Report no. VNTSC-TFM-10-01

Probabilistic Predictions of Traffic Demand for En Route Sectors Based on Individual Flight Data

Eugene P. Gilbo
Scott B. Smith

January 2010

Prepared for

Federal Aviation Administration

Office of System Operations Programs

600 Independence Avenue, SW

Washington, DC 20591

Prepared by

Volpe National Transportation Systems Center

Traffic Flow Management Division

55 Broadway

Cambridge, MA 02142

Acknowledgments

The authors would like to thank Midori Tanino and John McCarron of the FAA for their support and encouragement. Within the Volpe Center, we would like to thank Rick Oiesen and Ken Howard for reviewing the report and their valuable comments and suggestions that improved its content. We also thank Mary Costello for editing this report.

Executive Summary

Introduction. The Traffic Flow Management System (TFMS) predicts the demand for each sector, and traffic managers use these predictions to spot possible congestion and to take measures to prevent it. These predictions of sector demand, however, are currently made in a deterministic way with no regard for the uncertainty in the predictions. The purpose of this report is to discuss how these deterministic predictions can be misleading, how probabilistic predictions of aggregate traffic demand counts can be made that take account of the uncertainty in predicting timing events for individual flights, and how this can lead to better information on which traffic managers can base their decisions.

The major questions addressed by this study are as follows:

- What are the errors in TFMS predictions of sector entry time and time-in-sector for individual flights?
- What is the probability that an individual flight will enter a sector during a given one-minute interval?
- Given a set of predictions for several flights, what is the probability distribution for the number of flights to enter a sector during a specified one-minute interval?
- Given these flight predictions, what is the probabilistic demand count prediction (number of flights in a sector during a given minute)?
- How do these probabilistic and deterministic demand count predictions compare?

Empirical Analysis. The empirical portion of this study is based on statistical analysis of historical TFMS data on approximately 39,000 flights for 16 en route sectors during the week of April 10-16, 2009.

Major findings are the following:

- Sector entry time predictions for active flights are much more accurate than for proposed flights, i.e., flights still on the ground. See Figure ES-1.
- For active flights, the distributions of prediction errors for sector entry times were either symmetric or were slightly asymmetric with perhaps a slightly higher probability for active flights to enter a sector earlier than predicted. See Figure ES-1.
- For proposed flights, the distributions were asymmetric with a higher probability for a flight to arrive later than predicted. See Figure ES-1.
- The prediction errors for the amount of time that a flight spends in a sector were about the same for active and proposed flights, and the error for both was much smaller than for the sector entry time. See Figure ES-2.

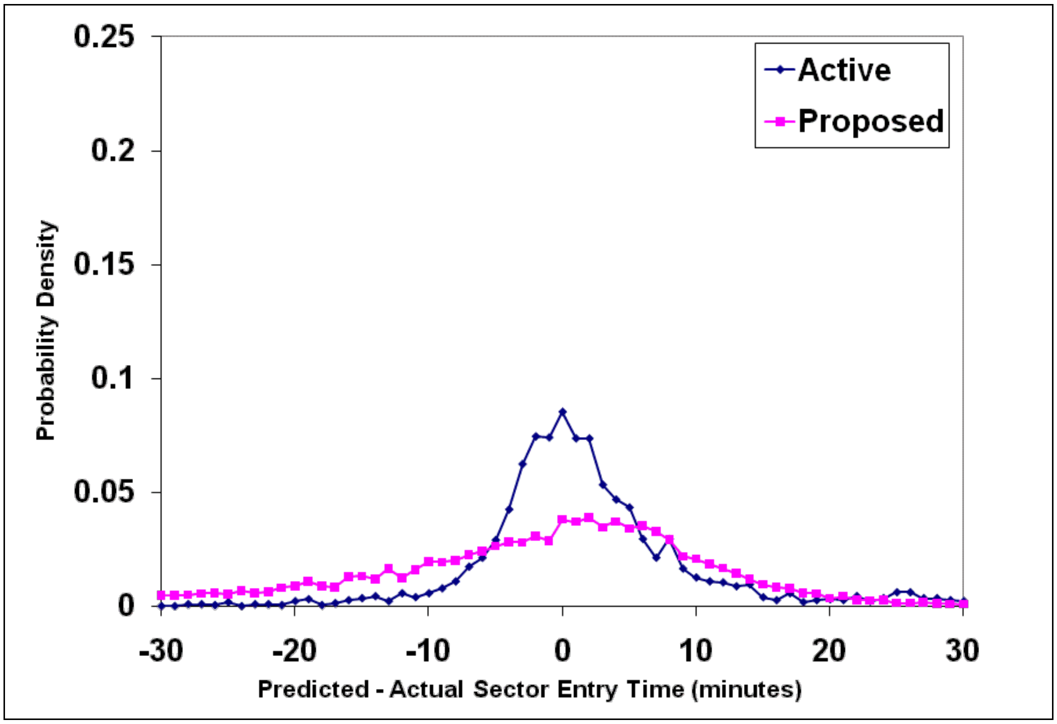

Figure ES-1 Distribution of Errors in Sector Entry Time Predictions (April 2009 Data)

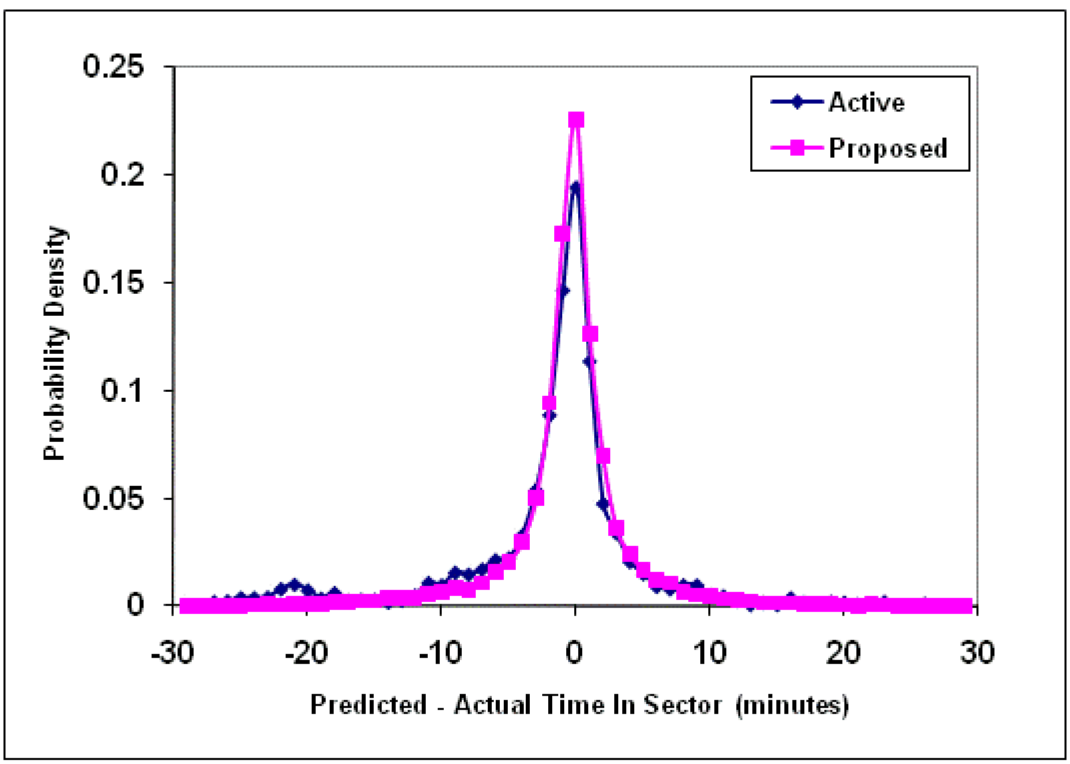

Figure ES-2 Distribution of Errors in Time-in-Sector Prediction: (April 2009 Data)

Theoretical analysis. The report presents a novel analytical approach to translating characteristics of uncertainty in predicting times of sector boundary crossings for individual flights into characteristics of uncertainty in predicting sector demand counts. These characteristics depend on both deterministic traffic predictions and parameters of errors in predicting times of flights' sector boundary crossings. Our approach considers the status of individual flights (active or proposed) in terms of both the characteristics of errors in flights' timing predictions and the composition of deterministic demand predictions. The theoretical analysis derives a probability distribution for a flight to enter a sector during a given minute. Based on this, the mean and standard deviation is derived for both the number of flights *entering* a sector during a given minute and the number of flights *present* in a sector during a given minute. This last derivation provides the basis for a probabilistic prediction of sector demand, with both an expected value and a confidence interval (Figure ES-3). In this figure, the diamonds represent the expected value of probabilistic predictions. Each diamond is surrounded by a confidence interval (the short vertical line that represents one standard deviation above and below the expected probabilistic prediction). The yellow bars are the original deterministic predictions.

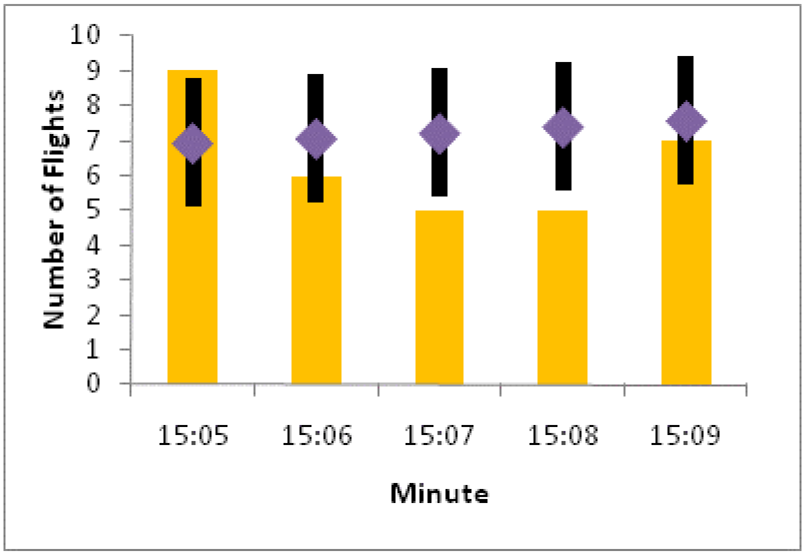

Figure ES-3 Hypothetical Probabilistic Demand Predictions

Figure ES-3 shows that expected demand values can be close to deterministic predictions (e.g., the prediction for 15:09) or far from deterministic predictions. For 15:05, 15:07 and 15:08, deterministic predictions are even outside the confidence intervals, i.e., the deterministic values are highly unlikely and have a very low probability to be true (correct).

