally 20 minutes for aircraft-to-aircraft problems. The detection and notification processes reflect operational concerns and physical realities.

Notification time processing further manages the likelihood of erroneous alerts by varying the notification time based on the probability of the conflict occurring. Conflicts with a high probability of occurring are notified immediately and those with a low probability are delayed. The delay gives the system time to correct a "bad" prediction. This logic has been shown to reduce the nuisance alert rate by about 33%.[5,6]

Performance Measurement Overview

To calculate the TPMs, the URET application software is run with a set of recorded data (flight plans and flight tracks from the NAS Host Computer, also called "scenario data"), forecast weather data, and adaptation data. Output data, such as trajectories and predicted minimum separation, are then collected. Conflict and trajectory records are used to compute conflict warning times, which are used in the calculation of alert rates. Track and trajectory records are used to calculate position error. Position error is modeled and yields vertical prediction accuracy. Based on the error models, the probabilities of predicted problems are computed for a specified actual minimum separation distance, look-ahead time, and navigation equipage category. These are used to compute the alert rates for specified actual minimum separation distances.

The first level Functional Performance Assessment metrics (lateral, longitudinal and vertical error) are calculated as a function of prediction look-ahead time (the difference between the selected track point's time and the trajectory modeling start time in minutes) and, for vertical accuracy only, the vertical flight phase (in particular climb/descent). Secondary factors such as atmospheric conditions (forecast wind speed and direction, temperature and pressure), the ARTCC and its unique procedures, time of day, season and traffic loading are confounding factors that may influence the metrics.

Sample scenarios are often chosen from peak-level activity during a typical day. "Typical" means that:

- Preferential Routes are configured normally
- No weather-induced Severe Weather Avoidance Programs (SWAP) routes are in effect
- There are few or no ground delays.

In practice, almost all of the flights with trajectories during the sample are used in computing the metrics. However, identification of "outlier" flights that may distort metric values is a necessary preliminary step to gaining statistical insights from a sample. Outlier flights reflect: known URET design limitations, such as holds without clearances or military training routes; software errors; data errors; or unusual anomalies (e.g., bad data in a flight plan). Flights are excluded from the track-trajectory predictability computations when they have known discrepancies that could distort the nominal track-trajectory error statistics. Note that URET still processes such flights, though they are excluded from the functional performance assessment.

Engineering judgement is thus used to determine when to exclude a flight. Normally, a flight is excluded if it has large ($\geq$ 50 nm) lateral or longitudinal PE, or large ($\geq$ 10,000 ft) vertical PE and satisfies one or more of the following rules:

- Flights that executed a hold or operated in a military operating area
- Flights on a military training route
- Flights with an adaptation error (e.g., a fix that is incorrectly adapted hundreds of miles away from route)
- Flights affected by software errors that are being corrected in a System Discrepancy Report

Warning Time Statistics

Conflict warning time is the time between notification of a conflict and its predicted start time. Operationally, it characterizes the time a controller has to assess an alert and take corrective action, if necessary, before an aircraft pair loses separation or comes within alert tolerances. This is expressed as a mean and standard deviation for conflicts that are immediately notified and for conflicts with delayed notification.

Alert Rates and Vertical Accuracy

The concept of **alert rates** is used to characterize the accuracy of the conflict probe. We define a *False Alert* as an aircraft encounter with an actual minimum separation greater than 5.0 nautical miles for which a conflict is predicted, since the separation criteria in en route airspace is 5.0 nm. Experience has shown that such alerts are not necessarily operational nuisance alerts.[‡] URET's definition of a conflict includes both horizontal and vertical positions; however, vertical prediction error is not considered in the measurement process for False Alerts. Vertical prediction accuracy during climb and descent are considered in parallel, as discussed below. Horizontal TPMs are parameterized on the minimum horizontal separation, while Vertical TPMs are parameterized on the lookahead time, measured in minutes. A combination of both TPMs into a two-parameter system, using both horizontal and vertical accuracy data, would be complex, yet potentially useful.

