

# **User Request Evaluation Tool (URET) Interfacility Conflict Probe Performance Assessment**

**William C. Arthur  
Dr. Michael P. McLaughlin**

**The MITRE Corporation  
Center for Advanced Aviation System Development  
1820 Dolley Madison Boulevard  
McLean, Virginia 22102  
Phone/fax/e mail  
703-883-5971/703-883-3348/warthur@mitre.org  
703-883-7658/703-883-1911/mpmcl@mitre.org**

**© 1997 The MITRE Corporation**

This is the copyright work of The MITRE Corporation, and was produced for the U. S. Government under Contract Number DTFA01-93-C-00001, and is subject to Federal Acquisition Regulation Clause 52.227-14, Rights in Data--General, Alt. III (JUN 1987) and Alt. IV (JUN 1987). No other use other than that granted to the U. S. Government, or to those acting on behalf of the U. S. Government, under that Clause is authorized without the express written permission of The MITRE Corporation. For further information, please contact the The MITRE Corporation, Contracts Office, 1820 Dolley Madison Blvd., McLean, VA 22102, (703) 883-6000.

The contents of this material reflect the views of the authors. Neither the Federal Aviation Administration nor the Department of Transportation makes any warranty or guarantee, or promise, expressed or implied, concerning the content or accuracy of the views expressed herein.

## Abstract

The Federal Aviation Administration (FAA) and The MITRE Corporation's Center for Advanced Aviation System Development (CAASD) have developed a set of en route decision support capabilities known as the User Request Evaluation Tool (URET) prototype. These capabilities include a flight plan based conflict probe that continuously checks for strategic conflicts, a trial planning function that allows a controller to evaluate resolutions or pilot requests before they are issued as clearances, and electronic flight data displays that provide an effective means of managing flight information at the sector. The URET prototype is the basis for the Free-Flight Phase 1 (FFP1) conflict probe that will be deployed for large-scale national trials. The prototype has been in daily use at Radar Associate sector positions since 1997 in the Indianapolis (ZID) and Memphis (ZME) Air Route Traffic Control Centers (ARTCCs). Daily use is defined as controllers using URET on a regular scheduled basis (currently 12 hours per day, 5 days per week). Beginning in 1998, an interfacility (IFA) capability that provides data exchange between multiple ARTCCs was added to the daily use system. IFA operation allows controllers to have more accurate and timely trajectory and conflict data beyond the center boundary, and provides a capability for automated coordination of trial plans between centers.

This paper describes quantitative metrics and models that can be used to determine FFP1 conflict probe accuracy requirements as well as to evaluate the prototype system upgrades. The metrics herein are technical performance measures to evaluate functional performance. Operational performance metrics that are based on controller assessments are outside the scope of this paper. However, the operational consensus that has emerged from the URET field evaluations to date is that the URET information is perceived to be "highly accurate" in the context of (1) the sector operations that it supports and (2) its relative value as compared to the current automation system capabilities in supporting those same operations. Technical performance measures, however quantified, also need to be interpreted in that context to ensure analytically complete and consistent

conclusions with respect to the net contribution of the tools to operational performance. This paper presents the results of the particular methodology currently being used at CAASD to quantify technical performance of the URET conflict probe. Technical performance is shown to vary over a range determined by a number of factors in the ATC operational environment, but the experience to date has been that the tolerance on suitable operational performance is consistent with the technical performance that has been achievable. Any specific instances of performance outside the tolerances for operational suitability are pursued for the purpose of understanding and correcting such variances. Work on consistent correlation and transformation of technical performance metrics to operational performance metrics is in progress, and is also reported in this paper. In the interim, results of operational performance evaluations to date are separately reported in this ATM-98 seminar [1] and the 1997 US/Europe ATM R&D Conference [2].

Two complementary approaches to technical performance analysis are described in this paper. One approach uses performance metrics derived from recorded real-world scenarios. These scenarios include controller actions to maintain separation, thus models are used to estimate accuracy metrics. Initial results were presented at the 1997 US/Europe ATM R&D Conference [2]. The second approach, an En Route Conflict Probe Testbed, provides quantitative measures for the URET system with synthesized scenarios in which there is no controller intervention. In this approach, track data is synthetically generated from stochastic models of air traffic and track reports can have a loss of FAA en route separation minima.

## Introduction

### Background

The FAA has developed new decision support systems (DSSs) that assist controllers to better meet the needs of the airspace users. The URET is a prototype DSS developed for the en route sector team that includes the following:

- Host Computer System Interface
- Trajectory Modeling
- Conformance Monitoring and Reconformance
- Current Plan and Trial Plan processing
- Automated Problem Detection (APD)
- Computer-Human Interface
- Interfacility Communication

URET identifies potential aircraft and airspace separation problems up to 20 minutes in advance, and provides a computer human interface to allow effective exchange of information between controller and the DSS.