Conclusions. The empirical and theoretical analysis leads to several conclusions.

- The deterministic TFMS sector demand predictions are sometimes misleading. For example, the deterministic prediction sometimes falls outside of the confidence interval that is derived from the probabilistic prediction. (See, for example, the times 15:05, 15:07 and 15:08 in Figure ES-3.)
- A major advantage of the probabilistic approach over the deterministic approach is that the former allows a confidence interval to be determined for a demand prediction. In other words, this gives one an idea of how reliable the prediction is. The deterministic approach does not provide a confidence interval.

- Looking at the demand for a single minute, as TFMS now does when measuring the demand for a 15-minute interval, is not adequate since the prediction errors for flights are much greater than a single minute. This means that the probabilistic demand prediction for a minute depends not only on the deterministic demand for that minute but also on the deterministic demands from surrounding minutes.
- A satisfactory probabilistic demand prediction for one-minute sector demand can be made from a weighted average of the deterministic predicted demands from surrounding minutes as an expected value, accompanied by the area of uncertainty around the expected demand. The area is usually measured by the number of standard deviations depending on the desired confidence level

Summary. This report started with giving a probabilistic characterization of when a single flight enters a sector, built upon that to probabilistically characterize the total number of flights entering a sector, and, finally, built upon that to probabilistically characterize the number of flights in a sector, which is what traffic managers are most interested in. Analytical expressions were derived for these various probability distributions and their parameters, and empirical analysis provided specific examples for real sectors.

The next step for this analysis, not undertaken in this report, is to determine exactly how these findings can be incorporated into TFMS's Monitor/Alert function and, thus, to lead to better decision making by traffic managers who are dealing with sector congestion.

Table of Contents

1. Introduction

Federal Aviation Administration (FAA) Air Traffic Flow Management (TFM) decision-making is based primarily on a comparison of deterministic demand and capacity threshold predictions at various National Airspace System (NAS) elements such as airports, fixes and en-route sectors. The current Traffic Flow Management System (TFMS) and its decision-support tools do not take into account the stochastic nature of the predictions.

Taking into account uncertainty in predictions and moving from deterministic to probabilistic TFM is an important part of the Next Generation Air Transportation System (NextGen) program that will help TFM specialists make better and more realistic decisions.

During the past few years, the concept of probabilistic TFM has matured, with research producing evidence of the potential benefits of the transition from current deterministic to probabilistic TFM. Many organizations, such as FAA, NASA, MITRE, Metron Aviation, Volpe Center and others, are currently engaged in research on probabilistic TFM. Research is in two major directions: uncertainty in traffic demand predictions and uncertainty in predicting the capacity of NAS elements. Each of those directions addresses two major elements of the NAS: airports and airspace. Although airports and airspace are equally important in the NAS and are generally subject to similar sources of uncertainty in traffic demand and capacity predictions, there are substantial differences in measuring both demand and capacity in airports and airspace. Both demand and capacity are better defined and measured for airports than for en route sectors. Moreover, because of the differences in measuring traffic demand in airports and in sectors in the current TFMS, the TFM specialists noticed that demand predictions in sectors are more uncertain, more volatile and less reliable than those for airports.

Several publications have presented the concepts and potential applications of probabilistic TFM as well as modeling and benefit analysis (see [1]-[18]). A general concept of probabilistic TFM and representation of uncertainty in air traffic demand and capacity predictions for identifying and managing congestion in NAS elements are described in [9]-[12]. Research results presented in [13]-[15] provide an important contribution to probabilistic TFM describing a constructive approach to incorporating probabilistic weather forecast into probabilistic TFM, as well as design of the modeling tool for evaluating the TFM strategies. References [16]-[18] describe a methodology for sequential decision-making approach to probabilistic TFM that makes it possible to update TFM strategies in accordance with updated probabilistic forecasts on demand and capacity.

Probabilistic TFM concepts will evolve to everyday TFM practice only if the decision support tools for probabilistic TFM are built on realistic data that reflects and quantifies existing uncertainty in the aviation system, namely uncertainty in predicting traffic demand and capacity for strategic TFM and tactical Air Traffic Control (ATC). This is why thorough data analysis along with analytical tools are needed to analyze the sources of uncertainty, characterization of uncertainty and estimation of its parameters, the ways of reducing uncertainty and incorporating uncertainty into the TFM decision-making process. Our research has been focused on developing a methodology that allows for a quantitative representation of uncertainty in air traffic demand predictions for NAS elements.

This report continues our previous research on the estimation and characterization of uncertainty in aggregate traffic demand predictions in the FAA Traffic Flow Management System (TFMS). The major goal of our research in this direction is to make CORRECT predictions of traffic demand (i.e., what would have happened without ATC interference) to better identify potential congestion at NAS elements and improve Monitor/Alert functions to help TFM specialists to make decisions on triggering traffic management initiatives (TMIs), such as ground delay programs (GDPs) or reroutes.

Our previous research [1]-[3] was focused on analysis of accuracy of 15-minute aggregate traffic demand predictions in TFMS and on the development of a regression model aimed at improving the accuracy and stability of those aggregate predictions. The regression model improved both the accuracy of demand predictions and the stability and accuracy of TFMS Monitor/Alert. The analysis dealt with aggregate demand counts and did not explicitly consider uncertainty in the prediction of events for individual flights. The next step in our research, reported in [4] and [5], was focused on analyzing uncertainty in predictions of airport arrival times for individual flights and developing a methodology for translating the uncertainty in estimated time of arrival (ETA) for individual flights into uncertainty in aggregate 15-minute traffic demand predictions for arrival airports. As a result, the methodology was developed for probabilistic traffic demand predictions at airports with quantitative characteristics of uncertainty in the predictions.

This report is devoted to probabilistic representation and characterization of uncertainty in predicting traffic demands in en route sectors. A basic idea and a methodology for analyzing probabilistic properties of aggregate traffic demand counts via characteristics of uncertainty in individual flight times prediction was presented in [6]. A statistical simulation model for analyzing sector load uncertainty on 15-minute basis was described in [7]. The probabilistic representation includes a mean (expected) value of a one-minute demand counts and the area of uncertainty around the mean value restricted by specific percentiles (e.g., uncertainty area between 25th and 75th percentiles). The report presents a methodology and results of analysis that demonstrate how uncertainty in prediction of arrival times for individual flights translates into uncertainty in prediction of aggregate traffic demand counts in sectors. Like our previous work, it is based on a statistical analysis of current TFMS data. The translation of the characteristics of uncertainty in TFMS predictions of times for individual flights into characteristics of uncertainty in predictions for aggregate demand counts is a challenging problem. However, it is much more complicated for sectors than for airports because of the differences in measuring traffic demands for these NAS elements.

The particular properties of traffic demand in the en route sectors that are different from traffic demand at airports as follows:

1. Traffic demand in sectors is based on one-minute aggregate counts vs. 15-minute arrival or departure counts at airports.
2. Current TFMS determines traffic demand in sectors on a 15-minute basis and considers the maximum one-minute count within a 15-minute interval as the traffic demand for the sector for entire 15-minute interval, while a 15-minute traffic demand for airports includes all flights with ETAs or ETDs (Estimated Time of Departure) within the 15-minute interval.
3. The measurement of sector demand on a one-minute basis, vs. 15-minute counts for airport demand, makes one-minute sector demand counts especially uncertain and unstable.
4. Significant fractions of one-minute demand counts in adjacent one-minute intervals in a sector might contain the same flights while, at airports, adjacent 15-minute intervals contain different flights.
5. Because successive one-minute counts in a sector contain many of the same flights, the correlation in a time series of successive one-minute predictions of flight counts within a sector is expected to be much higher than in a series of 15-minute traffic demand predictions at airports.

The purpose of this report is to

- Analyze the errors in sector entry time predictions for individual flights and characterize the accuracy of TFMS flight-by-flight prediction data
- Use these characteristics of uncertainty to determine the probability for an aircraft to enter a sector within a specific one-minute interval and to be in a sector at any given time
- Use these probabilities of individual flights' events for probabilistic predictions of aggregate one-minute counts of flights crossing sector boundaries and one-minute traffic demand counts of flights in sectors.

The report has been organized as follows.

- Section 2 describes the data that was used in this study.
- Section 3 presents an empirical analysis of TFMS one-minute sector count predictions
- Section 4 presents an empirical analysis of flight-by-flight sector entry and occupancy times.
- Section 5 uses the probability distributions of errors in predictions of sector entry times of individual flights to derive probabilistic predictions of one-minute counts of flights entering a sector as well as one-minute traffic demand counts within a sector.. From this, probabilistic demand predictions for sector counts, including mean demand and the area of uncertainty around the mean demand, are obtained.
- Section 6 concludes the report by summarizing the findings.
- The detailed analytical approach and results for probabilistic traffic demand for sectors are presented in Appendices A, B, and C

2. Examined Flight-by-Flight Data

TFMS continuously updates information on the status of each flight in the system and predicts each flight's time and location at various points along its origin-destination route. The TFMS flight list for a NAS element can be requested at any time. The flight list shows the flight's status (airborne or still on the ground), estimated time of departure (or actual departure time for an active flight), and estimated time of arrival at a NAS element, including destination airport. TFMS also collects flight-by-flight historical data that include both predictions and what actually happened.