False Alert probabilities are computed for a range of actual minimum separation distances between 5.01 and 20 nm. Probabilities are also given for all the subset of those encounters for which URET predicts **red conflicts**, i.e., conflicts with an actual minimum separation distance between trajectory center lines is less than or equal to 5.0 nm.

A *Missed Alert* is an aircraft encounter with an actual minimum separation less than or equal to 5.0 nm for which no conflict is predicted at a **threshold time** prior to the time of minimum separation. Thresholds of one minute and five minutes have been used in our performance analyses. As with the False Alert rate, vertical prediction error is not considered by this metric. Missed Alert rates are computed for **all** predicted conflicts for threshold times of one and five minutes.

In the vertical dimension, thresholds of 500 and 1500 ft are used in URET performance analyses. The probability that an aircraft is within the threshold value of its predicted altitude is computed for specific look-ahead times. This computation uses the mean and the standard deviation from the track-trajectory deviation data that are computed for climb and descent flight phase trajectories at each minute of look-ahead time from 0 to 20 minutes. Coefficients of a polynomial fit to the empirically derived means and standard deviations are computed for climb and descent flight phases. Vertical accuracy, modeled by a normal distribution, is computed for each specified look-ahead time and altitude tolerance.

[‡] In order for the probe to consistently notify when the closest point of approach is just under the 5 nm limit, the probability of notifying an alert for barely wider separations must be fairly large. This at least will be true when there is error in both the aircraft's track itself and in the modeled trajectories of some of the aircraft.

In summary, we are measuring a set of probabilities in the horizontal dimension (alert rates) and in the vertical dimension (climb and descent phase vertical prediction accuracy). The next section describes the mathematical models that transform empirical track-trajectory deviations and predicted conflicts into a set of alert rate functional performance metrics. The vertical prediction accuracy method is then outlined, and results of both sets of TPM calculations are given.

Functional Performance Evaluation Process

Empirical Position Error Modeling

The recorded air traffic data used in the functional performance assessment cannot be used directly to assess URET performance, since there are generally no real "conflicts" where aircraft violate the separation criteria that URET uses to generate an alert. As there are insufficient conflicts to assess, mathematical models representing the observed deviations between URET-derived trajectories and actual flight paths are created. These models are then used to determine the probability distributions of missed and false alerts. Therefore, the missed and false alert rates are primarily a function of the trajectory modeling accuracy. The computation of TPMs requires summarizing and modeling the deviations between the aircraft tracks and modeled trajectories in a scenario in order to capture the magnitude of the deviations as a function of time.

Trajectories are based on flight plans and updated based on amendments, interim altitudes, track reports, and reconformance logic. Each track report (x, y, altitude, time) is associated with exactly one flight plan trajectory. The track report may also be associated with one or more reconformance trajectories. Let L be the number of minutes from the time a trajectory was generated to a later event of interest. Generally, we use integer values of L and associate non-integer durations with the nearest L. Software tools create a PE file that contains the lateral and longitudinal track-trajectory differences at L = 0, 1, 2, … 40, for each trajectory that a track position is associated with.[2,4]

Position error records also include the equipage indicator, the value of L and the number of reconformances R since the last clearance trajectory in the lateral and longitudinal dimensions. The error measurements contained in the PE file are the following:

- *Lateral Deviation* The signed horizontal distance between the track reported position and its projection onto the associated trajectory segment. A negative value indicates that the track is on the left-hand side of the segment.

- *Longitudinal Deviation* The signed along-route distance between the track projection and the trajectory-estimated position. A negative value indicates that the track is behind the trajectory estimate.
- *Vertical Deviation* The signed vertical distance between the track reported altitude and the trajectory-estimated altitude. A negative value indicates that the track is below the trajectory estimate.

