# MEASURING UNCERTAINTY IN AIRSPACE DEMAND PREDICTIONS FOR TRAFFIC FLOW MANAGEMENT APPLICATIONS

Craig R. Wanke*, Michael B. Callaham**, Daniel P. Greenbaum†, Anthony J. Masalonis**

The MITRE Corporation
Center for Advanced Aviation System Development
7515 Colshire Drive, McLean, VA USA 22102

## Abstract

Traffic flow management (TFM) in the U.S. is the process by which the Federal Aviation Administration (FAA), with the participation of airspace users, seeks to balance the capacity of airspace and airport resources with the demand for these resources. This is a difficult process, complicated by the presence of severe weather or unusually high demand. TFM in en-route airspace is concerned with managing airspace demand, specifically the number of flights handled by air traffic control (ATC) sectors; a sector is the volume of airspace managed by an air traffic controller or controller team. Therefore, effective decision-making requires accurate sector demand predictions. While it is commonly accepted that the sector demand predictions used by current and proposed TFM decision support systems contain significant uncertainty, this uncertainty is typically not quantified or taken into account in any meaningful way. The work described here is focused on measuring the uncertainty in sector demand predictions under current operational conditions, and on applying those measurements towards improving the performance and human factors of TFM decision support systems.

## Background

### The Role of Demand Predictions in TFM

Traffic managers have many options when trying to address excess demand on a resource. For excess airport demand, a ground delay program is often used, in which arrival "slots" are rationed among airspace users, and flights are assigned delayed departure times such that available arrival capacity will be efficiently used. En route sector congestion, resulting from unusually high demand or when available airspace is limited due to hazardous weather, can be controlled several ways. Flights can be rerouted around hazardous weather and/or congested areas. Access to airspace can be limited by imposing miles-in-trail (MIT) restrictions at the airspace boundary, by applying ground delay, or in extreme cases by halting departures to some destinations (ground stop).

Decision support tools for TFM, therefore, must provide predictions of resource demand. Ideally, predictions should be provided based both on the current traffic situation and on proposed traffic management strategies, so that candidate solutions can be developed and compared. For example, the Enhanced Traffic Management System (ETMS)[1] used in the U.S. National Airspace System (NAS) provides real-time resource demand estimates based on predicted aircraft trajectories. In the near future, ETMS will be capable of predicting resource demand as it would be affected by proposed reroute strategies[2], and research continues towards more sophisticated strategy impact assessment capabilities.[3,4]

### What is Airspace Demand?

Airspace demand can be literally defined as the number and distribution of airspace users – aircraft – that seek to use a chunk of airspace (typically, a single ATC sector). A slightly more detailed definition is required in the context of real-time TFM decision support. The ideal demand estimate for traffic managers would provide the expected number of aircraft in a sector at a specific future time, and be based initially on best-known aircraft intent in the form of flight plans or, failing that, flight schedules. It would include the impact of TFM decisions that have already been made, but assume that no further actions will be taken and that sectors have infinite capacity. In other words, the demand prediction should indicate what would happen in a sector if, from this moment forward, no TFM or ATC actions were taken to control capacity. This

*Senior Principal Engineer, Senior Member AIAA
**Lead Engineer, Senior Member AIAA
†Senior Engineer

estimate provides the basis for choosing those actions required to keep sector loads within capacity.

An important consequence of this definition is that it is difficult to determine the accuracy of demand predictions using recorded sector load data. For example, it may be accurate to predict, one hour in advance, that a sector will contain 10 more aircraft than its capacity allows. However, it is likely that, when the timeframe of interest arrives, the actual sector occupancy will be at or near the capacity value. TFM or ATC actions will have been taken to maintain safe sector loading; in fact, they may have been taken as a direct result of the +10 demand prediction. Therefore, it is incorrect to compare the one-hour demand prediction to the actual sector loading to determine "prediction error." In short, the quantity being predicted – sector demand – doesn't actually occur, so there is no "truth" data against which to calculate prediction error. Alternate methods must be employed; one such method was developed for this study.

Sector Load Predictions in the NAS

The ETMS provides demand predictions for most National Airspace System (NAS) sectors in 15-minute bins, for prediction look-ahead times (LAT) of several hours. This information is available for particular sectors by user request, or (as of November 2002) in a collected form on a Sector Count Monitor (SCM) display as illustrated in Figure 1. Note that this screenshot was taken from the MITRE/FAA

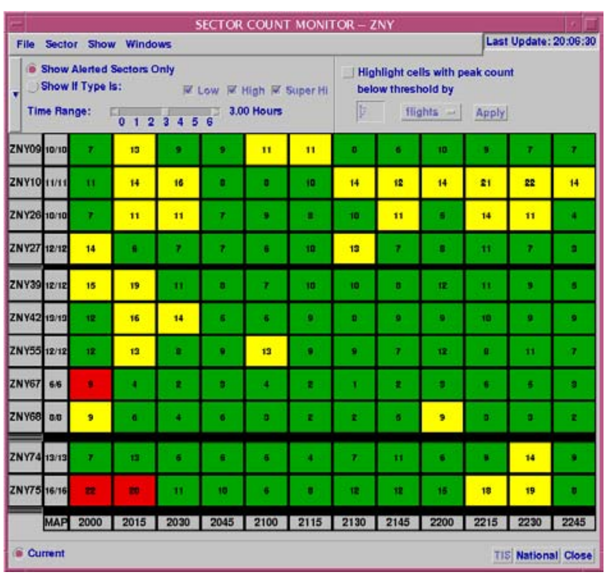


**Figure 1: NAS Sector Load Predictions.** This example shows predicted traffic demand in the New York ARTCC (ZNY).

Collaborative Routing Coordination Tools (CRCT) prototype,[3] which was used to develop this display prior to implementation in the ETMS.