From Friday April 10 to Thursday April 16, 2009, TFMS List Requests were repeatedly run for the following sixteen en-route sectors: ZBW02, ZBW09, ZBW17, ZBW20, ZBW46, ZID82, ZID83, ZID86, ZLC06, ZLC16, ZMP20, ZOB57, ZOB67, ZOB77, ZSE14, ZTL43. Figure 2-1 shows their approximate locations, noting that some may have been combined with other sectors at the time the snapshot for this figure was created.

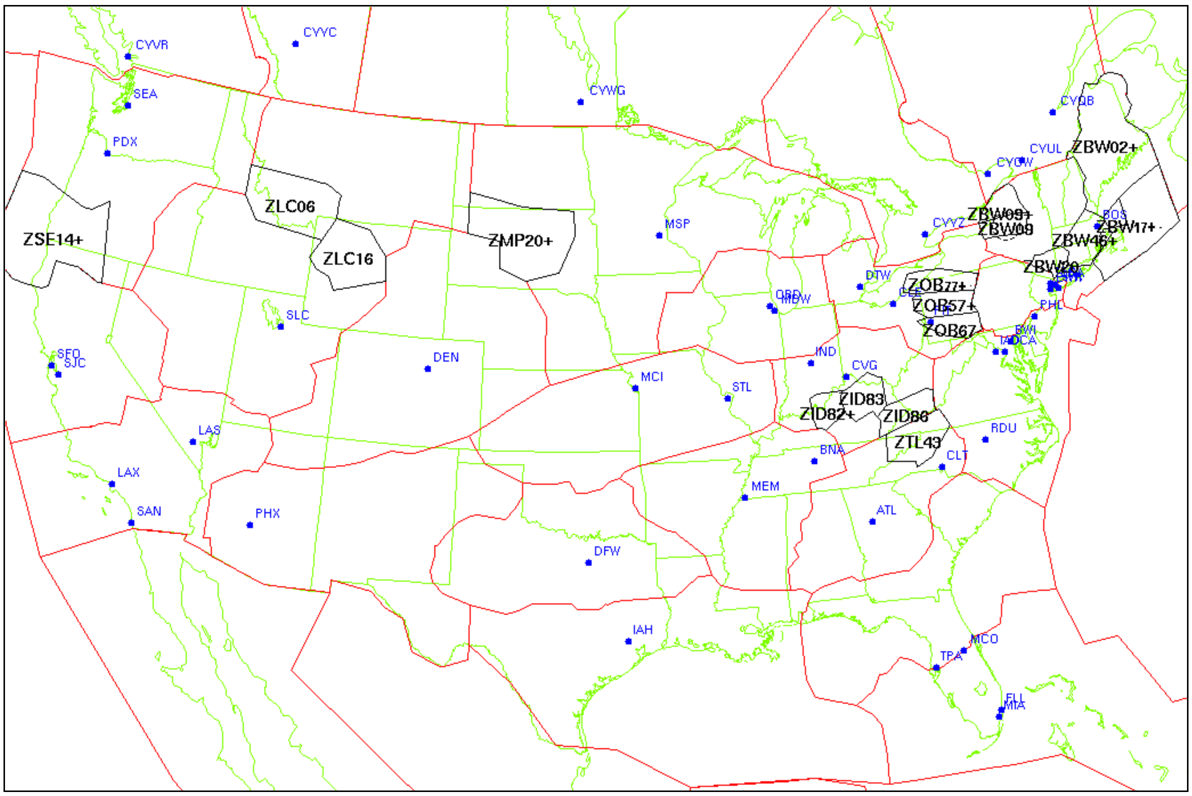

Figure 2-1 Examined Sectors

These repeated list requests, which were run once every 15 minutes, generated flight-by-flight predictions for airport arrival, sector entry and sector exit. On 4/10 through 4/12, data was gathered for 1230Z through 2230Z each day. On 4/13 through 4/16, data was gathered for 1130Z through 2230Z each day. For the 16 sectors on April 10-16, there were approximately 39,000 flights, with 834,000 observations.

Statistical analysis of flight data was performed to characterize uncertainty (errors) in flight event predictions. Separate analysis was conducted for active (airborne) flights and for proposed flights (that are still on the ground) to characterize the difference in the accuracy of predictions depending on the flight's status.

3. Analysis of TFMS One-Minute Traffic Demand Predictions

TFMS uses minute-by-minute sector count predictions when determining whether the sector count is above the monitor alert parameter (MAP). The number of flights in a sector is calculated for each minute; then an alert is raised if that number exceeds the MAP. Typically, 15-minute intervals are used when presenting alerts, so that an alert is raised if the maximum of the 1-minute predictions in that 15-minute interval exceeds the MAP.

Both sector counts (number of flights in a sector during a minute) and sector entries (number of flights entering a sector in one minute) are of interest to air traffic managers. The next section reviews minute-by-minute counts. Section 3.2 reviews sector entries.

3.1 Sector Count Predictions

Figure 3-1 is a bubble chart of these predictions for a one hour look-ahead time. The x-axis has the forecast count and the y-axis, the actual count. The area of each bubble is proportional to the number of observations for a given x-y pair. If predictions were perfectly accurate, all of the bubbles would lie on the 45-degree line. In reality, although the predictions tend to be fairly accurate on average (with a roughly equal number of observations above and below the 45-degree line), there are significant positive and negative errors.

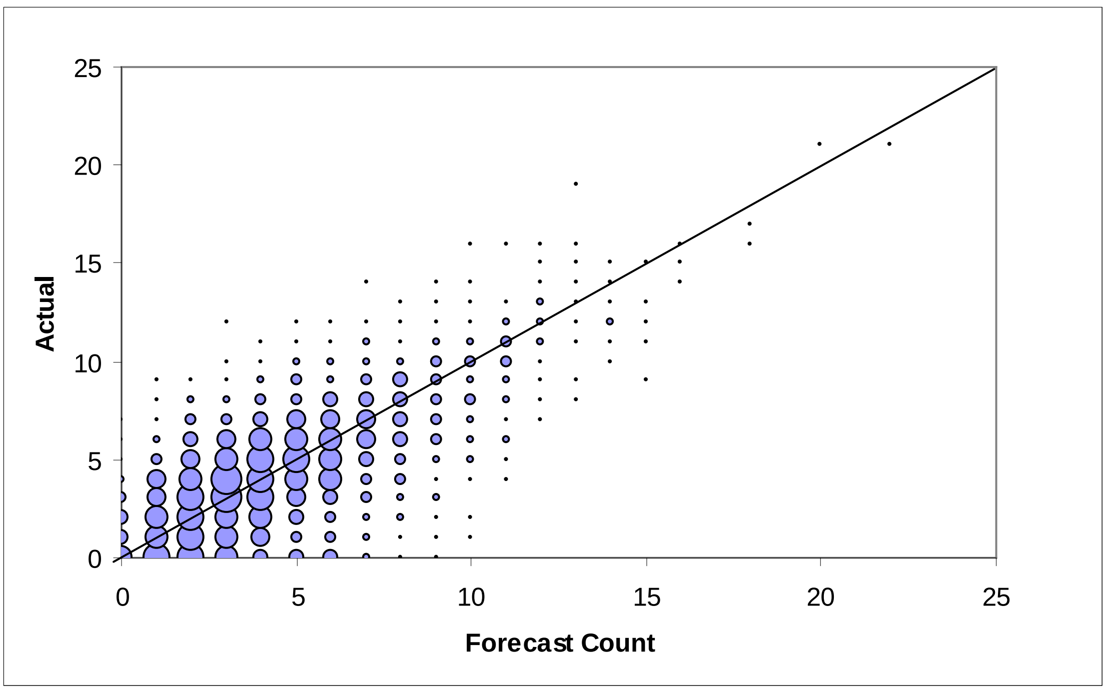

Figure 3-1 Sector Count Predictions

The next six figures illustrate the minute-by-minute volatility of the counts, using data from April 2009. All of these figures are predictions, with a look-ahead time of approximately 90 minutes. The first figure of each pair illustrates the minute-by-minute predictions over a 12-hour period. The second figure zooms in on a two- hour subset of that day, so that the minute-by-minute fluctuations are visible.