URET field trials began in 1996, starting as a controller initiated conflict probe and quickly progressing to one which continuously checks for conflicts. Several major system upgrades have been evaluated and implemented to incorporate controller feedback. URET daily use began in 1997 in the Indianapolis (ZID) and Memphis (ZME) Air Route Traffic Control Centers (ARTCCs). Daily use is defined as controllers using URET on a regular scheduled basis (currently 12 hours per day, 5 days per week). Interfacility (IFA) URET began daily use operations in May 1998.

The URET conflict probe is the prototype for the Free-Flight Phase 1 (FFP1) conflict probe. FFP1 embodies a core set of capabilities from several types of applications that are to be deployed and operationally integrated at a limited number of sites in the 2000-2002 timeframe. The purpose is to begin achieving near-term benefits and to reduce the risk of future national deployment. FFP1 is the first step in the evolutionary development and deployment of

functionality that falls under the general heading of NAS Modernization capabilities.

### Interfacility URET

Interfacility URET is an extension of URET capabilities to provide continuity of service for a parameter distance (200 nm) outside an ARTCC's boundary. Each URET system stores data for its own airspace and for neighboring facilities so that trajectory modeling and conflict detection are consistent in each facility. Real-time events such as flight data, position data, reconformance data, and status information are exchanged between URET processors. The URET processor in each center computes trajectories and predicted problems for flights in its own center and flights within the parameter distance of the center boundary. Effectively, the accuracy from a single site is extended across facility boundaries.

Initial IFA field evaluations confirmed the operational acceptability [3]. Controllers found the trajectories and conflict probe to be operationally accurate. However, controllers also identified system discrepancies that were resolved in subsequent versions of the system. Conflict notification rules were modified to include notification to more than one sector near center boundaries. Trajectory modeling was corrected to better ensure consistency and continuity of information across facility boundaries.

In addition to field evaluations, system upgrades are evaluated for technical performance that characterizes the system accuracy and for areas of improvements. Two complementary approaches are described below: actual recorded traffic performance measures and synthetic (testbed) traffic performance measures.

### Recorded Scenario Performance

Recorded scenario functional performance metrics are based on actual air traffic scenarios (track and clearances) that are replayed through URET. These metrics characterize the system accuracy and provide a basis for comparing different URET versions. Recorded air traffic scenarios include the effects of clearances issued by controllers.

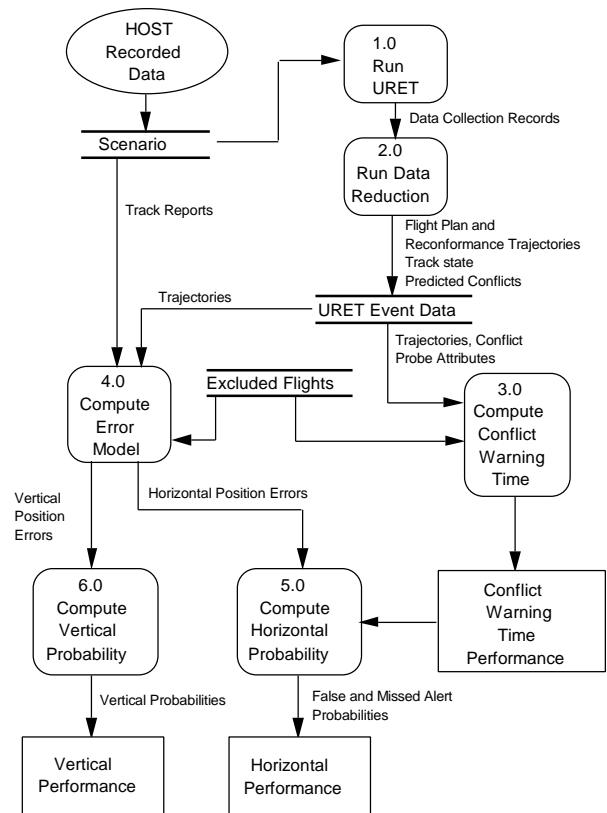
The technical performance measures (TPM) used in this section to describe functional performance of the URET conflict probe are as follows:

- *False alert rate (horizontal):* Aircraft encounter with an actual minimum separation greater than 5 nm for which a conflict is predicted. Vertical prediction error is not considered by this metric.
- *Missed alert rate (horizontal):* Aircraft encounter with an actual minimum separation less than or equal to 5 nm for which no conflict is predicted at a threshold time prior to the time of minimum separation. Vertical prediction error is not considered by this metric.
- *Vertical prediction error:* The probability of predicting an altitude within a given tolerance of the actual aircraft altitude during vertical transition at specified lookahead times.
- *Conflict warning time:* The time between Conflict Notification Time (time when a predicted conflict is displayable to a controller) and Conflict Start Time for a given predicted conflict.