Each box in the SCM matrix represents a 15-minute period, and the number in the box represents the maximum predicted traffic count for any single minute within that 15-minute span. This value is often referred to as maximum instantaneous aircraft count (IAC) or simply "peak count" for the interval. The horizontal axis indicates increasing LAT (corresponding to 2000 to 2245 UTC, in this case). Each row of the matrix represents predictions for single sector (e.g. ZNY09). Next to the sector name are two sector alert thresholds (e.g. "18/18"), although currently, only one is used. This threshold is called the Monitor/Alert Parameter (MAP) and is compared to the peak count to determine whether a sector should be alerted.

When the peak count is predicted to exceed the MAP for a sector, the corresponding box is colored yellow or red. Red alerts indicate that, of the aircraft involved in the peak count, enough are already airborne to exceed the MAP even if pre-departure flights are not counted. Otherwise, the alert will be yellow.

As previously mentioned, the ETMS uses predicted aircraft trajectories to predict sector load. Many data sources are used in the prediction of aircraft trajectories. These include filed flight plans, flight schedules and historical routes (for flights which have not yet filed plans), wind forecasts, track reports, aircraft performance characteristics, and many others. All of these data sources, and the trajectory models themselves, are uncertain, and therefore the demand predictions are uncertain.

Why Measure Demand Prediction Uncertainty?

While it is widely accepted within the traffic management community that sector demand predictions are uncertain, the magnitude of that uncertainty is not well-understood. ETMS sector load predictions include a crude estimate of uncertainty, in that alerts are differentiated into "red" and "yellow" based on whether or not all the aircraft involved are airborne. This is based on the assumption that departure time uncertainty is the largest source of uncertainty in the predictions. However, there are other uncertainty factors, and the yellow/red distinction does not consider how many of the aircraft involved in the over-capacity prediction are currently airborne. Finally, the yellow/red distinction is purely relative. It does not indicate, in an absolute sense, how uncertain the demand prediction is.

This information is important in decision making. What does a particular alert actually mean? Given the deterministic traffic count data provided by current tools, it is difficult to judge the importance of taking action. Intuition suggests that a predicted peak count of 3 over the sector threshold is likely to indicate an actionable problem. However, some uncertainty measure would greatly aid in making the decision. For example, does the +3 prediction mean that there is an 80% probability that the demand exceeds capacity, or a 20% probability?

It is an open research question whether it is desirable to present probabilities such as these directly to traffic managers. ATC personnel do think in terms of probability; in a human-in-the-loop simulation study by Masalonis et. al.,[5] controllers were presented with hypothetical traffic situations and asked to rate the probability that they would take action to resolve a conflict. Some participants used a full range of probabilities, depending on the situation, while others tended to provide probabilities of either 0% or 100% in most situations. There was evidence of correlation between probability ratings and conflict severity.

Even if probability is not directly presented, the automation's knowledge of the probability of an event of interest (for en route TFM, exceeding the MAP) can be used to modify alerting thresholds and symbology to better convey the uncertainty of predictions, as is done by color-coding in the ETMS. This strategy is also used in the User Request Evaluation Tool (URET), a conflict detection capability for en route ATC.[6] Another approach is performance metric alerting (PMA), developed by Yang and Kuchar.[7] In the PMA approach, the probabilities of missed alert and of false detection are calculated in real time, and used to determine whether or not to issue an alert.

## Related Traffic Prediction Work

This work builds on a great deal of recent research in the general area of traffic prediction for TFM and ATC applications. Work has been done to identify, rank, and quantify sources of sector demand uncertainty.[8,9] Aircraft trajectory prediction accuracy under operational conditions has also been studied in detail.[10,11,12,13] These studies provide a basis for identifying and understanding the mechanisms that produce demand prediction uncertainty.

Sector count predictions have also been directly studied by comparing predictions to actually-experienced sector aircraft counts.[14,15] This approach provides a useful starting point, but is subject to the limitations discussed earlier. Most recently, techniques have been developed for probabilistic prediction of sector demand,[16,17] based on knowledge of the uncertainty in the individual aircraft trajectory predictions which make up the sector load prediction.

## Approach

Simulating aircraft trajectories, as done by Mueller, et. al.[17], is a powerful and flexible approach to quantifying demand uncertainty. By explicitly modeling the various component uncertainties in trajectory prediction, Monte-Carlo simulation techniques can be applied to model arbitrary traffic and airspace situations. This approach avoids the previously-described difficulty with using actual traffic. Since the uncertainties in modeling are individually modeled, simulation can be used to identify the effect of individual uncertainty components on the overall demand uncertainty. Furthermore, the distributions of the components can be controlled to simulate possible future changes in the operating environment (e.g. better data sources) and thereby evaluate the potential benefits of reducing the component uncertainties.

These benefits come at a price. It is feasible and instructive to use this approach for specific examples, such as in Ref. 17 where it is applied to a single sector, for a single aircraft type on a single route. However, it is difficult and expensive to develop general purpose simulation models that work under a wide range of conditions. There are many aircraft types, which exhibit different degrees of performance modeling uncertainty. There are many types of uncertainty components, some of which are quite difficult to model (for example, in-flight route amendments) and some of which may interact in unknown ways. Different airspace types have different route and altitude patterns, producing different levels of uncertainty.

The work discussed here requires uncertainty estimates for present-day demand predictions, under a variety of traffic conditions and for a variety of airspace types. While the simulation method provides the best and most flexible option for studying demand uncertainty, it was deemed too difficult for the initial phase of this work. A different approach was chosen, based on empirical measurements of sector load uncertainty.

As noted, in the general case, actual traffic cannot be used directly to evaluate demand predictions. However, there are specific cases in which no significant TFM actions are taken, and in these cases, the measured traffic could be used to evaluate demand predictions. It was hypothesized that these cases would be those in which predicted sector peak counts were significantly less than the MAP, thereby not requiring TFM actions to control congestion.