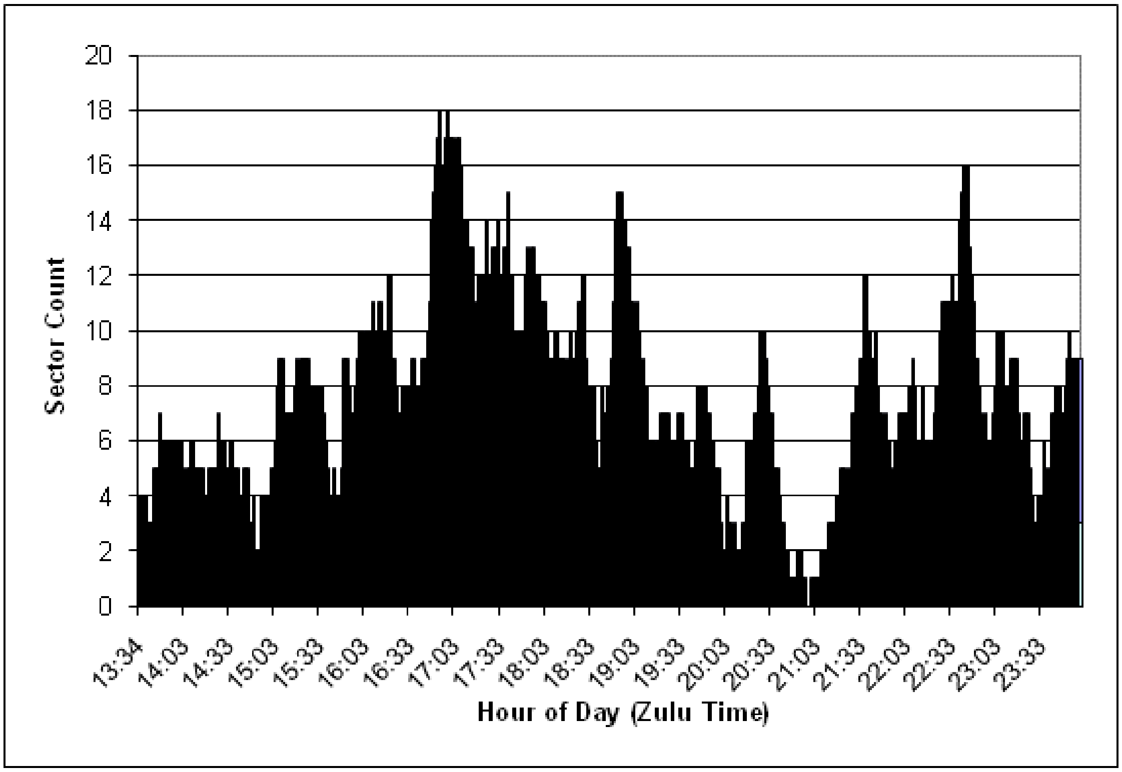

Figure 3-2 ZBW02 on 4/10/2009

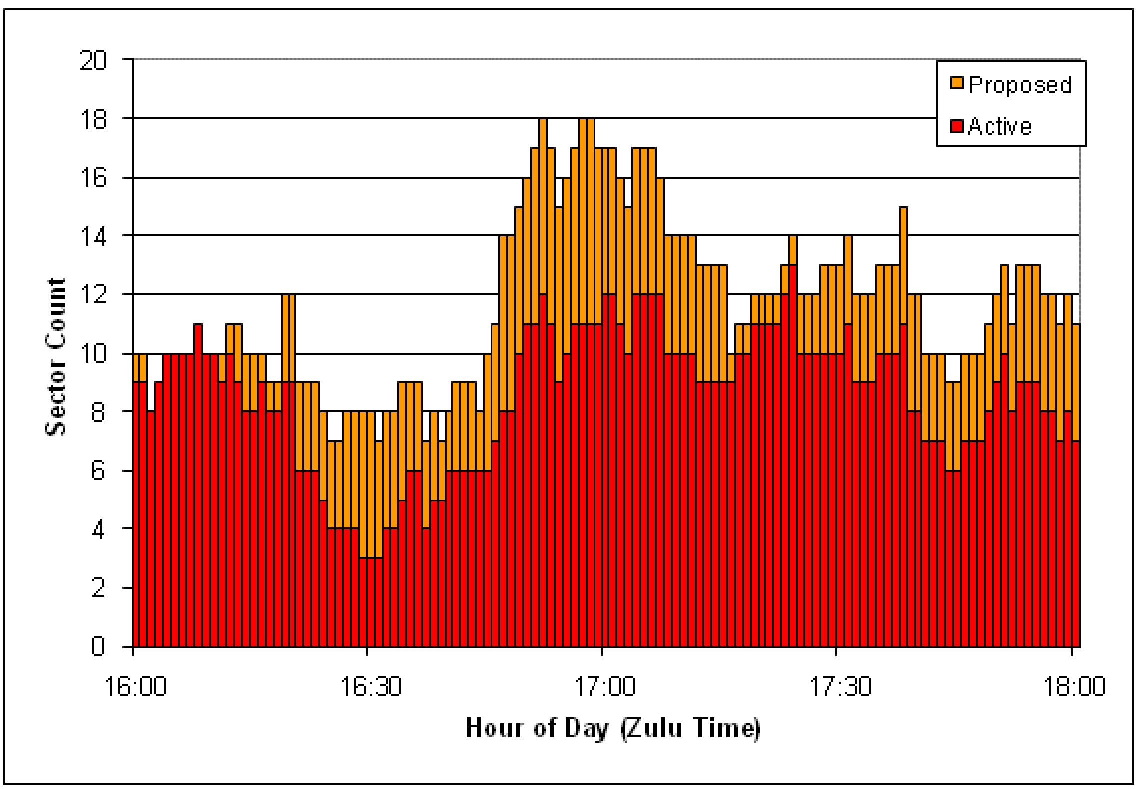

Figure 3-3 ZBW02 on 4/10/2009 1600 – 1800Z

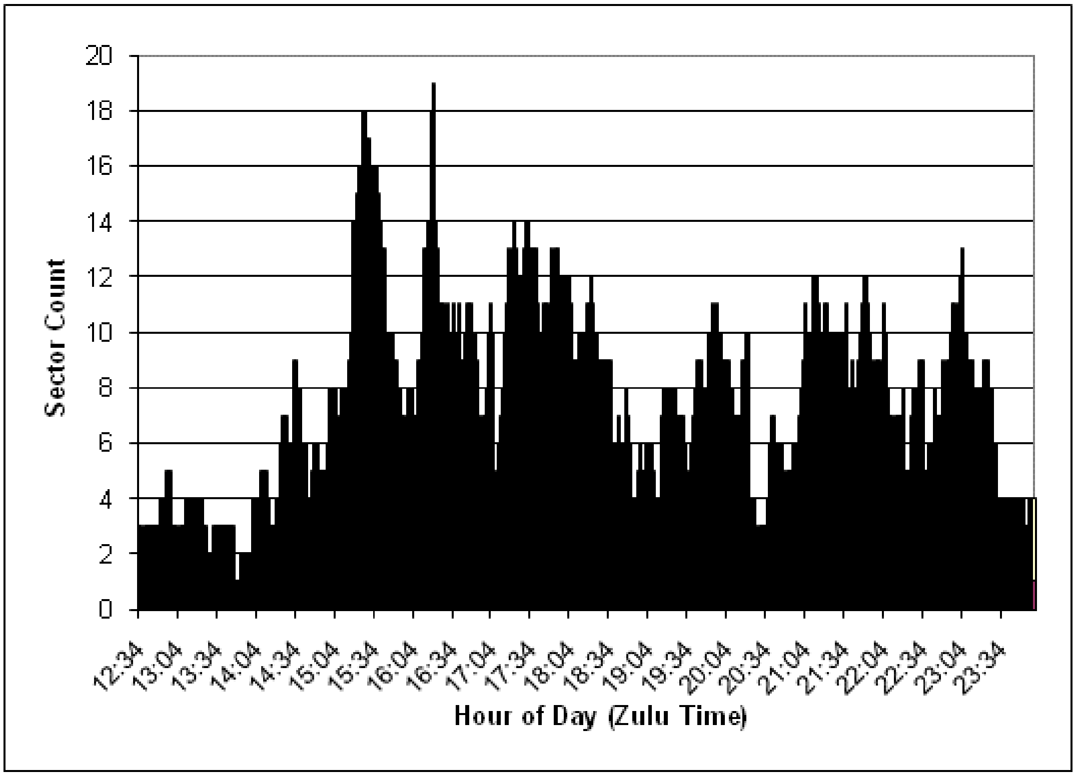

Figure 3-4 ZMP20 on 4/13/2009

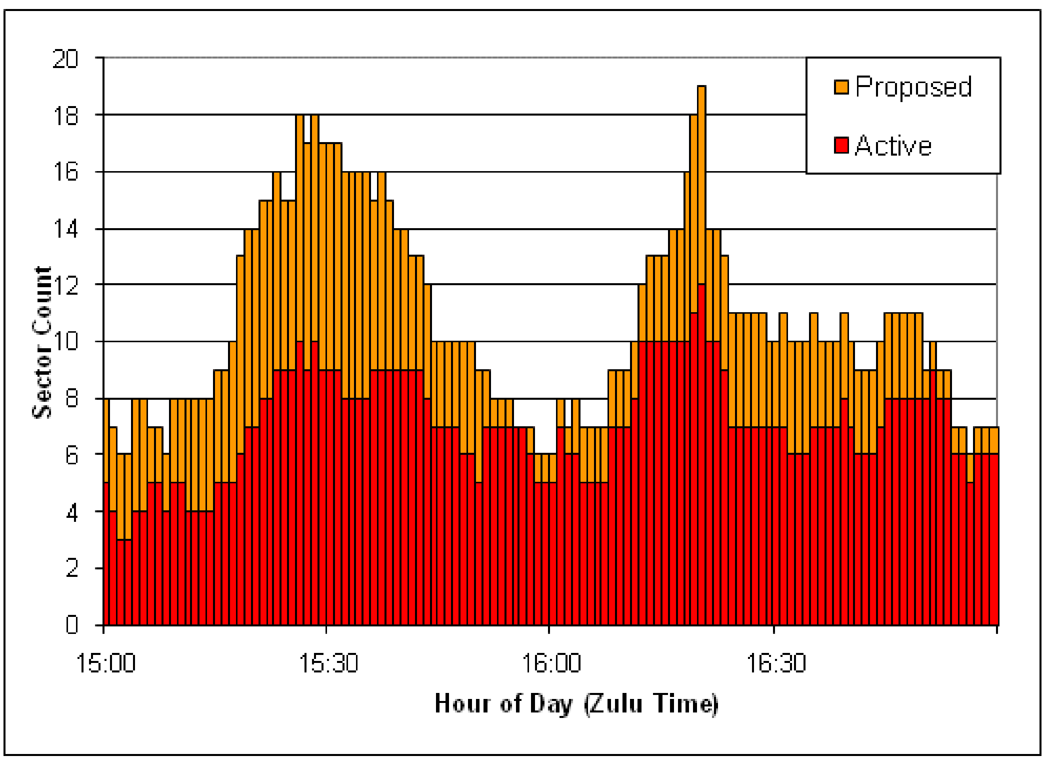

Figure 3-5 ZMP20 on 4/13/2009 1500Z - 1700Z

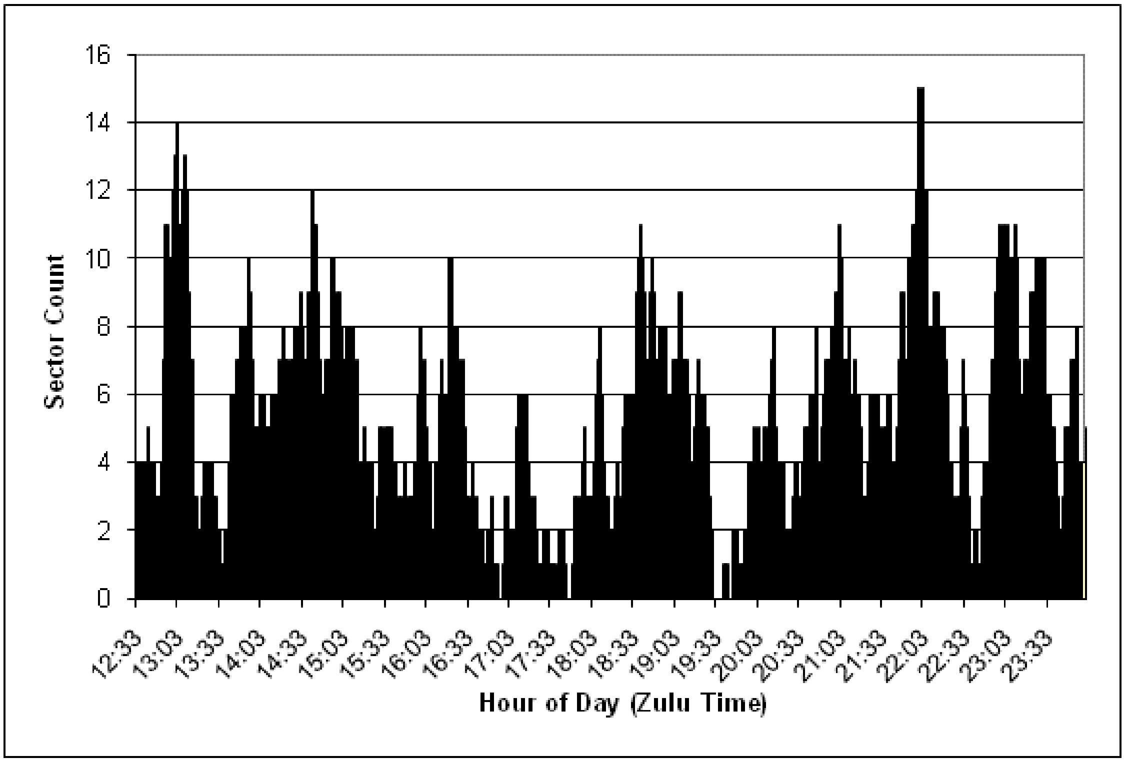

Figure 3-6 ZOB57 on 4/13/2009

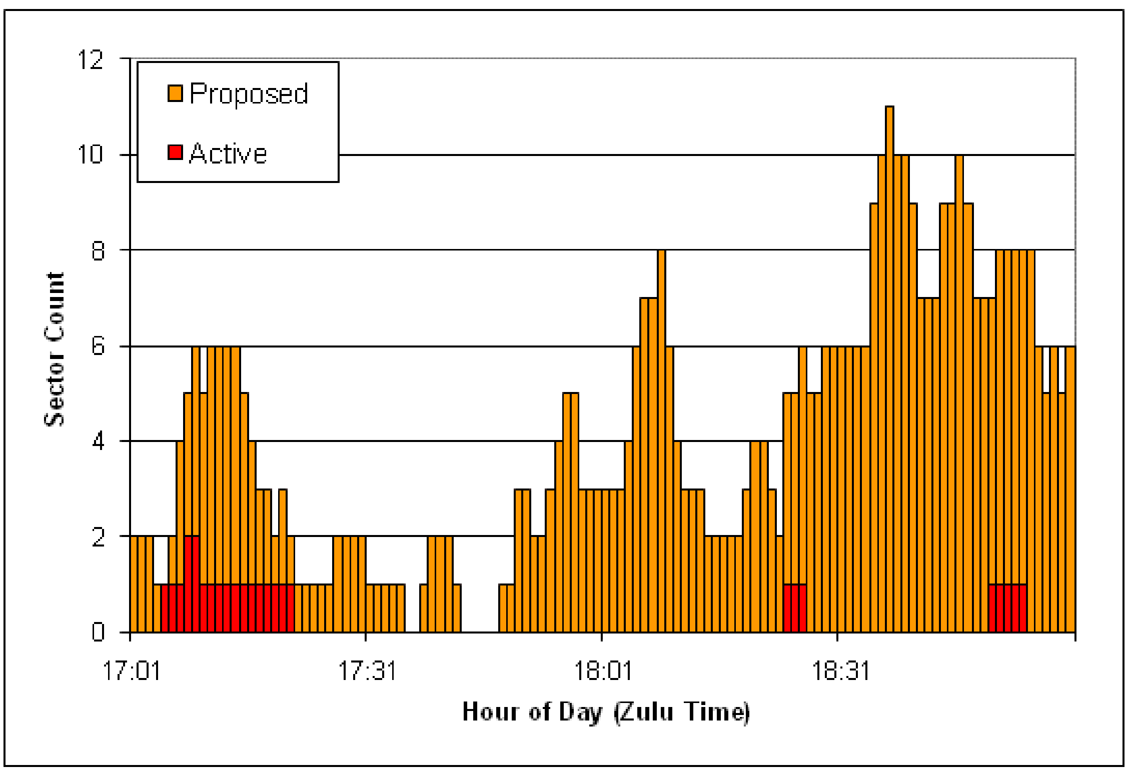

Figure 3-7 ZOB57 on 4/13/2009 1700Z - 1900Z

The autocorrelation between counts for minute x and minute x-1 is extremely high: 0.96. The distribution of these count differences is shown in Table 3-1and in Figure 3-8.

Table 3-1 Change in Sector Count from One Minute to the Next

Count – Prior minute Count	Number of Observations
-5	5
-4	59
-3	410