### Scenario-Based Performance Model

Scenarios are extracted from the Host computer system tapes that contain actual traffic data (e.g., clearance messages, track data). Since a recorded air traffic scenario includes controller actions to maintain separation, mathematical models were developed to transform empirical measurements into the set of metrics described above.

The measurement and analysis process is divided into six steps (figure 1), described next.



**Figure 1. Recorded Scenario Functional Performance Process**

**Step One.** The URET application software runs with the extracted scenario data (e.g., track and flight plan messages) and records its responses (e.g., trajectories, conflicts predicted).

**Step Two.** The data collection files are reduced into URET Event files.

Prior to computing performance statistics, flights that have abnormal processing or known discrepancies that can result in extremely large deviations (outlier points) are identified and excluded. Flights are excluded from the performance results if the track data indicates a hold, the flight is on a military training route, or the flight is subject to an adaptation or software error that creates an outlier deviation. In most scenarios, less than one percent of ZID flights and less than a few percent of ZME flights are excluded.

**Step Three.** Compute Conflict Warning Time derives warning time from the conflict and trajectory

records in the URET Event files. Unique problems that reach notification time (time at which a problem is displayable to a controller) are identified, and the attributes of predicted minimum separation distance and conflict warning time are recorded. Unique problems are categorized as immediately notified and not immediately notified by comparing the time of the first conflict probe record to the notification time. Output from this step is the conflict warning time summary data.

The distribution of predicted conflict warning times are grouped into one-minute bins. This distribution is used in Step 5 as a weight for the false alert average over time for a given actual minimum separation distance.

**Step Four.** The Compute Error Model is divided into three submodels. First, the track-trajectory differences are determined by comparing the track position data in the scenario file to the trajectory-estimated flight plan position in the URET Event file. Deviation is measured as the signed difference in the lateral, longitudinal, and vertical dimensions between a given track reported position and altitude, and its associated trajectory-predicted position and altitude.

Second, horizontal trajectory predictability is computed by partitioning the track-trajectory deviation measurements based on the following attributes:

- Navigational Equipage [area navigation equipage (RNAV) and non-RNAV]
- Look-ahead Time (0 to 40 minutes)
- Prior Reconformances. The cumulative number of reconformances in each dimension since the last Host flight plan update are partitioned into three categories: 0, 1, and 2+ reconformances. The average track-trajectory error is computed from the data samples from each above reconformance state in each dimension.

Vertical trajectory predictability is computed by partitioning the track-trajectory deviation measurements based on the following attributes:

- Look-ahead Time (0 to 40 minutes)
- Flight phase (climbing or descending)
- Prior reconformances

Third, the coefficients of a polynomial function are computed to decrease error associated with the use of empirical data. For lateral and longitudinal dimensions, the RMS error is used. For climbs and descents, both the mean and standard deviation errors are fit by polynomials. A weighted least-square approximation is used.

**Step Five.** Compute Horizontal Probability determines the probability of a predicted problem for a specified actual minimum separation distance, lookahead time, and navigation equipage category. Vertical probability is not included in this function. Parameters specify the range of lookahead times, actual minimum separation distances, navigational equipage, conformance bounds (tolerances around the nominal trajectory position that define where the aircraft can be; 1.5 nm longitudinal; lateral is 2.5 nm RNAV and 3.5 nm non-RNAV), distribution of initial positions, and coefficients to the lateral and longitudinal deviation polynomial functions.

False alert rates for actual minimum separation distances of 5 to 20 nm are computed as weighted averages. For each actual minimum separation distance, the false alert rates are averaged by the fraction of flights with RNAV and non-RNAV equipage indicators, and by the fraction of problems reaching notification time at each lookahead time (from step 3).

A problem is a missed alert if it is not predicted at the last probe before a threshold time. Two threshold times are used: one minute and five minutes. Missed alert rates are computed as one minus the average probability of alert over the times where the last reconformance of either flight in a problem would occur. The average time between reconformances in URET has been approximately five minutes. Thus, for the 5 minute threshold, times 5 through 9 minutes are used. For the 1 minute threshold, times 1 through 5 are used (modeled as uniformly distributed over these times).

**Step Six.** Compute Vertical Probability computes the probability that a track altitude will be within a distance of a flight level altitude at a given lookahead time for trajectory segments that are in a climb or descent phase of flight. The mean and standard deviation from the track-trajectory deviation data are computed for climb and descent flight phase trajectories at each minute of lookahead time from 0 to 40 minutes. Coefficients of a polynomial are determined from the empirically derived mean and standard deviations in each flight phase (from step 4). The probability of predicting an altitude, assuming a normal distribution, is computed for each specified lookahead time and altitude tolerance.