This hypothesis has some weaknesses. For example, a weather-related reroute may place traffic into a sector that was not predicted to have high traffic loads. Or, a sector with light demand may be adjacent to another with heavy demand, and a miles-in-trail restriction applied to control the heavy-demand sector could affect the light-demand sector as well. It was assumed that, with a sufficient quantity of data, these effects should not have a major impact on the analysis.

By measuring the uncertainty directly – assuming that the light-demand hypothesis holds – then all significant components of uncertainty will be included in the estimate. This eliminates the modeling work required for trajectory simulation. The obvious disadvantage is that none of the trajectory uncertainty components are individually modeled, so only the overall demand uncertainty in the present-day environment can be usefully studied. Also, vast quantities of data are required to span the traffic conditions of interest.

Data Collection

Fortunately, as part of other MITRE research, approximately a full year of sector demand predictions for the entire NAS have been recorded. This is enough data to get statistically-interesting samples for most traffic conditions of interest. These predictions were made by the CRCT prototype, which uses traffic prediction algorithms similar to those used by the ETMS. Details of these algorithms are available in Holly, et. al.[18]

The predictions are in the form represented in Figure 1, namely peak counts in 15-minute intervals. Every 15 minutes, predictions with time horizons (LAT) of 0 to 6 hours ahead are recorded. In other words, for each NAS sector and each 15 minute period spanned by the data, predictions are available from 6 hours down to zero time in advance, at 15-minute LAT intervals. The data spans 286 days, from 23 Jan 2002 to 20 Sep 2002, includes predictions for the 754 NAS sectors with established MAP values, and includes over 400 million total demand predictions.

One complication is that the actual number of aircraft that were present in the sector at any given time was not available, since radar surveillance data was not recorded for all of the predictions. Therefore, the zero LAT predictions (made at the beginning of the 15-minute period of interest) were used as a proxy for the actual peak aircraft count. Analysis of a limited set of data for which radar surveillance data was available indicated that this assumption is sufficiently accurate for this study, with the possible exception of predictions for sectors which handle predominantly departure flows. This will be discussed in detail later.

Defining Light Demand

Specific criteria are required to apply the light-demand hypothesis. In particular, how light is light enough? Insight was gained through exploratory data analysis. Traffic predictions at varying LAT were compared with the zero-LAT predictions, and the mean difference was calculated. This was done for several LAT values, and plotted against the predicted peak count *relative to the MAP* in Figure 2. A prediction of 15 for a sector with a MAP of 12 is classified as "+3", in the same bin as a prediction of 18 for a sector with a MAP of 15. Note that the mean difference measures the bias in the predictions, not the uncertainty of the predictions or the shape of the prediction error distribution.

The results show two interesting trends that have reasonable operational explanations. At low sector loads, the peak count estimate is approximately unbiased for short LAT. As LAT increases, the peak count estimates develop an increasingly lower bias. In other words, the peaks are under-predicted on the average. This is due to an input data feature. At short LAT, most or all flights that may enter the sector have filed flight plans. At longer times, some percentage of these flights (primarily General Aviation) have not filed flight plans, and have no regular operation schedule, and so are unknown to the prediction algorithms.

The second obvious trend is the increasing over-prediction bias as the predicted sector load increases. This is likely to be the effect of TFM actions. For example, for predictions of 5 aircraft over MAP, the mean difference is 4. In this case, the average actual peak was 1 aircraft over MAP rather than 5 over, presumably because actions were taken to manage the congestion before it occurred. Note also that the curves for different LAT are quite close together at high sector loads relative to MAP. This is consistent with the hypothesis, since the effect of sector congestion management would be to force traffic counts down to the MAP, suppressing other sources of prediction uncertainty such as the previously-noted missing flight plans.

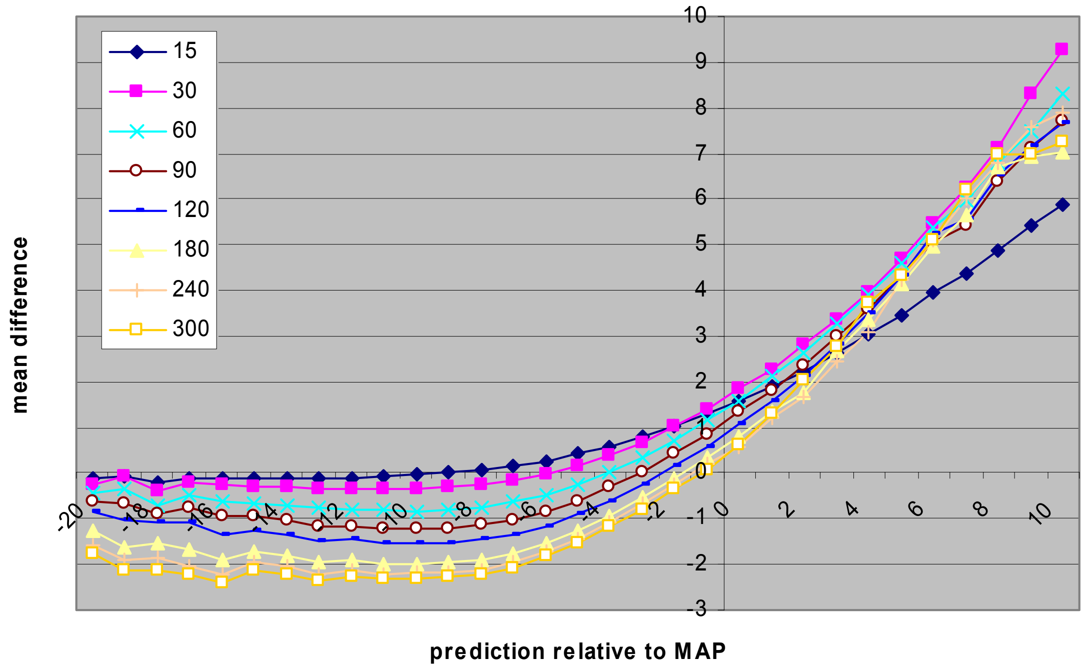


**Figure 2: Bias Trends in Demand Predictions by Prediction Look-Ahead Time (LAT) in Minutes**

Another interesting feature is that the 15-minute LAT curve does not climb as steeply as the longer LAT curves as predicted counts increase above MAP. The most likely explanation is that, at 15 minutes before the event, there is not enough time to take actions that would bring the peak count back down to the MAP.