-2	2333
-1	10824
0	31568
1	11050
2	2313
3	403
4	54
5	7

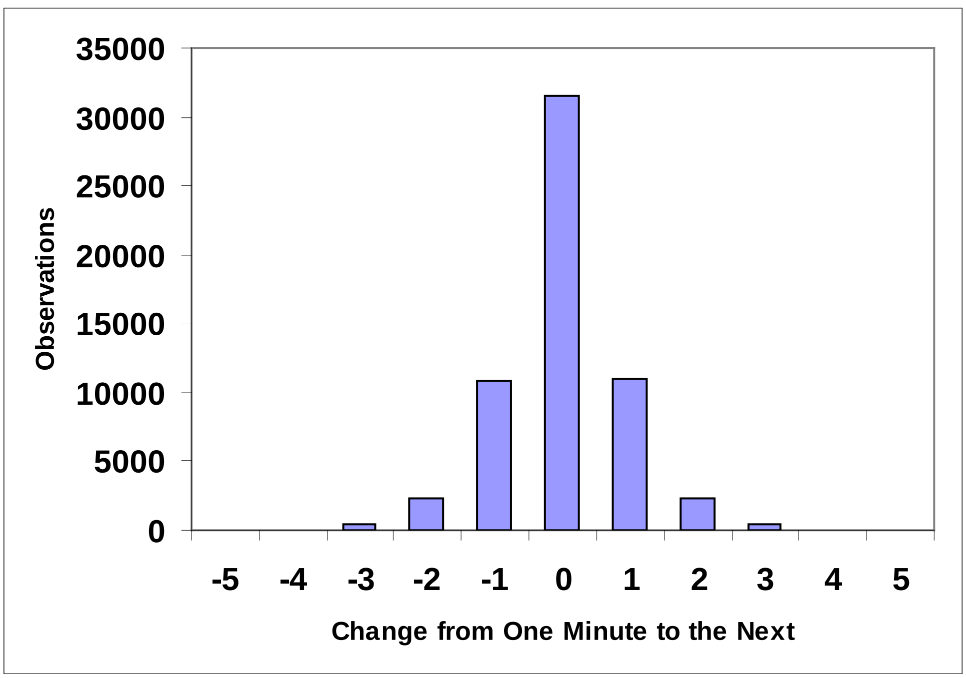

Figure 3-8 Change in Sector Count from One Minute to the Next

Finally, Table 3-2 and Figure 3-9 show the correlations between counts taken k minutes apart, where k ranges from 1 to 15. As expected, the correlation is higher for low k and is also higher for those sectors which have longer occupancy times (time that an aircraft spends in a sector). For example, the time-in-sector for ZMP20 (light blue line) is often 30 minutes, so the correlation remains high. The time-in-sector for ZOB57 tends to be much shorter (5 - 20 minutes), so the autocorrelation decreases more rapidly as the number of minutes k increases. Time-in-sector will be further discussed in section 4.2.