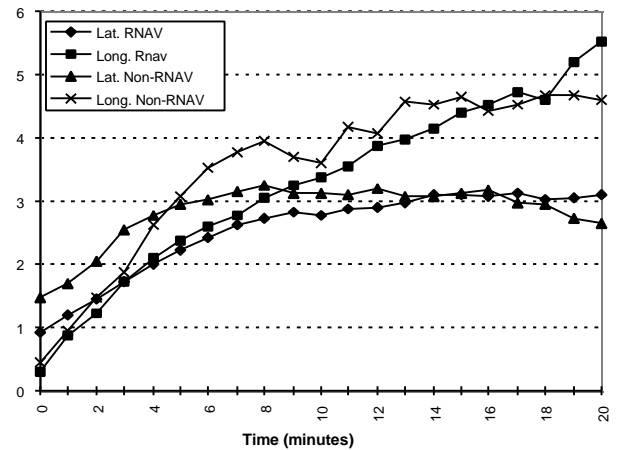
### Recorded Scenario Performance Results

The below results are from a URET Interfacility (IFA) system run for a five hour air traffic scenario from the Indianapolis and Memphis centers on 18 December 1997 [4]. The scenario included over 6,000 flights, and 800,000 track reports, and 18,000 Host messages (e.g., flight plans) as well as voice clearances not recorded as a Host message.

Performance results for the four TPMs are as follows.

#### *False Alert Rate*

The false alert probabilities are specified for a range of actual minimum separation distances of greater than 5 nm to 20 nm. False and missed alert rates are derived from the horizontal track-trajectory errors (figure 2, before smoothing). Note, these errors include the effects of controller voice clearances as well as other factors such as wind and aircraft characteristic uncertainties.



**Figure 2. Horizontal Track-Trajectory RMS Error**

Figure 3 shows the false alert probabilities (rates) for each actual minimum separation distance modeled. The probability of an alert for a specified actual minimum separation distance is given for all predicted conflicts and for the subset of those defined as red conflicts (conflicts with a predicted minimum separation distance less than or equal to 5 nm between trajectory centerlines). This figure also contains an “ideal” probability of a predicted problem for each actual minimum separation distance. The ideal probability curve is derived from the probability of a predicted problem where 1) URET conformance bound APD algorithms are used, and 2) all flights have area-navigation equipment, and are modeled with minimal realistic position error that is normally distributed within the lateral conformance bound and longitudinal error. The ideal probability curve indicates that “better” performance is greater than URET performance for actual minimum separation distances less than the crossover point and better performance is less than URET performance for actual minimum separation distances greater than the crossover point. For example, for an actual minimum separation distance of 7 nm, URET had a probability of alert of about .8 (80 percent alert rate); better performance would be to have a probability of alert greater than .8.

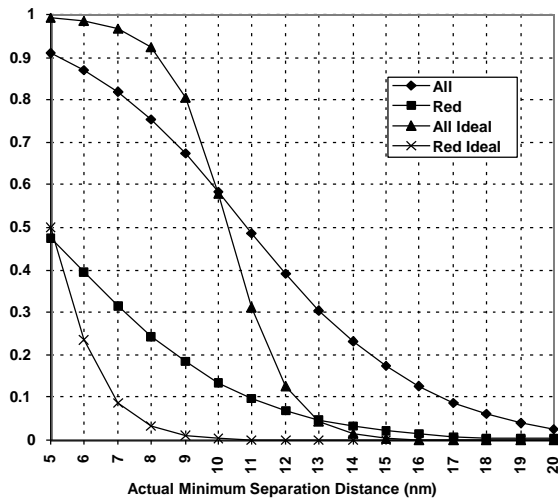


Figure 3. False-Alert Rate

*Missed Alert Rate*

A missed alert is defined as an aircraft encounter with an actual minimum separation distance less than or equal to 5 nm for which no conflict is predicted at a threshold time prior to the time of minimum separation. Vertical prediction error is not considered by this metric.

The TPM for average missed alert rates (figure 4) characterizes results from the recorded scenario data. Recorded scenarios include the effect of voice clearances that increase the lateral errors beyond that of pilotage error; only a small percentage of route and speed clearances are entered as Host messages [2]. Thus, three additional cases are modeled in figure 4 to approximate the average missed alert rate for flights without vectors and with improved (ideal) data (i.e., modeled with lateral deviations inside the conformance bounds and decreasing longitudinal errors).

The probability of an average missed conflict is specified in figure 4 as an average value for actual minimum separation distances between 0 nm and 5 nm (uniformly distributed). Average missed alert rates are specified for two thresholds (1 and 5 minutes), and for all and red problems (red missed alert rate is derived as if only red problems are notified). URET notifies all problems, not just red problems. Better performance is less than or equal to these probabilities.