One can conclude from this plot that peak traffic load predictions relative to MAP of -7 or less indicate situations with little or no TFM action. Therefore, for the analyses presented below, the data set was reduced to include only cases satisfying that criterion. This reduced the total data set by approximately one-eighth, to around 350 million observations.

The converse of this conclusion is that comparing predicted peaks to actual peaks at MAP-relative loads of -6 or greater is not an effective way to evaluate demand prediction accuracy, since the sector load data has been affected by significant TFM actions. It may work for comparing relative performance between conditions (e.g. different prediction approaches, or differences in input data sources such as wind forecasts), but is not effective for measuring demand prediction accuracy in absolute terms.

Factors Affecting Demand Predictions

The reduced data used to measure and classify the demand prediction uncertainty as a function of several factors. Table 1 summarizes the factors chosen for initial consideration, though not all have been studied at the time of this writing. These factors are not all independent, nor are they equally important. For example, the number of scheduled airline flights changes according to the day of the week. However, this effect is primarily seen as an overall change in traffic levels, and may be adequately captured by conditioning the prediction uncertainty on the predicted value. Also, even seemingly independent factors may have interaction effects.

After exploratory analysis, three variables were identified as having strong effects on uncertainty distributions and were therefore chosen for initial study. These are highlighted in Table 1, and include (1) LAT, (2) predicted peak count, and (3) sector traffic type.

**Table 1: Factors Affecting Demand Predictions**

| Category | Class | Values |
|---|---|---|
| Airspace | Sector | Individual Sector (754 total)<br>Altitude Class (Low, High, Super High)<br>**Primary Traffic Type (Departures, Arrivals, En Route, Mixed)** |
| | ARTCC | Individually (21 total) |
| Time | Day | Day of week |
| | Time-of-Day | Hour of day, local time |
| | Time-of-Year | Season |
| Prediction | LAT | **15 minute intervals, 0 to 6 hours** |
| | Value | **Absolute number** or relative to MAP |
| Weather | Severe WX | Location (in sector, near sector, none) |
| | Jet Stream | Location, direction, strength |

This third classification requires some explanation. Sectors were assigned to one of four traffic categories – (A) arrival, (D) departure, (E) en route, and (M) mixed – based on the predominant type of traffic passing through the sector on a representative traffic day. A 24-hour CRCT simulation was run on archived ETMS data from 16 August 2002. At each hour of the simulation, a list of flights predicted to enter each sector over the subsequent six hours was generated from modeled flight trajectories and FAA-provided sector geometries. Sectors were categorized by whether flights typically entered the sector at the beginning, at the end, or in the middle of their routes. In order to ensure that analyzed trajectories extended completely from origin to destination, only flights departing and arriving within the USA were considered.

Because the set of pre-departure flights predicted to pass through a sector in the near future tends to over-sample departures relative to arrivals, only flights predicted to pass through the sector between five-and six hours after the prediction time were included. Sectors which were the first or last sector entered by at least 70% of those flights entering the sector were considered departure and arrival sectors respectively. Sectors for which at least 70% of those flights entering the sector did so as neither the first nor last sector on the route were considered en route sectors. All remaining sectors were considered mixed traffic sectors. Of 754 sectors categorized, 47 were found to be arrival sectors, 49 were departure sectors, 406 were en route sectors, and 252 were mixed traffic sectors.*

## Results

The reduced data set was analyzed in terms of *demand prediction error,* defined as the predicted peak value minus the zero-LAT peak value. Thus, positive error values indicate over-prediction, and negative values indicate under-prediction. Prediction error was analyzed as a function of LAT, prediction value, and sector traffic type, and was visualized several ways. Also, the resulting distributions were used to draw conclusions about the characteristics of current strategies for sector load alerting.

### Prediction Error Distribution

One simple way to study a distribution is via percentiles.[19] Figure 3 is a "box-and-whisker" plot of the percentiles of prediction error distribution for mixed sectors, tabulated for four LAT values. The lower and upper bounds of each box represent the 25th and 75th percentiles of the distribution, respectively, and the horizontal line within the box indicates the median (50th percentile). In other words, 25% of the prediction error values fall below the box and 25% fall above the box. The lower and upper whiskers (horizontal lines connected to the boxes by dashed vertical lines) show the 10th and 90th percentiles, respectively.

It is clear that predictions are much better at short LAT, where the middle 50% of the distribution falls between -1 and +1, than at long LAT. Also, there is a systematic and asymmetric bias toward under-prediction at longer LAT, as indicated by the decreasing median and 25th percentile values. It is also interesting that the upper whisker is 2 for all four LAT values plotted, while the lower whisker drops to –6 at 300 minute LAT, indicating that a significant percentage of the predictions are very low with respect to the demand.

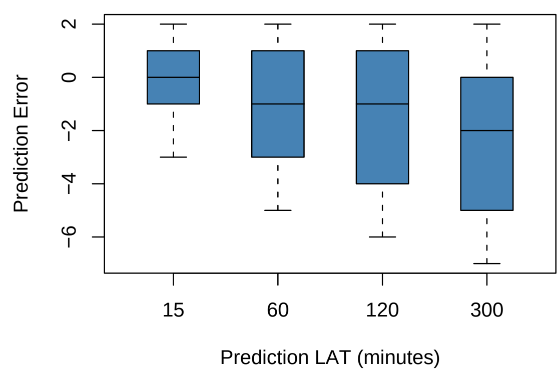


**Figure 3: Percentiles of Prediction Error Distribution for Mixed Sectors**

Because the predictions take discrete values, the percentiles are of limited use for comparing uncertainty among many conditions. Figure 4 shows the mean and standard deviation of the prediction error for mixed sectors, as a function of both the predicted value and of LAT. This provides a more sensitive way to compare several prediction curves.