Table 3-2 Autocorrelation Between Sector Counts Taken k Minutes Apart

k	All	ZBW02	ZMP20	ZOB57
1	0.962	0.977	0.985	0.928
2	0.924	0.955	0.970	0.859
3	0.886	0.931	0.953	0.784
5	0.818	0.883	0.919	0.667
10	0.692	0.765	0.835	0.478
15	0.607	0.655	0.754	0.335

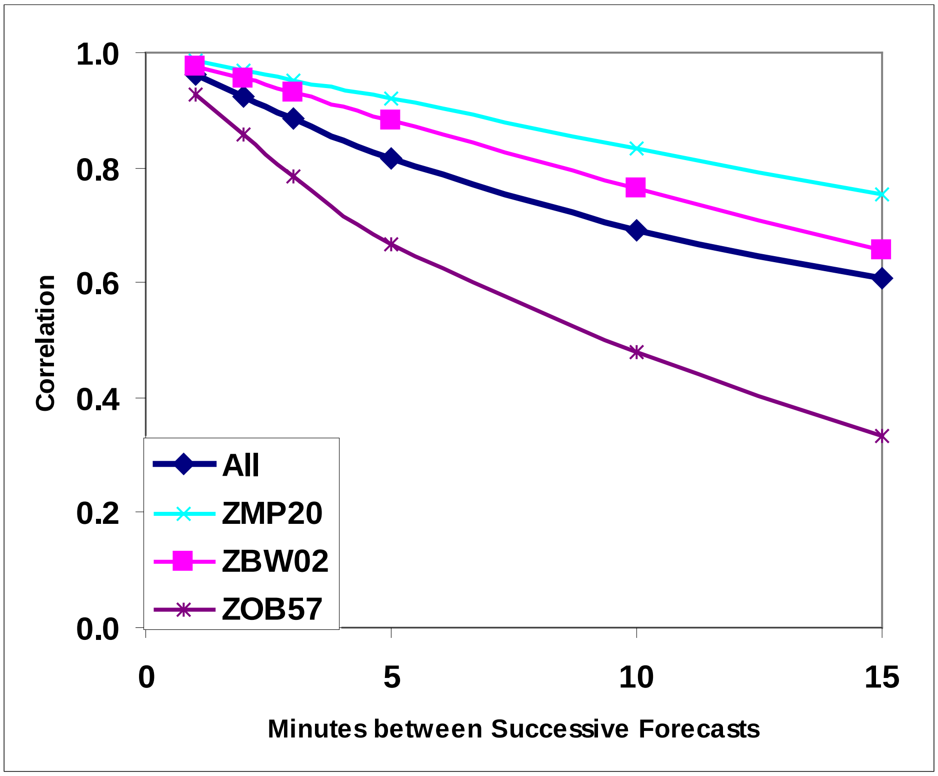

Figure 3-9 Autocorrelation Between Sector Counts Taken k Minutes Apart

3.2 Correlation Between Sector Entry Predictions

In addition to the number of flights in a sector during a particular minute, discussed earlier, the number of flights entering a sector during a minute is also of interest for two reasons. First, predictions of sector entry are of a special interest to air traffic managers and controllers because this signals the need to hand off the flight from one sector to the next. Second, sector entry counts explicitly affect traffic demand counts within a sector.

One aspect of the one-minute sector entry counts that will affect probabilistic predictions of sector demand is a small correlation from one minute to the next. One would expect little correlation, because the one-minute intervals contain completely different flights, each of which has random errors in sector entry time prediction not correlated with the prediction errors in other flights. To check this supposition, a data analysis was performed.   The analysis revealed very little correlation from one minute to the next in the number of flights entering a sector.

4. Analysis of TFMS Data on Flight-by-Flight Sector Entry and Occupancy Times

This section provides a descriptive statistical analysis of the accuracy of flight-by-flight TFMS predictions that will be used in later sections. The figures of merit used are as follows:

- The accuracy of predicted flight arrival times at a NAS element, which is measured as an error (predicted – actual) in the arrival time. For the sector data, this is the (predicted – actual) sector entry time. The error is positive if the flight is early and negative if it is late.
- The prediction accuracy of time-in-sector.

4.1 Flight Arrival Times at a Sector

Our prior report in 2007 presented analysis of sector entry times, based on 2007 data. In 2009, additional data was collected and the same analysis performed. Table 4-1 shows the distributions of **(**predicted – actual**)** sector entry time, divided into the following time buckets.

- Less than -180 minutes (flights more than 3 hours late)
- -180 to -61 minutes (flights 1 – 3 hours late)
- -60 to -15 minutes (flights less than 1 hour late)
- On time (flights that are within 15 minutes of the predicted arrival time)
- 15 – 60 minutes (flights less than 1 hour early)
- 61 – 180 minutes (flights 1 – 3 hours early).

No flights were more than 3 hours early.

The LAT column shows Look-Ahead Time (LAT) range for predictions. In the Flight Status column, letter A denotes active, i.e., airborne flights, and letter P denotes the proposed flights that are still on the ground at origin airports. The results in this table are not surprising. Flights in the air tended to be early, while flights on the ground tended to be late due to departure delays. For example, Table 4-1 shows that approximately 2 to 24 percent of active flights enter sectors minutes earlier than predicted and only 1 to 3 percent of active flights enter sectors later than predicted. However, more than 30% of the flights still on the ground cross sector entry boundaries later than predicted, and roughly 6 to 10% enter sectors earlier than predicted. Note that "early" and "late" here refer to the sign (positive or negative) of the (predicted – actual) time, and is related to the latest known flight plan. For example, if the original (schedule) flight plan resulted in estimated sector entry time of 1300Z, but its origin airport departure was delayed by two hours so that its updated flight plan results in a new estimated sector entry time of 1500Z and the flight actually enters the sector at 1435Z, it would show up in this table as "early," even though according to the original schedule, the flight is nearly two hours late.

Table 4-1 Distribution of Sector Entry Time Errors

Month	LAT	Flight Status	Number of Obs.	Time Bucket					
				Late (minutes)			On Time[1]	Early (minutes)	
				Over 180	61 – 180	15 - 60		15 - 60	61-180
April 2007	1-2 hr	A	3194	0.1%	0.2%	1.0%	94.7%	3.7%	0.2%
June 2007	1-2 hr	A	5690	0.0%	0.2%	1.6%	96.5%	1.6%	0.1%
April 2007	2-3 hr	A	1853	0.0%	0.0%	2.9%	80.2%	16.2%	0.6%
June 2007	2-3 hr	A	4204	0.0%	0.0%	1.8%	74.6%	23.2%	0.3%
April 2007	1-2 hr	P	10474	0.6%	3.9%	26.0%	63.1%	6.2%	0.2%
June 2007	1-2 hr	P	10691	0.9%	5.8%	26.0%	60.9%	6.4%	0.1%
April 2007	2-3 hr	P	4962	0.5%	5.3%	28.2%	56.5%	8.5%	0.9%
June 2007	2-3 hr	P	5443	1.2%	7.3%	28.6%	52.5%	9.4%	1.0%
April 2009	1.5 hr	A	2322	0.2%	0.4%	2.1%	88.5%	8.2%	0.6%
April 2009	1.5 hr	P	9868	0.0%	1.6%	18.4%	73.6%	6.2%	0.2%

The bias in the estimate deserves comment. The late bias for proposed flights is understandable, since TFMS is conservative in adjusting predicted take-off times for a delayed flight. For active (airborne) flights, there appears to be a slight bias the other way, with some flights entering a sector earlier than predicted by TFMS.

Table 4-2 shows the average and standard deviation of sector entry time error, using 90-minute look-ahead time and the 2009 data. There were 17 active flights (less than 1% of the total) with reported errors greater than 90 minutes. Since it is difficult to imagine an airborne flight being this far off plan (for fuel considerations if nothing else), these 17 flights were removed as outliers. For the remaining active flights, Table 4-2 indicates a median standard deviation of approximately 8 minutes, while for proposed flights, the median standard deviation is approximately 18 minutes.

[1] "On-time" is considered to be within 15 minutes (early or late) of the predicted time.

Table 4-2 Sector Entry Time Error by Sector: Average and Standard Deviation (April 2009)

	Active Flights			Proposed Flights		
	Count	Average Error	Standard Deviation	Count	Average Error	Standard Deviation
ZBW02	491	13.4	16.2	495	-3.8	19.0
ZBW09	119	5.3	8.0	448	-1.1	16.6
ZBW20	209	0.7	6.5	1209	-7.2	18.9
ZBW46	110	6.1	8.2	501	-6.7	22.1
ZID82	60	-0.5	6.5	515	-5.2	19.3
ZID83	70	-0.6	5.2	641	-4.5	18.5
ZID86	108	-3.9	9.5	1096	-4.2	17.7
ZLC06	130	-1.8	5.4	463	-2.5	17.9
ZLC16	109	-0.2	6.1	263	-4.3	21.3
ZMP20	388	-1.3	4.1	379	-0.1	16.4
ZOB57	73	1.7	14.7	778	-4.1	18.4
ZOB67	93	0.0	10.0	938	-2.0	18.3
ZOB77	60	2.5	9.8	237	-4.1	18.2
ZSE14	81	-5.0	11.6	520	-3.4	12.3
ZTL43	106	-0.9	7.2	970	-4.4	18.8
Total	2207	3.0	11.8	9453	-4.1	18.4

The next two figures show the distributions of sector entry time predictions for April 2009 data with 90 minute look-ahead time. Separate curves are shown for active (airborne) and proposed (still on the ground) flights. Both the probability densities (Figure 4-1) and cumulative distributions (Figure 4-2) are shown.

The 2007 data in our previous report showed that the range of sector entry errors (measured by standard deviation) was between 5 and 32 minutes, depending on the flight status and the sector. It was in the 10 minute range for active flights and the 25 minute range for proposed flights.

The 2009 data, as shown in Figures 4-1, 4-2 and Table 4-2, tells a similar story:

- For active flights, the distributions of prediction errors are asymmetric with a heavier right-hand tail; i.e., the distribution is more biased toward earlier arrivals than predicted.
- For the proposed flights, the distributions of prediction errors are asymmetric with heavy left-hand tails, which reflect the tendency for flights on the ground to, on average, enter sectors later than predicted.
- The probability density plot (Figure 4-1) shows a number of outliers (the "bumps" in the plots at values far from zero).
- In 2009, the standard deviation of error continued to be substantially lower for active than for proposed flights. It was in the 8 minute range for active flights, and in the 18 minute range for proposed flights.
- The mode for both distributions is very close to zero. For the most part, the prediction errors are caused by uncertainty in the system, not by an incorrect prediction of central tendency.