If the lateral deviation error is modeled within the conformance bounds (since controllers know when a vector is issued), the all problem missed alert rate is approximately half the rate of the recorded scenario rate. As better aircraft characteristic speed data becomes available, the missed alert rate is due more to errors in the forecast winds aloft. Two estimates of longitudinal error growth rates representing different levels of accuracy in the forecast winds aloft are modeled: .25 nm/minute and .12 nm/minute longitudinal error rates.

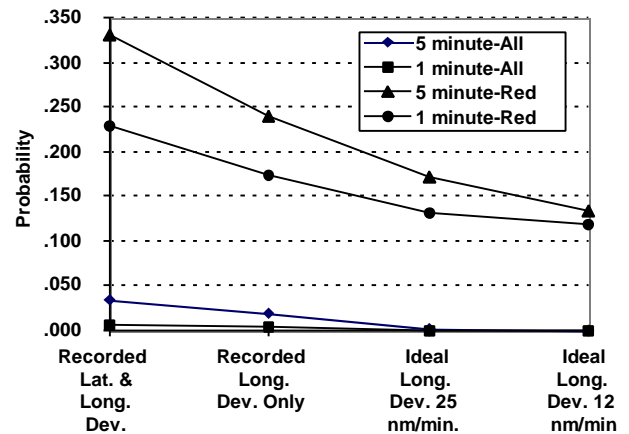
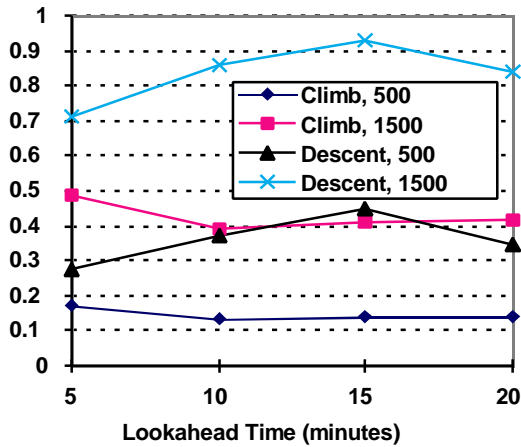


Figure 4. Average Missed-Alert Rate, Recorded and Ideal Horizontal Error Models

*Vertical Prediction Error*

The probability of predicting an altitude within a given tolerance of the actual aircraft altitude during vertical transition is specified in figure 5. Probabilities are given for climbs and descents; for tolerances of 500 and 1,500 feet; and for lookahead times of 5, 10, 15, and 20 minutes. Better performance is for values where the vertical probability of predicting an altitude is greater than or equal to these probabilities.



**Figure 5. Vertical Prediction Rate**

### *Conflict Warning Time*

Conflict Warning Time is the time between Conflict Notification Time and Conflict Start Time for a given predicted conflict.

Conflict warning times are specified for unique problems that reach notification time (table 1). Conflicts that do not reach notification time are not considered here. Values are provided for all unique conflicts and for the subset of those defined as red conflicts. Each of those two categories is again divided, where values are provided for all cases and for the subset of cases for which notification of the conflict is delayed.

**Table 1. Conflict Warning Time**

Warning Time Attribute	All Cases	Delayed Notification Cases
	<b>All Unique Conflicts</b>	
Total Number	1730	859
Mean (minutes)	8.53	11.62
Standard deviation (min.)	4.60	2.79
Percent > 5 minutes	73	96
<b>Red Unique Conflicts</b>		
Total Number	643	382
Mean (minutes)	9.88	12.69
Standard deviation (min.)	5.06	2.88
Percent > 5 minutes	79	98

### **Improvements Using Empirical Data**

URET performance can be improved by more accurate aircraft characteristic data. URET initially used characteristic data derived from performance manuals (37 types); all other types (over 200) were mapped to the climb or descent profiles for types with data.

URET models climb and descent profiles for an aircraft type (e.g., A300). Aircraft type profiles are stored in the aircraft characteristic data file with a speed and a gradient specified between two altitudes for a set of altitude layers. Climb profiles are specified for different temperatures. Trajectory modeling (TJM) uses this data, forecast winds and temperatures, and ATC restrictions to build climb and descent profiles.

Empirical data from track reports was used to modify the aircraft characteristic data resulting in significant improvements in the accuracy of the data. Initial changes to climb and descent gradients have been implemented; further updates to both speed and gradient characteristics are in progress. Updating the aircraft characteristic data provides more accurate and more stable trajectories. For example, initial climb gradient changes resulted in 6 percent fewer false alerts and 4.5 percent fewer reconformances.

The potential improvement that can be achieved from empirical data is illustrated for one aircraft type (table 2). Before modification, the average TJM distance to achieve an altitude is much different than the average empirical distance (table 2, first set). The aircraft characteristic profile gradients are modified to become steeper or shallower using the set one inputs. Next, additional scenarios are run to determine if the updated profile data remains more accurate. In this second scenario set, the differences between updated TJM and empirical data are much smaller. Overall, the updated TJM trajectory is a better predictor for flights with this aircraft type.