---

* This categorization is only for the purposes of this study. From an FAA procedural perspective, there are more than 47 sectors which designed to primarily handle arrivals, and more than 49 which are designed as departure sectors. The 70% criterion used here for primary traffic type may be too restrictive, and may be reevaluated.

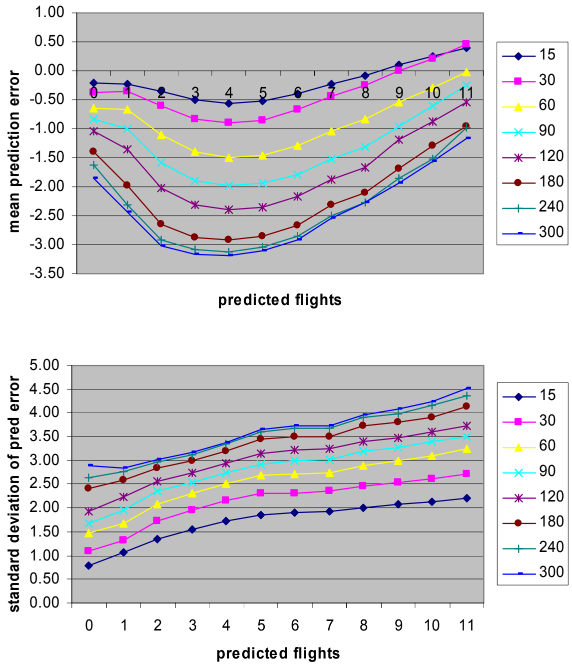


**Figure 4: Mean and Standard Deviation of Prediction Errors for Mixed Sectors by LAT (min)**

The mean value (upper plot) shows increasing under-prediction with increasing LAT, as seen previously in Figure 2. However, since the error is plotted as a function of predicted absolute count (not relative to MAP) the curves are not as flat as in Figure 2. For predicted values near zero flights, under-prediction is less likely, so means are higher. From predicted values of 4 upward, the mean prediction error steadily increases with increasing predicted value.

Standard deviations (lower plot) are a measure of the "spread" of the prediction error distribution, and larger standard deviations indicate higher uncertainty. As expected, standard deviations increase steadily with increasing LAT. This increase is substantial. For a prediction of 8 aircraft, the standard deviation increases from 2 at 15 minute LAT to almost 4 at 5 hour LAT.

Figure 5 also shows mean and standard deviation, for a single value of LAT (60 minutes), but parameterized by sector type. It is apparent that departure sectors have different prediction characteristics from other sectors. The primary feature of departure sectors is that predictions for them involve a much greater proportion of pre-departure flights than for other sector types. Therefore, it is expected that predictions for departure sectors should be more uncertain, since departure time predictions are highly uncertain.[8,9] In this analysis, this is shown by the much higher standard deviation in departure sectors. Similarly, predictions are most accurate for arrival sectors, since a higher proportion of involved flights are airborne 60 minutes from the period being predicted.

However, it must be noted that the analysis technique used here relies on the assumption that zero-LAT peak predictions are an accurate proxy for the actual peak aircraft count. This assumption is weakest for departure sectors in that, even at the beginning of a 15-minute interval of interest, some flights involved in the peak prediction for that interval will not yet have departed. If delays at the airports involved are unusually high, the prediction algorithms may model an inordinately large number of flights as leaving within the next 15 minutes, since according to those flights' filed plans and typical ground movement times, they should have already departed.

Therefore, the meaning of the zero-LAT predictions for departure sectors is rather different than for other sectors. It can be argued that the zero-LAT prediction is the best possible estimate for actual departure sector demand, since late flights normally want to depart as soon as possible, and hence it is still useful for this analysis. Regardless, the uncertainty distributions for departure sectors shown here are clearly more affected by the zero-LAT assumption than predictions for the

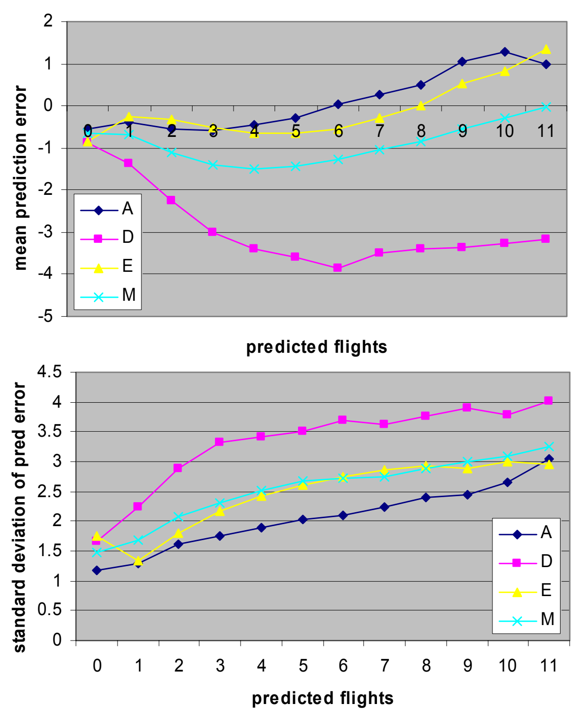


**Figure 5: Mean and Standard Deviation of Prediction Errors for 60 min LAT by Sector Type.**
A = arrival, D = departure, E = en route, M = mixed.

other sector types, for which zero-LAT predictions are made up almost entirely of airborne aircraft, and are hence very accurate representations of the actually-experienced peak sector count.