Let E be the subset of PE records used in the analysis. For a track report to be used for such measurements, it must satisfy all of the following criteria:

- A trajectory exists and continuous, reasonable track data is being received[†]
- The report contains valid position and altitude data
- The report is within a parameter distance (50 nm) from the route
- The report is not from an aircraft in "vertical drift", i.e., the flight is in cruise phase, but the reported altitude is more than a parameter distance away from the modeled altitude
- The aircraft is under control of an ARTCC

The URET-derived lateral, longitudinal and vertical standard deviation data are computed from this subset E of the PE records as follows. A subset $A_F \subset E$ has an Area Navigation (RNAV) equipage indicator and a subset $A_X \subset E$ is non-RNAV. Flights with Area Navigation are capable of more accurately adhering to their filed flight plan, and they are modeled as doing so in URET. Finally, the three reconformance states are defined as follows:

- 0 "Zero" reconformance state is defined as when both the track and trajectory have zero reconformances since the last clearance trajectory (R=0).
- 1 "One" reconformance state is defined as when both the track and trajectory have one prior reconformance since the last clearance trajectory (R = 1).
- 2+ "More than one" reconformance state is defined as when both the track and trajectory have more than one reconformance since the last clearance trajectory (R ≥ 2).

[‡] Tracks in category A can be downgraded to category B, D or F. Category B indicates that a trajectory exists and initial, intermittent or frequently unreasonable track data is being received. Category D identifies aircraft in hold status. Category F indicates the aircraft is under the control of another ARTCC.

To summarize, the components of the PE records "E" used for track-trajectory deviation measurements are:

$$E = A_X \cup A_F$$
$$A_X = A_X^0 \cup A_X^1 \cup A_X^{2+}$$
$$A_F = A_F^0 \cup A_F^1 \cup A_F^{2+}$$

Subsets A_F and A_X are further broken down into components corresponding to the time elapsed from trajectory creation L = 0, 1, 2,…, 40 minutes. Averages are computed using the percentage of flights with and without-RNAV equipage, and the percentages of records in the 0, 1 or 2+ reconformance states. The partitioning by the three reconformance states is done since these states directly affect the application of URET's notification logic.

The average root mean square (RMS) error is calculated for each subset and is used to calculate the probability of a predicted problem. However, to eliminate some of the error due to the use of empirical data, the data is fit by a third order polynomial. A weighted least-square methodology is used unless the error is determined to be systematic. If the weighted third order polynomial has a systematic error, an unweighted third order polynomial is used. If both the weighted and unweighted third order polynomials have systematic errors, a second order polynomial is used (weighted first, unweighted second). The weight is set to an estimate for the resolution of the track position divided by the square root of the number of sample points. Figure 2 illustrates examples of how lateral and longitudinal track-trajectory deviations, for RNAV-equipped flights, are modeled by polynomials.

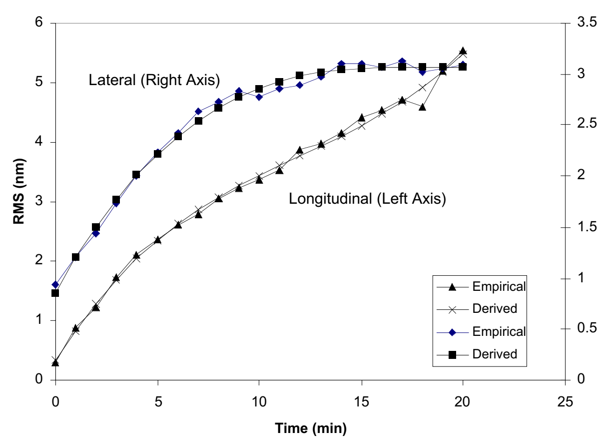

Fig. 2 Lateral and Longitudinal Deviations

Computing Probability of Predicted Problems

Once the position error has been characterized for the scenario, the next step is to compute the probability of a predicted problem for a specified actual minimum separation distance, look-ahead time and navigational equipage category. Input parameters specify the range of look-ahead times, actual minimum separation distances, navigational equipage, and coefficients of the derived track-trajectory deviation polynomial functions. Also specified are the initial position distributions based on track-trajectory deviation statistics.