**Table 2. Using Empirical Data to Improve Aircraft Characteristic Data for One Aircraft Type, Average Climb Distance to Altitude**

Altitude	First Scenario Set		Second Scenario Set	
	Empirical Distance (nm)	Initial TJM (nm)	Empirical Distance (nm)	Updated TJM (nm)
5000	9.4	5.6	7.9	8.4
10000	17.7	14.8	17.0	17.9
15000	32.1	25.6	32.2	32.3
20000	47.4	39.4	48.1	48.6
25000	68.0	58.9	68.7	67.7
30000	91.1	82.7	93.4	90.8
35000	114.4	113.9	118.0	114.1
40000	142.9	178.3	152.1	142.9

## En Route Conflict Probe Testbed

A complementary approach to performance results from real-world recorded scenarios is one where traffic scenarios are generated synthetically, but in a manner that accurately reproduces those characteristics observed in actual traffic. This capability is referred to, in this paper, as the *testbed* [5, 6].

The testbed serves both to validate the recorded scenario estimates and to provide enhanced functionality. It includes features that characterize the ideal capability for evaluating the *predictions* of a conflict probe. Three features stand out as especially desirable as described next.

First, synthetic traffic scenarios yield perfect knowledge of ground truth by construction. Moreover, in contrast to recorded data, in which conflict probe (CP) predictions are confounded (made unverifiable) by subsequent actions of controllers, it is easy to examine these synthetic tracks to find all the instances of true conflicts, however these might be defined. Finally, the amount and nature of the synthetic input is limited only by the availability of the necessary computer resources. Assuring that these synthetic tracks are indistinguishable, on average, from observed traffic is a straightforward modeling task, with statistical measures that are well-known.

## Testbed Components

A high-level flow diagram of the testbed, using the URET conflict probe as an example, is shown in figure 6. There are three primary (software) components: the Scenario Generator, the Conflict Analysis module, and the Performance Analysis module.

### Scenario Generator

Synthetic traffic scenarios are created using a set of flight plans, plus adaptation and weather (wind) data. Typically, a recorded set of flight plans is taken from field observations and input without modification. This ensures that the scenario, usually about 12 hours of traffic for an entire center, is realistic. The adaptation and wind data are those appropriate to the chosen site and date.

The output of the Scenario Generator is a pair of ASCII files, one with data/format matching that of the probe's input file for a scenario run. Synthetic data may thus be presented to the conflict probe without having to modify the latter in any way. The second file is a "ground-truth" file containing errorless track data for all flights in the scenario.

The Scenario Generator itself comprises three models, partitioned commensurately with the dimensionality of the problem and termed the Lateral, Longitudinal, and Vertical models.

The **lateral model** takes a flight plan and outputs a sequence of X,Y points, in center coordinates, at twelve-second intervals, assuming a nominal speed of 400 knots. The time field is modified and the altitude filled in with subsequent models.

In the lateral model, a flight plan is first expanded into a *reference trajectory*, a sequence of line segments connecting the associated fixes. This trajectory is then perturbed using empirical, stochastic models describing how aircraft are observed to follow their planned routes. Basically, these are models for offsets and correlated noise, validated using large datasets of recorded tracks.