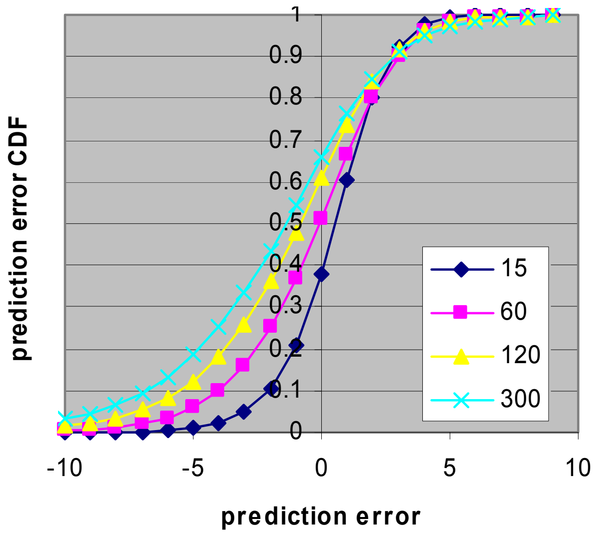


**Figure 6: CDF for Mixed Sectors by LAT**

A complete way to visualize the prediction uncertainty is to display the entire distribution. Figure 6 shows the cumulative distribution function (CDF) for all predictions for mixed sectors, parameterized by LAT. Each point (y-axis value) on the CDF indicates the probability that the prediction error is equal to or less than the corresponding x-axis value. This means that an unbiased prediction would have a CDF value of 0.5 for a prediction error of zero, and that the steeper the slope of the CDF, the more precise the prediction. As expected, the longer LAT curves in Figure 6 are flatter

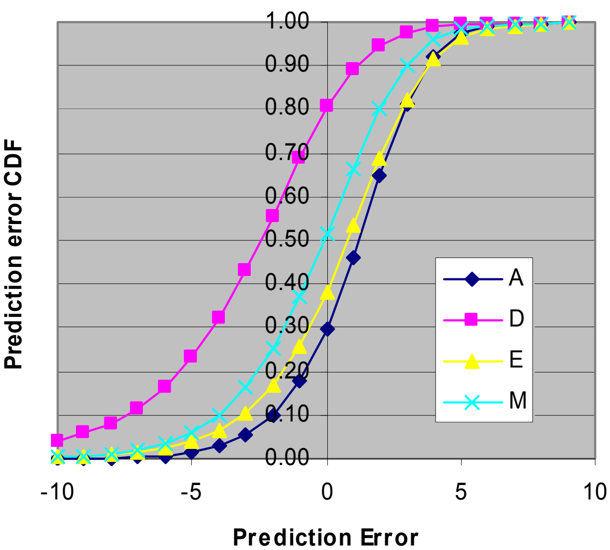


**Figure 7: CDF for 60 min LAT by Sector Type**

than the 15 minute LAT curve, and show a higher probability of errors less than zero (under-prediction) than errors greater than zero.

Figure 7 shows the CDF for 60 minute LAT, classified by sector type. As in Figure 5, it is apparent that arrival sector predictions are most accurate, with the steepest distribution, while the departure sector predictions are the least accurate.

One final method for observing the distributions is shown in Figure 8. Shown are contours of the 25$^{th}$ and 75$^{th}$ percentiles of the distributions, varying by LAT and by prediction value relative to MAP.[†] The distance between the 25% and 75% curves widens with increasing LAT, indicating increasing uncertainty. Also, the interval between the curves shifts downward, showing the increasing tendency towards under-prediction at high LAT. Finally, the distance between the 25% and 75% curves widens with increasing relative count (i.e. as the predicted count gets closer to the MAP). This effect is more noticeable at the 25$^{th}$ percentile than at the 75$^{th}$ percentile error contour.

<u>Application to Sector Alerting</u>

Once the uncertainty distributions have been characterized, they can be used to evaluate sector alert probabilities under various conditions. For example, if a particular value of MAP is assumed, then the meaning of a particular prediction can be evaluated in terms of the probability that the "actual" sector demand corresponding to that prediction will exceed the MAP.

Figure 9 illustrates application of this technique to mixed sectors with a MAP of eight aircraft. Using the distributions shown in Figure 6, the probability of demand exceeding MAP was computed for different predicted peak values and for several LAT values. As for the CDF, steeper curves indicate less uncertainty in the predictions, so it is apparent (and unsurprising) that there is less uncertainty in shorter LAT predictions. A perfect prediction would have zero probability for predicted peaks of 8 or less, and a probability of 1 for predicted peaks of 9 or more.

[†] This plot differs from Figure 3 in that the 25$^{th}$ and 75$^{th}$ percentile contours are calculated from the PDFs (as in Figure 6) rather than selecting the 25$^{th}$ percentile value directly from the prediction data, which would produce only discrete values. This allows a more sensitive comparison among the different distributions.