In order to calculate the False and Missed Alert TPMs, a multi-stage process is used to arrive at four "Matrices":

1. False Alerts, Red problems, RNAV equipage
2. Missed Alerts, all problems, RNAV equipage
3. False Alerts, Red problems, non-RNAV equipage
4. Missed Alerts, all problems, non-RNAV equipage

(Typically, the Missed Alert rate is calculated for Red problems as well, as is the False Alert rate for all problems, but these metrics tend to be less useful than the ones we discuss here.) Each Matrix contains the probability of an alert p_{ij} for a given actual minimum separation distance i at a given time to minimum separation j. The computed RNAV and non-RNAV probabilities are averaged into a single probability by weighting each by the percentage of RNAV and non-RNAV flights in the scenario, respectively.

The cumulative probability *cumprob* is equal to **C1 + C2 + C3** computed in Algorithm 1. Probabilities for *all* alerts are given by *cumprob* averaged over the number of RNAV and non-RNAV cases; and for *red* alerts are given by the average of **C4** over the number of cases.

For each actual minimum separation distance, the probability of an alert at each look-ahead time is averaged based on the percentage of flights reaching notification at that look-ahead time.

Vertical Prediction Accuracy

Vertical accuracy is divided by climb and descent phase. Each is sampled at discrete look-ahead times of one minute up to a 20-minute look-ahead time. The probability of predicting that a reported altitude will be within a tolerance (500 or 1500 feet) of the trajectory altitude in each phase is computed through the following steps.

1. The URET derived mean and standard deviation of vertical position error are computed. These are used to compute coefficients for polynomials that estimate the means and standard deviations as functions of time.
2. A weighted least-squares method is used to compute the coefficients.
3. The lowest degree polynomial between orders 3 and 5 that satisfies a) a test of statistical significance and b) the squared correlation coefficient > .90 is used to fit the means of the data across look-ahead times. The polynomial is selected the same way for the standard deviations.
4. Step 3 is repeated for climb and descent data separately.
5. The polynomials are used to compute the estimated mean and standard deviation of a normal distribution at a look-ahead time. The probability that the predicted trajectory altitude would be within a tolerance (500 or 1500 feet) of a track altitude is computed at look-ahead times between 1 and 20 minutes.

Step 1 is accomplished by using the PE file described in the previous section. Again, this file includes the track-trajectory difference at each minute of look-ahead time for each trajectory that the track position is associated with. It also includes the flight phase, and the number of reconformances since the last flight plan trajectory.

Conflict Warning Time

The conflict model identifies the unique problems that reach notification time and those that are removed before reaching notification time. Predicted problems are categorized as "immediately notified" if the first occurrence of the problem has a notification time that is the same as the probe time; otherwise predicted problems are categorized as "not immediately notified." This process is described in Brudnicki.[1] Once the data is sorted, the mean and standard deviation of conflict warning time is computed for:

- All unique problems that reach notification time
- Red unique problems that reach notification time
- All unique problems that reach notification time and are not immediately notified
- Red unique problems that reach notification time and are not immediately notified.

Application to URET

Each operational release of the URET software has undergone a Functional Performance Assessment. Generally, two operational air traffic scenarios have been used in each assessment to determine the effects of different data samples. Each scenario contains five hours of data for over 5,000 flights. Commercial,

Algorithm 1 - Probability of Predicted Problems

$l1$ = longitudinal conformance bound (nm)
$l2$ = lateral conformance bound (nm)

Loop over times to minimum separation *tsep* (20, 19, …,1 for false alerts; 9,8,…,1 for missed alerts)

Loop over minimum separation distance (5, 6…20 or 0, 1, … 5)

Loop over encounter angles (fix subject aircraft angle ϕ_1 at 90, object aircraft angle varies). Each encounter angle is weighted equally (weight = 1/6) when averaged in the probability calculation.

Object aircraft heading:

$$\phi_2 = 120{,}150{,}180{,}210{,}240{,}270 \text{ (head-on)}$$

Absolute difference between subject and object aircraft headings:

$$\phi = |\phi_1 - \phi_2|$$

Aircraft velocities: $v_1 = v_2 = 480kt$

For each encounter angle ϕ calculate the "miss criterion", or "tolerance". This is given by the formula for *seph* (horizontal separation), below.