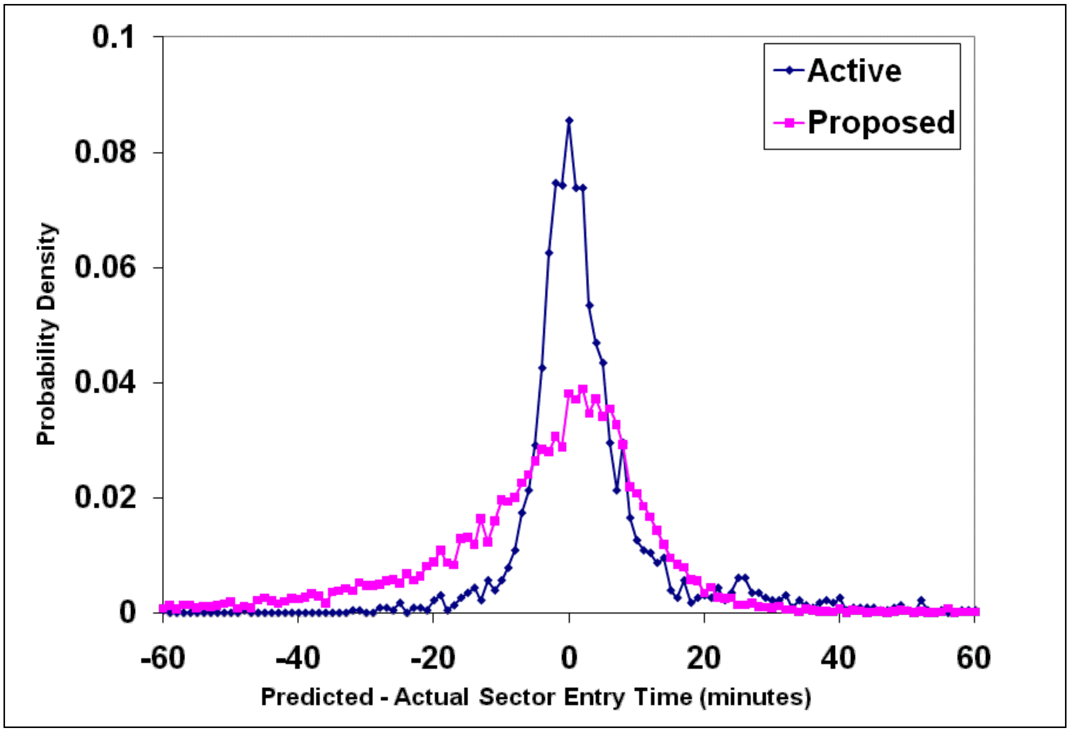

Figure 4-1 Distribution of Errors in Sector Entry Time Predictions: April 2009 Data

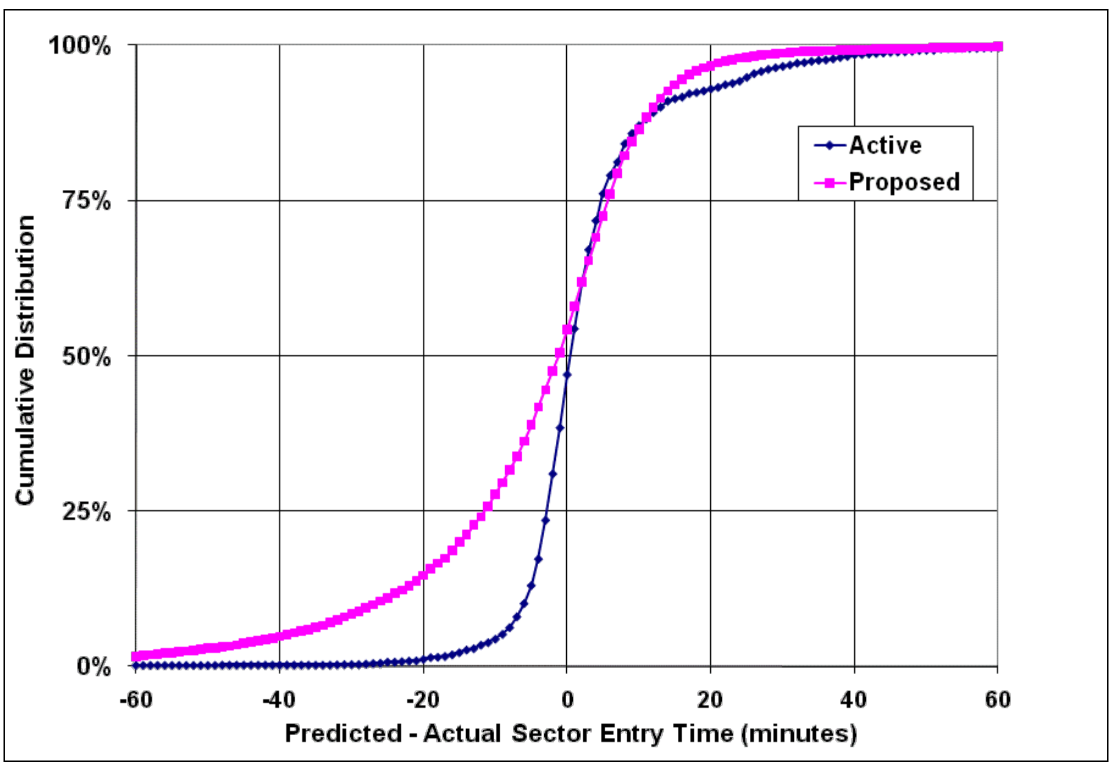

Figure 4-2 Cumulative Distribution of Errors in Sector Entry Time Predictions: April 2009 Data

4.2 Time-In-Sector

Sectors vary in size and in the manner flights traverse them. As a result, there are significant differences among the sectors in time-in-sector for a flight.

Figure 4-3 shows the time-in-sector distribution for three sectors. For these three sectors, the distributions are simple, with one primary peak, and perhaps one secondary peak. For other sectors (Figure 4-4), the distributions are more complex, with multiple peaks.

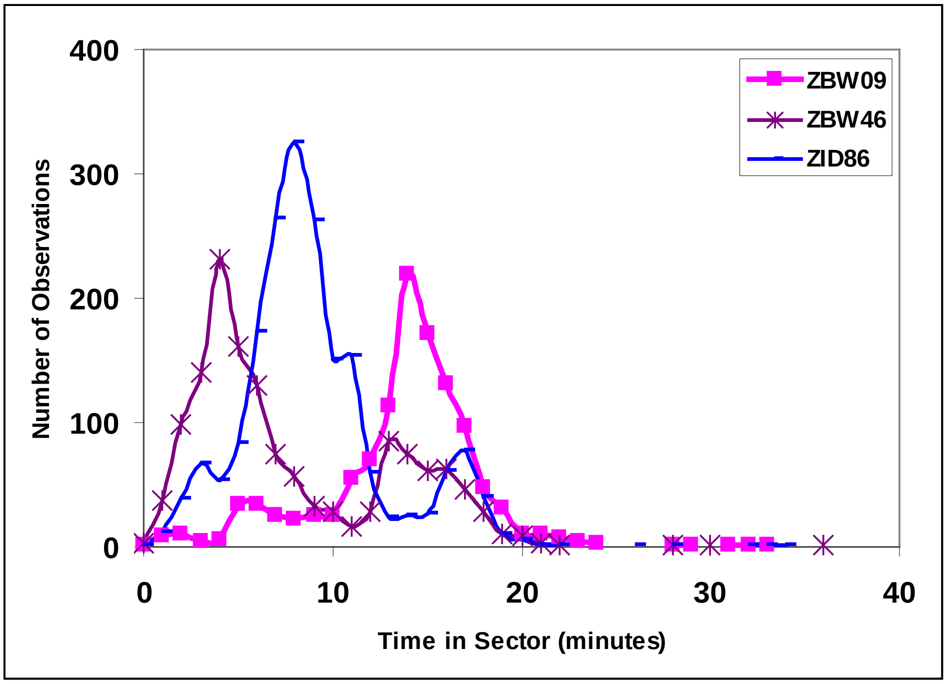

Figure 4-3 Time-in-Sector Distributions

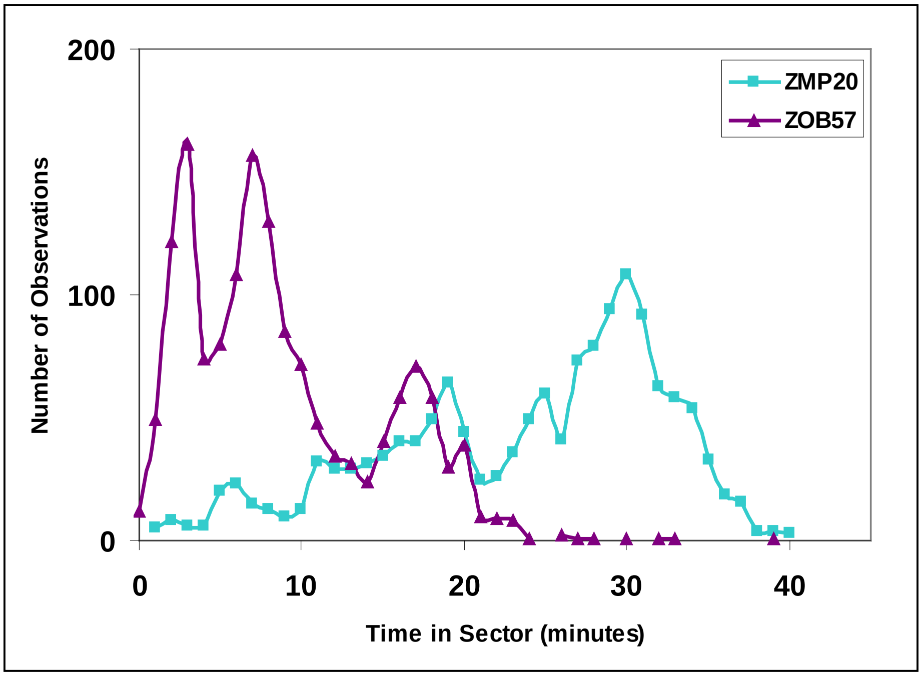

Figure 4-4 Time in Sector: ZMP20 and ZOB57

Two aspects of this chart should be noted. First, some sectors have much longer in-sector times than others. For example, ZMP20 is a large sector, with typical in-sector times of 30 minutes. ZOB57 is a much smaller sector, with smaller in-sector times. Second, some of the sectors have multi-modal distributions, with two or more peaks. For example, most traffic is in ZOB57 for 3 – 8 minutes, but a few flights are present for 15-20 minutes. The next two figures illustrate why this might be. Figure 4-5 shows the flights whose routes include ZMP20. These flights appear to be in an east-west flow, traversing the 300-mile length of the sector. Flights whose routes include ZOB57 (Figure 4-6) appear to be in a mix of flows. Many are east-west (perhaps traversing a corner of the sector) while others are north-south. Furthermore, ZOB57 is a much smaller sector than ZMP20.