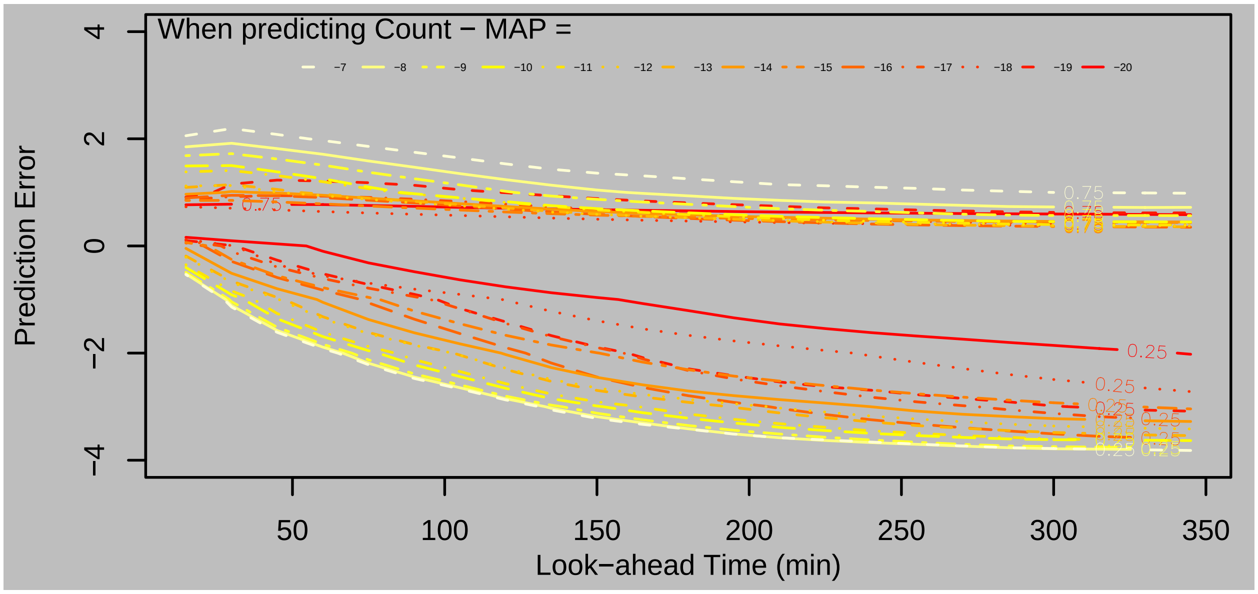


**Figure 8: Probability Contours for Demand Predictions, All Sectors, by LAT and Relative Prediction Value**

It is interesting to look at several predicted values, and compare the probability of actual demand exceeding the MAP among different LAT values. Recall that, using the ETMS alerting rules, a predicted peak of 9 or more would trigger an alert.

For a predicted peak of 6 (MAP – 2) at 15-min LAT, there is only a 12% probability that the actual demand exceeds MAP, while for 60-min LAT, that probability is 28%, and for 120-min LAT, 41%. This trend is due to the systematic under-prediction of demand at longer LAT, as observed earlier.

These probabilities, for predictions less than or equal to

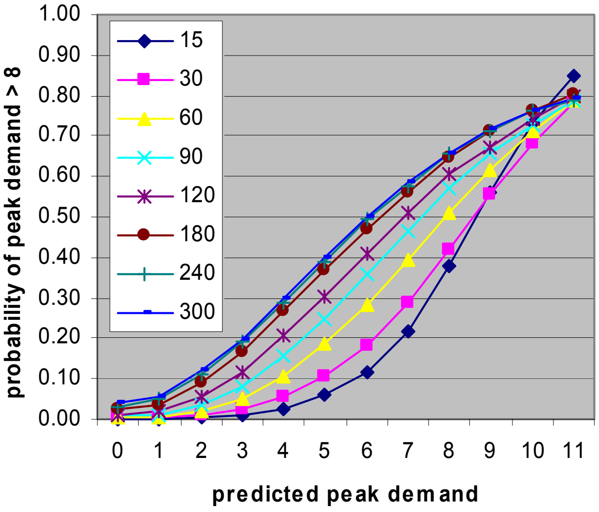


**Figure 9: Probability of Demand Exceeding 8 for Mixed Sectors by LAT (minutes)**

the MAP, can be interpreted as the likelihood of missed alerts, and they increase steadily as LAT increases. Of course, as time progresses and LAT decreases, alerts that were "missed" earlier may be detected. A missed alert in this context, therefore, does not necessarily have serious consequences. Also, peak IAC is a "noisy" metric[‡], and the same prediction algorithms shown here may perform differently if different criteria for sector alerting are chosen. This is a potential area for future research.

For predicted peak count of 11 (MAP + 3), the percentage trend reverses. At 15-minute LAT, a predicted peak of 11 indicates an 85% probability of exceeding MAP. At 120-minute LAT, this prediction indicates an 80% probability of exceeding MAP. Even at 240-minute LAT, this probability is 79%. Therefore, a prediction of 11 indicates a high likelihood that the demand exceeds sector capacity, even when predicted several hours in advance. Such a prediction can be considered reliable for traffic management applications.

At predicted peaks above the MAP, the percentage indicates the likelihood of correct alerting, with the balance (i.e. 100% - 85% = 15% for 15-minute LAT) being the likelihood of false alert.

[‡] The peak IAC in an interval is sensitive to small changes. For example, a small change in airspeed for an aircraft which is projected to enter the sector may push it into or out of the peak minute. As a consequence, the peak IAC fluctuates significantly as time passes and predictions are updated. A metric incorporating some form of time-smoothing would probably be more predictable than peak IAC, but would need to be carefully chosen for operational relevance.

This analysis can be repeated for different MAP values, sector types, weather conditions, or other variables of interest. The results could be used to establish alerting heuristics that are sensitive to the prediction conditions, and thereby establish the desired balance between missed and false alerts. They could also be used to develop alerting displays that are calibrated by the uncertainty in the prediction. For example, yellow and red alerts in the ETMS could be driven by the probability of exceeding the MAP rather than by the presence or absence of pre-departure flights in the prediction. This would provide more useful "alert reliability" information than the current ETMS technique. Or, new displays which directly present probabilistic predictions to traffic managers could be developed. Finally, knowing the validity of predictions under varying conditions and LAT may affect traffic management procedures. For example, given a situation under which predictions are highly uncertain, it may be advisable to delay traffic management actions until the demand is more precisely known.

Extension to Higher Traffic Loads

Because only light-demand predictions were studied, the measured distributions are only valid for predicted peak counts up to 11 aircraft. However, MAP values in the NAS range from 6 to 25 with the majority falling between 12 and 18, inclusive. Also, the prediction error distributions clearly vary with predicted peak value (Figure 4). Therefore, the distributions presented above are not directly applicable to the majority of NAS sectors.

Observing the trends in Figures 4, 5, and 8, it seems that for predictions greater than four flights, the distribution parameters (at least the mean and standard deviation, and perhaps other measures) seem to increase steadily and linearly with increasing prediction count. The change in characteristics for predictions fewer than four flights is due to the one-sided nature of the prediction; the number of flights predicted and experienced must be equal to or greater than zero.