Ratio of velocities:

$r = \min(v_1 / v_2, v_2 / v_1)$ = 1 when $v_1 = v_2$.

r, ϕ determine the formula (A, B or C) for calculating seph.

A. $$(\cos\phi < r) \Rightarrow seph = 5 + \frac{(1+r)*(l1*\sin(\phi) + l2*(1-\cos(\phi)))}{\sqrt{1-2*r*\cos(\phi)+r^2}}$$

B. $$(\cos\phi \ge r) \Rightarrow seph = 5 + \frac{(1+r)*l1*\sin(\phi) + \sqrt{(1-r)^2 * l2 * (1-\cos(\phi))}}{\sqrt{1-2*r*\cos(\phi)+r^2}}$$

C. $$(\phi_2 = 270) \Rightarrow seph = 5 + 2*l2$$

Determine initial aircraft positions x_i, y_i and velocities $\dot{x}_i, \dot{y}_i$ for desired msep and object aircraft angle

$$\dot{x}_2 = v_2 * \sin(\phi_2) \qquad \dot{y}_2 = v_2 * \cos(\phi_2)$$
$$x_1 = -v_1 * tsep \qquad y_1 = 0$$

If encounter is head-on,

$x_2 = -x_1$ $\qquad$ $y_2 = -$(minimum separation distance)

else $\quad x_2 = -tsep * \dot{x}_2 \qquad y_2 = -tsep * \dot{y}_2$

Calculate the time to minimum separation t_{min} as follows:

Encounter is head-on $\Rightarrow t_{\min} = \dfrac{x_2}{v_2}$

Else Calculate t_{min} by a subroutine "tracing" the relative paths until the difference between the minimum separation and the desired minimum separation is < .05 nm.

Fit polynomials for RMS error of track trajectory deviations (using weighted least squares) for the RNAV and non-RNAV data separately.

Algorithm 1, continued

Use the time to minimum separation, t_{min}, to calculate the longitudinal and lateral deviations.

$\sigma_{long} = P_{long}(t_{\min})$, where P is the RNAV or non-RNAV polynomial

$\sigma_{lat} = P_{lat}(t_{\min})$, where P is the RNAV or non-RNAV polynomial

Compute position of subject and object aircraft "boxes" of the dimensions ± three standard deviations in each dimension. Divide each box into cells of dimension $.25*\sigma_{lat}$ by $.25*\sigma_{long}$.

Determine the center of each cell and compute the relative minimum separation distance (msepr)

$$\min dx = (subx + t_{\min} * \dot{x}_1) - (objx + t_{\min} * \dot{x}_2)$$
$$\min dy = (suby + t_{\min} * \dot{y}_1) - (objy + t_{\min} * \dot{y}_2)$$
$$msepr = \sqrt{\min dx^2 + \min dy^2}$$

… between each subject cell and all object cells for each subject cell. Initialize C1, …, C4 to 0. Add the product of the subject cell's likelihood and the object cell's likelihood to C1, C2, or C3 if the separation between cell centers is less than *seph*. Add the result to:

C1, if $\dfrac{msepr}{seph} \le .333$

C2, if $.333 < \dfrac{msepr}{seph} \le .667$

C3, if $.667 < \dfrac{msepr}{seph} \le 1.0$

and also to C4, if $msepr \le 5$ nm.

In the diagram in Figure 3, the closest "cells" representing possible locations of the subject and object aircraft, satisfy msepr < seph, and are used in the probability calculation. The values of C3 and C4, though not illustrated in the figure, are determined the same way as C1 and C2. This figure simplifies the 12 x 12 cell grid of both the modeled subject and object flight.