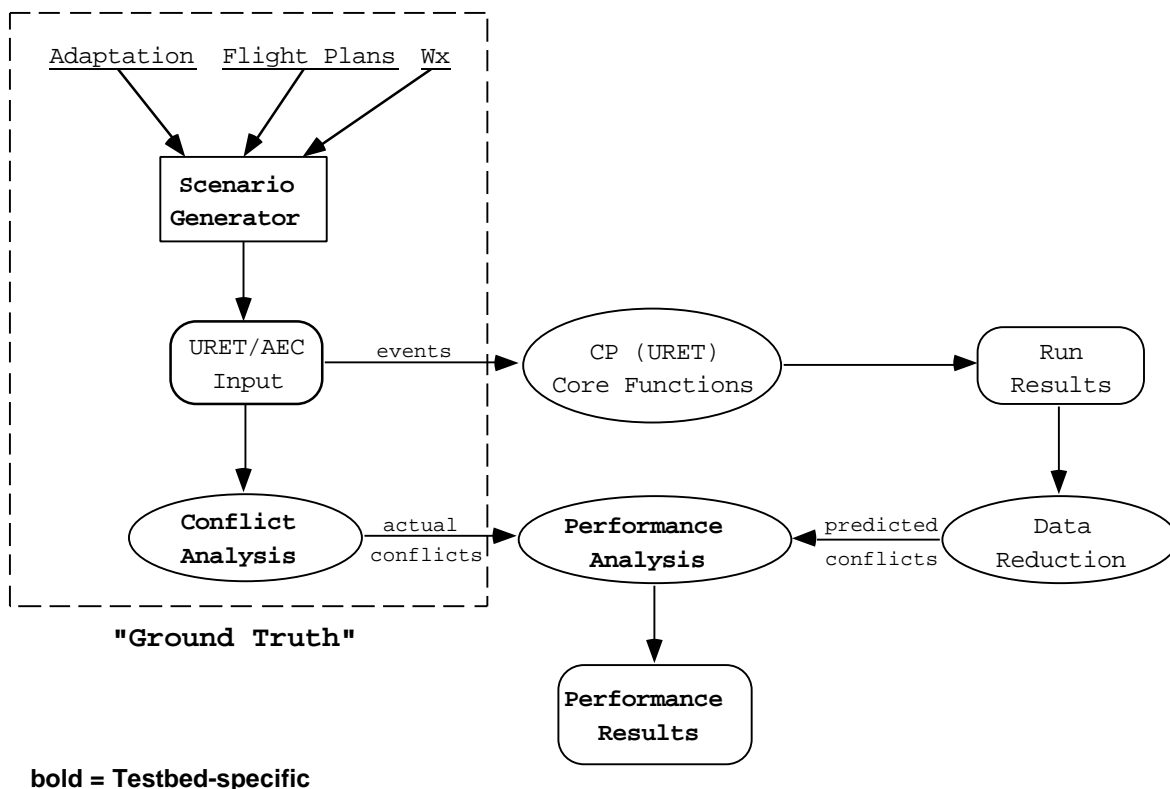
The purpose of the **longitudinal model** is to compute the correct value for the time at which the aircraft reaches the X,Y points specified by the lateral model.

This is done using speed information recorded in the flight plan, wind data and, again, empirical noise models derived from observed tracks.

The **vertical model**, which is strongly correlated with the longitudinal model during ascent and descent phases of flight, is analogous to the lateral model. It uses vertical information from the flight plan and any additional altitude messages, along with empirical models, to determine the sequence of altitudes corresponding to the X,Y,T vectors already computed.

As noted, these synthetic tracks are formatted into a file that can be directly input to the conflict probe.

Since the data given to the probe are usually corrupted by (modeled) radar/tracker noise, a second file is created without such noise. This errorless “ground-truth” file is used to identify any conflicts in the scenario.



**Figure 6. Conflict Probe Testbed Flow Diagram**

### *Conflict Analysis*

The testbed is focused on assessing the accuracy of conflict-probe predictions. This requires definitions for two categories as follows:

- The set of observed conflicts
- The associated metrics

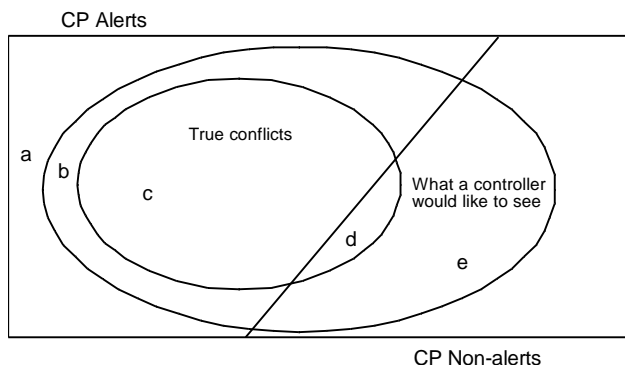
The set of observed conflicts is derived from a list of constraints and factors pertaining to the output of the

conflict probe and their significance. In particular, the testbed assesses only the *initial* prediction of the probe with respect to any given aircraft pair, conditioned on that conflict's having been *notified* to a controller. If the probe subsequently modifies this prediction, the modifications are ignored.

Conflict analysis consists of examining the ground-truth file to identify any *true conflicts* between aircraft pairs or any *near conflicts* that do not meet the strict definition of “conflict” but which might

nevertheless be of operational significance to controllers.

The open-ended universe of discourse is depicted in figure 7.



**Figure 7. Categories Of Conflict Probe (CP) Results**

Here, a true conflict, subset c+d, is defined to be a violation of FAA en route separation minima [7], viz., 5 nautical miles laterally and 1,000 feet vertically (2,000 feet if above FL290). “What (additional encounters) controllers would like to see,” subset b+e, is currently not defined with equivalent mathematical precision. The formulation involves many factors, including lateral and vertical flight

characteristics, encounter geometry, the traffic situation, and other “operational” factors; and is beyond the scope of this paper. For the time being, as a rough approximation, subset b+e is taken to include encounters with a separation of 10 nautical miles horizontally, plus the aforementioned vertical criteria, that are not already conflicts according to the FAA’s technical definition. Finally, as the figure indicates, it is also possible for a conflict probe to issue alerts of no interest to a controller (subset a).