One approach to modeling high demand prediction errors is to use this trend, extrapolating the distribution parameters, to develop approximate distributions for higher prediction counts. This is operationally reasonable, in that since we are purposely avoiding the influence of TFM and ATC actions for calculating demand, there is no reason to expect a radical change in the shape of the distribution as more aircraft are added.

Application to Simulation

One of the hardest tasks in the previously-described simulation approach to evaluating demand uncertainty is validating the simulation models. The statistical data assembled here could be used to validate simulation models, by selecting an appropriate subset of the data representing a particular set of conditions in the simulation.

## Conclusions

A technique has been developed for analyzing airspace demand predictions based on observed data from a prototype TFM decision support system. This technique has been used to quantify the uncertainty in present-day NAS airspace demand predictions. The resulting statistical distributions of prediction error can be used to evaluate the reliability of current sector load alerts, to develop new techniques to display and utilize sector demand predictions with known uncertainty, and to develop procedures that take this uncertainty into account when managing traffic in the NAS.

## References


1. *Enhanced Traffic Management System (ETMS) Reference Manual, Version 7.5*, Report No. VNTSC-DTS56-TMS-004, Volpe National Transportation Systems Center, U.S. Department of Transportation, November 2002.

2. Wanke, C, L. Carlson, J. DeArmon, S. Mayer, S. Miller, and G. Schultz, "New Tools for Strategic Route Planning: Collaborative Routing Coordination Tools," *Handbook of Airline Strategy,* edited by G. Butler and M. Keller, The McGraw-Hill Companies, New York, 2001.

3. Wanke, C., D. Berry, J. DeArmon, S. Miller, N. Taber, L. Rhodes, and M. Yee, "Decision Support for Complex Traffic Flow Management Actions: Integrated Impact Assessment," AIAA-2002-5015, *AIAA Guidance, Navigation, and Control Conference and Exhibit,* Monterey, CA, 5-8 August 2002.

4. Sridhar, B., G. Chatterji, S. Grabbe, and K. Sheth, "Integration of Traffic Flow Management Decisions," AIAA-2002-3748, *AIAA Guidance, Navigation, and Control Conference and Exhibit,* Monterey, CA, 5-8 August 2002.

5. Masalonis, A. J., J. A. Duley, S. M. Galster, D. J. Castano, U. Metzger, and R. Parasuraman, "Air traffic controller trust in a conflict probe during Free Flight," *Proceedings of the 42nd Annual*

*Meeting of Human Factors and Ergonomics Society*, Chicago, IL, 1998.

6. Brudnicki, Daniel J. and Alvin L. McFarland, 1997, "User Request Evaluation Tool (URET) Conflict Probe Results," *1st USA/Europe Air Traffic Management Research and Development Seminar,* Saclay, France, 17-20 June, 1997.

7. Yang, Lee C., and James K. Kuchar, "Performance Metric Alerting: A New Design Approach for Complex Alerting Problems," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans,* Vol. 32, No. 1, Jan 2002.

8. Wanke, Craig, *Practical Limitations of Trajectory-Based Traffic Flow Prediction for Traffic Flow Management.* MITRE Product 98W158, The MITRE Corporation, McLean, VA, August 1998.

9. Krozel, Jimmy, Dan Rosman, and Shon R. Grabbe, "Analysis of En Route Sector Demand Error Sources," AIAA 2000-5016, *AIAA Guidance, Navigation, and Control Conference and Exhibit,* Monterey, CA, USA, 5-8 August 2002.

10. Green, S.M., and R. Vivona, "Field Evaluation of Descent Advisor Trajectory Prediction Accuracy", AIAA 96-3764, *AIAA Guidance, Navigation, and Control Conference*, July, 1996.

11. Lindsay, K. S., and N. E. Rozen, *Functional Performance of the User Request Evaluation Tool (URET) Delivery 3.4 Release 2,* MITRE Technical Report 00W115, The MITRE Corporation, McLean, VA, December 2000.

12. Jackson, M. R., Y. J. Zhao, and R. A. Slattery, "Sensitivity of Trajectory Prediction in Air Traffic Management," *Journal of Guidance, Control, and Dynamics,* Vol. 22, No. 2, Mar-Apr 1999.

13. Mueller, E. R. and G. B. Chatterji, "Analysis of Aircraft Arrival and Departure Delay Characteristics," *AIAA Aircraft Technology, Integration, and Operations Forum,* Los Angeles, CA, Oct 1-3 2002.

14. Masalonis, A., C. Wanke, D. Greenbaum, C. Jackson, E. Beaton, S. Mahmassani, D. Chaloux, E. Cherniavsky, and J. Forren, *Preliminary Analysis of CRCT Traffic Prediction Performance,* MITRE Product 01W62, The MITRE Corporation, McLean, VA., March 2001.

15. Lindsay, Kenneth S., *Predictive Accuracy Analyses for the Collaborative Rerouting Coordination Tools (CRCT),* MITRE Technical Report 02W62, The MITRE Corporation, McLean, VA., Sept. 2002.

16. Meyn, Larry A., "Probabilistic Methods for Air Traffic Demand Forecasting," AIAA 2000-4766, *AIAA Guidance, Navigation, and Control Conference and Exhibit,* Monterey, CA, USA, 5-8 August 2002.

17. Mueller, K. Tysen, John A. Sorenson, and George J. Couluris, "Strategic Aircraft Trajectory Prediction Uncertainty and Statistical Sector Traffic Load Modeling," AIAA 2000-4765, *AIAA Guidance, Navigation, and Control Conference and Exhibit,* Monterey, CA, USA, 5-8 August 2002.

18. Holly, K., Y. Cabeza, M. Callaham, D. Greenbaum, A. Masalonis, and C. Wanke, *Feasibility of Using Air Traffic Complexity Metrics for TFM Decision Support,* MITRE Technical Report MTR02W55, The MITRE Corporation, McLean, VA, August 2002.

19. Tukey, J., *Exploratory Data Analysis,* Addison-Wesley, Reading, PA, 1977.