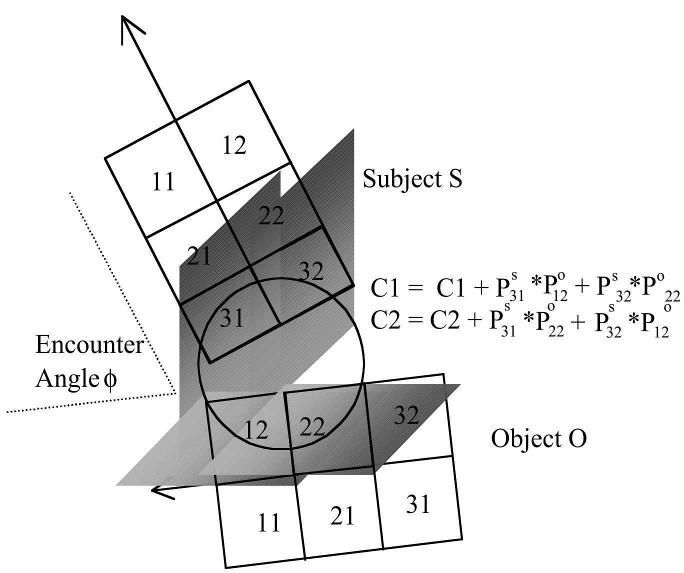

Fig. 3 Simplified Grid Model

military and General Aviation flights are represented. The scenarios consist of Host Computer flight plan, amendment, interim altitude and track messages.

Flights with more accurate navigation equipment follow their flight paths more closely than those without. URET conformance bounds are tighter for aircraft with more accurate navigation equipment. This impacts the alerts that URET notifies and consequently the results of some of the analyses performed as part of the Functional Performance Assessment. Therefore, the navigation equipment indicator is mapped to one of two categories: RNAV (the more accurate category) and non-RNAV.

Sufficient track data are required for flights in order for them to be used in computing TPMs. The scenarios that are used include flights at various stages along their routes. As a result, some flights in the scenarios have insufficient track data with which to compute the TPMs and are therefore not used.

Table 1 presents the conflict warning time attributes for all (and all red) notified problems for the two scenarios used in the functional performance[3] of URET. Figure 4 and Tables 2 and 3 present false and missed alert rates for the same scenarios and software version. This version has been in operational use at ZID and ZME since the latter part of 2000. The false alert rate at 5.01 nm is approximately 50%; it decreases monotonically to the 20 nm separation level, and is about 1% at 15 nm. The missed alert rates are shown for both the one-minute and five-minute thresholds. As can be seen from Tables 2 and 3, these rates increase from the 0 nm separation level to the 5 nm separation level.

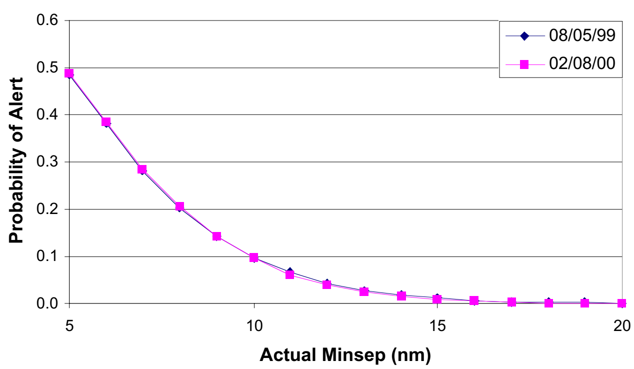

Fig. 4 False Alert Rates - Curves Nearly Coincident

Attribute	5 August 1999		8 February 2000	
	Red	All	Red	All
Mean Warning Time (min)	10.16	8.95	10.19	9.07
Standard Deviation (min)	4.84	4.54	5.08	4.71

Table 1 Warning Time - All Notified Conflicts

Actual Minimum Separation (nm)	All Predicted Problems	
	5 August 1999	8 February 2000
0	3.92×10^{-3}	5.61×10^{-3}
1	5.06×10^{-3}	6.93×10^{-3}
2	8.93×10^{-3}	1.18×10^{-2}
3	1.69×10^{-2}	2.16×10^{-2}
4	3.16×10^{-2}	3.82×10^{-2}
5	5.61×10^{-2}	6.56×10^{-2}

Table 2 Missed Alert Rates - 5 Minute Threshold

Note that the missed alert rates with the one-minute threshold are significantly lower than those for the five-minute threshold.