For any given scenario, the number of elements found in each of these subsets, and their ratios, constitute the primary testbed metrics.

*Performance Analysis*

As far as the testbed is concerned, the only input required from the conflict probe system is its list of predicted conflicts. The performance analysis module takes these predictions and compares them to ground truth. Some of the primary metrics, and their definitions, are listed in table 3, where the letters a-e refer to the number of elements in the respective subsets. Note, in particular, that not all the metrics are normalized to the same basis and are thus not necessarily additive. This is important when these metrics are interpreted as probabilities.

**Table 3. CP Prediction Metrics**

		Alerts		
		Successful	Missed	False
Technical		$\frac{c}{c+d}$	$\frac{d}{c+d}$	$\frac{a+b}{a+b+c}$
Operational		$\frac{b+c}{b+c+d+e}$	$\frac{d+e}{b+c+d+e}$	$\frac{a}{a+b+c}$

There are a large number of secondary metrics as well, related to factors such as warning time, conflict duration, etc. There is, for example, a set of metrics identical to those in table 3 but conditioned on a controller’s having received at least k minutes of advanced warning for the putative conflict. In

experiments to date, parameter k has been tentatively set equal to 5.

## Sample Results

[Since subset b+e (figure 7) is yet to be adequately defined, the results presented below are intended only to be illustrative, **not** definitive.]

The testbed was used with the conflict probe of a pre-interfacility version of URET. The scenario consisted of 960 flights, with flight plans corresponding to those seen in six (consecutive) hours of data recorded at Indianapolis Center. In one run, frequency counts for the various subsets, using the definitions given above, were those shown below:

Subset	Frequency (est.)
a	284
b	89
c	127
d	1
e	3

Hence, the probability of missing a true conflict is estimated as 1/128 while the probability of falsely predicting a true conflict is estimated as 373/500, and so forth. This is all based upon a single sample. Proper estimates would require many samples, from many centers, and should be accompanied by robust estimates of confidence limits.

As an example, the central 95-percent confidence interval for the ratio  $373/500 = 0.746$  is [0.706, 0.782] [8]. It should be emphasized that this value reflects only measurement variance. Any additional sources of variance (i.e., *process noise*) would tend to widen this estimated interval.

It must be noted that the single missed conflict ( $d = 1$ ) listed above was very marginal, lasting only 3.3 seconds. Its lateral miss distance was 4.7 nautical miles and its vertical miss distance was 996 feet during the conflict. The aircraft had been separated vertically when the higher one began a descent just before lateral separation was achieved.

## Discussion

The approach taken in developing the testbed has proven to be powerful and flexible. Ground truth is clearly known to perfection, input is unlimited, and

the usual confounding factors are eliminated *via* the conflict analysis module. That is, any conflicts, however defined, can be readily identified unambiguously without approximation. Overall, the testbed provides a very robust tool for investigating the performance of a conflict probe.

## References

[1] Kerns, K., and A. L. McFarland, Dec. 1998, *Conflict Probe Operational Evaluation and Benefits Assessment*, to be published, The MITRE Corporation, McLean, VA.

[2] Brudnicki, D. J., and A. L. McFarland, June 1997, *User Request Evaluation Tool (URET) Conflict Probe Performance and Benefits Assessment*, MP97W000112, The MITRE Corporation, McLean, VA.

[3] Brudnicki, D. J., et al., April 1998, *URET Evaluation Report (Interfacility Capability)*, MTR98W0000045, The MITRE Corporation, McLean, VA.

[4] Brudnicki, D. J., W. C. Arthur, and K. L. Lindsay, April 1998, *URET Scenario-based Functional Performance Requirements Document*, MTR98W0000044, The MITRE Corporation, McLean, VA.

[5] Conker, R. S., W. D. Love, and M. P. McLaughlin, September 1998, *Conflict Probe Testbed Update and Initial Assessment of Conflict Probe Performance*, WN98W0000086, The MITRE Corporation, McLean, VA.

[6] McLaughlin, M. P., D. J. Brudnicki, R. S. Conker, R. Lejeune, and W. D. Love, July 1997, *Traffic Scenarios for Probe Performance Evaluation: Testbed Concept and Methodology*, MP97W0000127, The MITRE Corporation, McLean, VA.

[7] FAA, 1989, *Air Traffic Control, FAA Handbook 7110.65*, U. S. Department of Transportation, Federal Aviation Administration, Washington, D.C.

[8] Nicholson, B. J., August 1985, "On the F-Distribution for Calculating Bayes Credible Intervals for Fraction Nonconforming," *IEEE Trans. on Reliability*, Vol. R-34, No. 3, pp. 227-228.