Actual Minimum Separation (nm)	All Predicted Problems	
	5 August 1999	8 February 2000
0	1.34×10^{-4}	2.98×10^{-4}
1	2.41×10^{-4}	4.97×10^{-4}
2	7.63×10^{-4}	1.22×10^{-3}
3	2.10×10^{-3}	3.21×10^{-3}
4	5.51×10^{-3}	7.84×10^{-3}
5	1.40×10^{-2}	1.88×10^{-2}

Table 3 Missed Alert Rates - 1 Minute Threshold

Figures 5 and 6 present URET vertical prediction accuracy during climb and descent for same scenarios. Insufficient data samples were available to calculate vertical prediction accuracy past the 18-minute and 15-minute look-ahead times, respectively.

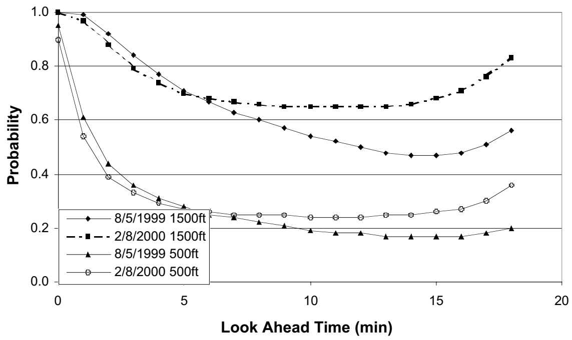

Fig. 5 Vertical Prediction Accuracy During Climb

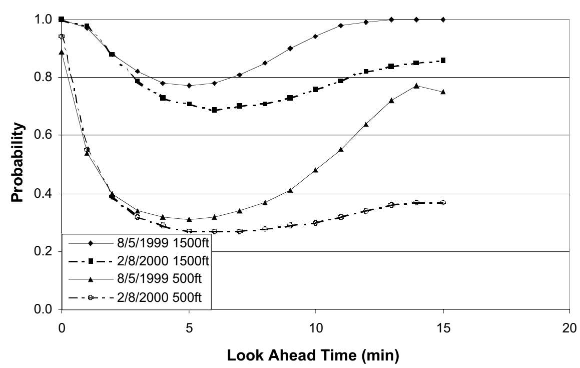

Fig. 6 Vertical Prediction Accuracy During Descent

Interpreting the Results

One of the trade-offs in interpreting the functional performance results of an operational capability such as URET is the dichotomy between missed alerts and false alerts. Ideally, missed alert rates should go to zero as soon as the actual miss distance (minimum separation) falls below the desired threshold (in this case the ATC separation standard of 5.0 nm). Likewise, the false alert rate should ideally go to zero as soon as the actual miss distance reaches above the threshold. This is easily achieved if one were to base the assessment of missed or false alert at the *time* of actual minimum miss distance, but this is not practical for a strategic conflict detection function that has to alert a controller *ahead of time*, with sufficient time to act. Consequently, if plotted as a function of actual miss distance, the probability of an alert would ideally resemble a discontinuous function at the ATC separation standard (with the probability of alerting the controller equal to 1.00 for miss distances below the threshold, and a probability of 0.00 otherwise.) In practice, however, that is not the case because of the uncertainties present in predicting aircraft positions, particularly at the long look-ahead times used by URET. Essentially, those uncertainties determine the functional performance of URET. Reduce the uncertainty and the resultant alert probability curve should converge toward the ideal. Realistically, that convergence stops at some point, and the distribution reaches some limit because of the uncertainties inherent in the NAS (wind forecasts, aircraft performance, route and amendment uncertainties, NAS surveillance performance, etc). That limit is consistent with the expectation that as the miss distance gets smaller, the probability of an alert is greater.

The approach taken to date is not necessarily to compare the resultant probability distribution to the ideal, but rather concentrate on: (A) the operational acceptability of URET and (B) the differences in the probability curves for different versions of URET.

Measuring Operational Utility

Although the URET TPMs provide a good method for assessing data and software modifications and for deriving system requirements, they do not necessarily translate into metrics for assessing operational acceptability. Indeed, a capability like URET must undergo qualitative and subjective analyses and evaluations to determine its operational utility. One of the goals of such evaluations is to determine the relationship (if any) between *quantitative* TPMs and *qualitative* operational acceptability. Once having achieved operational acceptability, the corresponding functional performance can be used to some extent as a means of quantifying that acceptability and as an operational benchmark.

To date, operational utility has been based on formal evaluations of URET in daily operations[7,8]. Based on feedback received from air traffic controllers during these evaluations, URET's operational utility has been widely accepted. As changes are made to the URET design, corresponding operational evaluations and functional performance analyses will both continue to be used to assess the qualitative and quantitative performance, respectively.

Conclusions

The Functional Performance Evaluation results have had an impact on URET's development, testing and acceptance as a prototype. Scenario data's influence in determining the outcomes of the evaluation is inescapable. The type of traffic, the distribution of flights and aircraft types, the time of day and especially the amount of controller clearance data entered into the Host Computer System all influence the outcomes. Such dependence is in fact a strength of this methodology. It allows us to use the TPMs to uncover scenario characteristics that have different impacts upon the conflict probe's accuracy. Without this scenario-dependent aspect of Functional Performance Assessment, probability distributions of different input parameters would have to be subjectively chosen; such choices could be inaccurate. This methodology, we believe, does not rely on many arbitrary choices to produce its results. Confidence in the evaluation process can be nearly as important as the validity of the evaluation results themselves.

Next, the process provides both relative and absolute information. For a given configuration of software and adaptation data for that software, a certain duration of traffic data for a specific center (ARTCC) or set of centers, and a certain quantity of excluded flights, PE in the various dimensions can be modeled. Then the TPMs (alert rate, vertical prediction accuracy, and warning times) can be generated. *Relative* to those TPMs, an identical sample, only run with a different software configuration, can be compared. This process has reinforced the decisions made to gradually modify or improve the trajectory subsystem, track subsystem or URET databases by showing the effect on performance results. An example would be improved modeling of aircraft climb rates and its impact on vertical prediction accuracy.

URET's extensive *operational* testing provides us with a valuable context within which to view the TPM results produced by functional performance analyses. The subjective acceptance of its use by controllers at Indianapolis and Memphis, the use of trial plans, the completeness of intent information entered into the Host and received by URET and other indications help us to measure the operational utility of the system. The TPMs do not measure the operational utility of URET's alert information, yet they are relevant. As we continue to investigate controller utility, we expect to see that the "absolute" (in the above sense) measurements made in the lab continue to correlate to the subjective, but nevertheless crucial, operational performance levels.

References

1. Brudnicki, D. J., Arthur, W. C., and Lindsay, K. S., "URET Scenario-based Functional Performance Requirements Document", MTR98W000044, The MITRE Corporation, April 1998.
2. Lindsay, K. S., "Algorithmic Performance Analysis for Deliveries D3A and D3.1 of the User Request Evaluation Tool", WN99W0000008, The MITRE Corporation, March 1999.
3. Rozen, N.R., and Lindsay, K. S., "Functional Performance of the User Request Evaluation Tool Delivery 3.3, Release 1", MTR00W0000025, The MITRE Corporation, April 2000.
4. Lindsay, K. S., "XEVAL Users Manual", MTR 97W000030, MITRE Corporation, January 1998.
5. Kirk, D. B., and Ricker, M. L., "User Request Evaluation Tool (URET) D3 Automated Problem Detection Algorithmic Definition", MTR 96W0000038R1, MITRE Corporation, December 1997.
6. Lindsay, K. S., "Results of a URET Operational Utility Experiment", WN99W0000081, The MITRE Corporation, January 2000.
7. Bowen, K. C., "Evaluation Report URET Delivery 3.1, Two-Way Host Interface", WN99W0000043, The MITRE Corporation, July 1999.
8. Brudnicki, D. J. and Bowen, K. C., "Evaluation Report URET Delivery 3.1, One-Way Host Interface Field Evaluations at Indianapolis and Memphis ARTCC", WN99W0000005, The MITRE Corporation, February 1999.