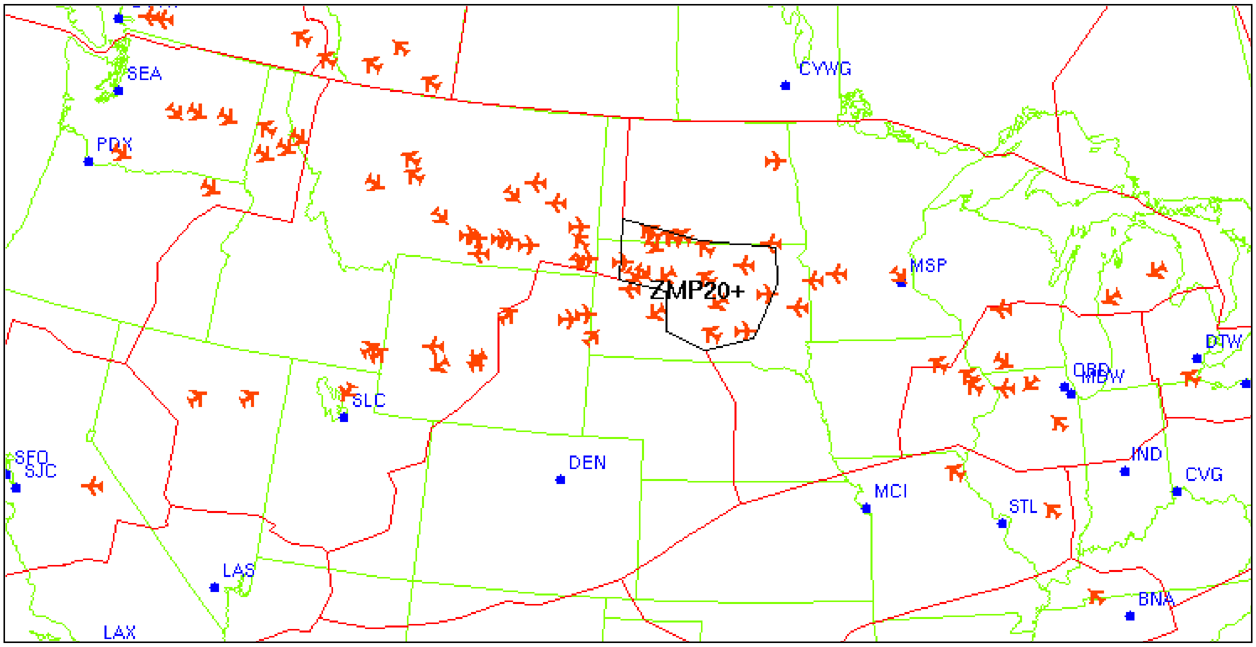

Figure 4-5 Flights Whose Routes Include ZMP20

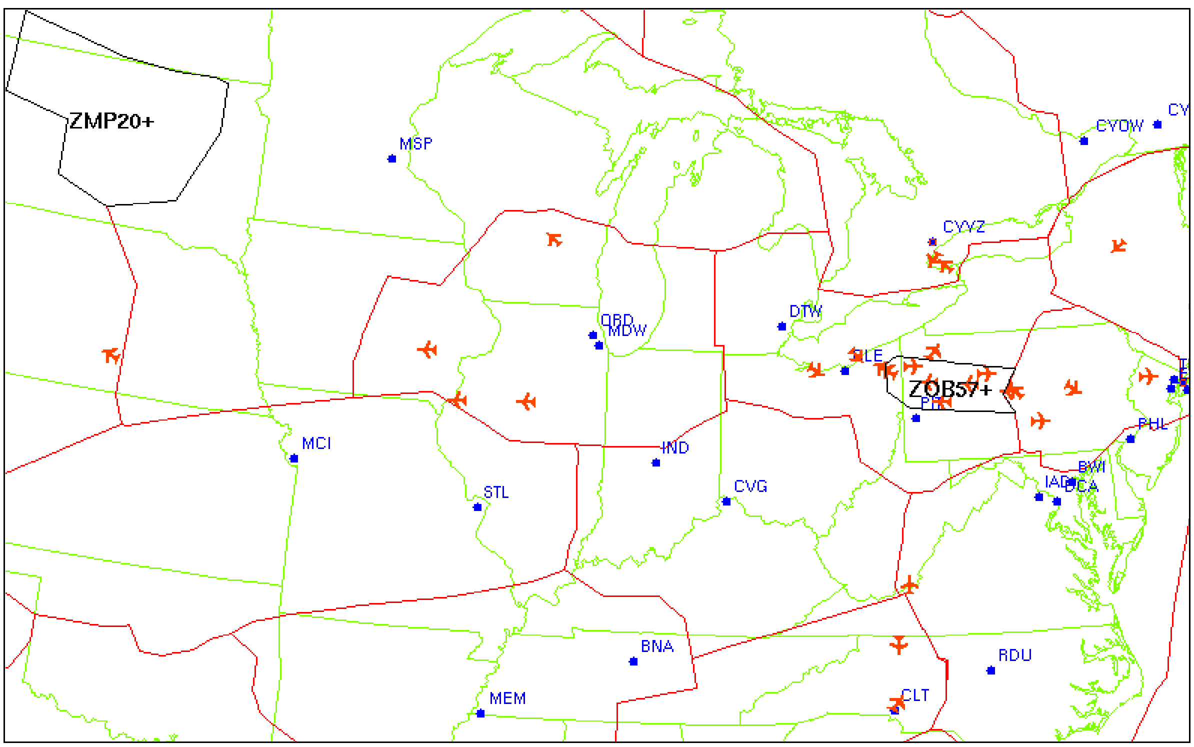

Figure 4-6 Flights Whose Routes Include ZOB57

The prediction accuracy for time-in-sector is much better than for sector entry time, and is not much different for active and proposed flights (Figure 4-7 and Figure 4-8). After outliers (time-in-sector errors greater than 15 minutes) were removed, the standard deviation of the time-in-sector error was 4.1 minutes for active flights and 3.7 minutes for proposed flights.

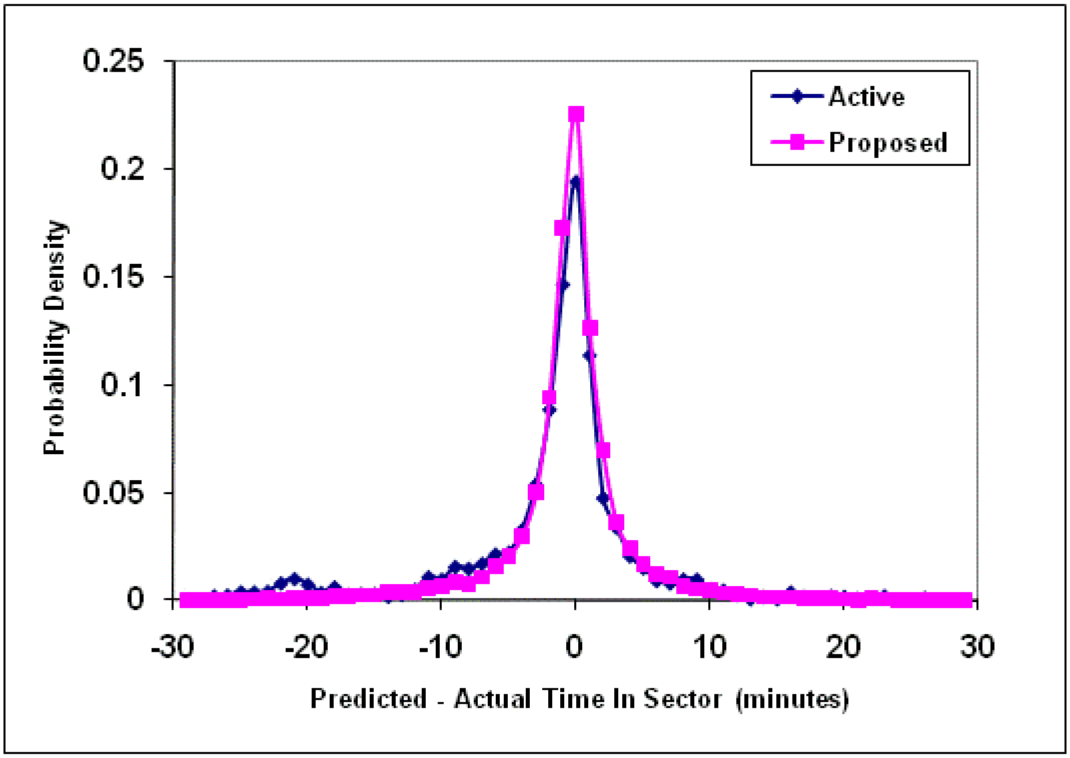

Figure 4-7 Probability Density: Time in Sector (April 2009)

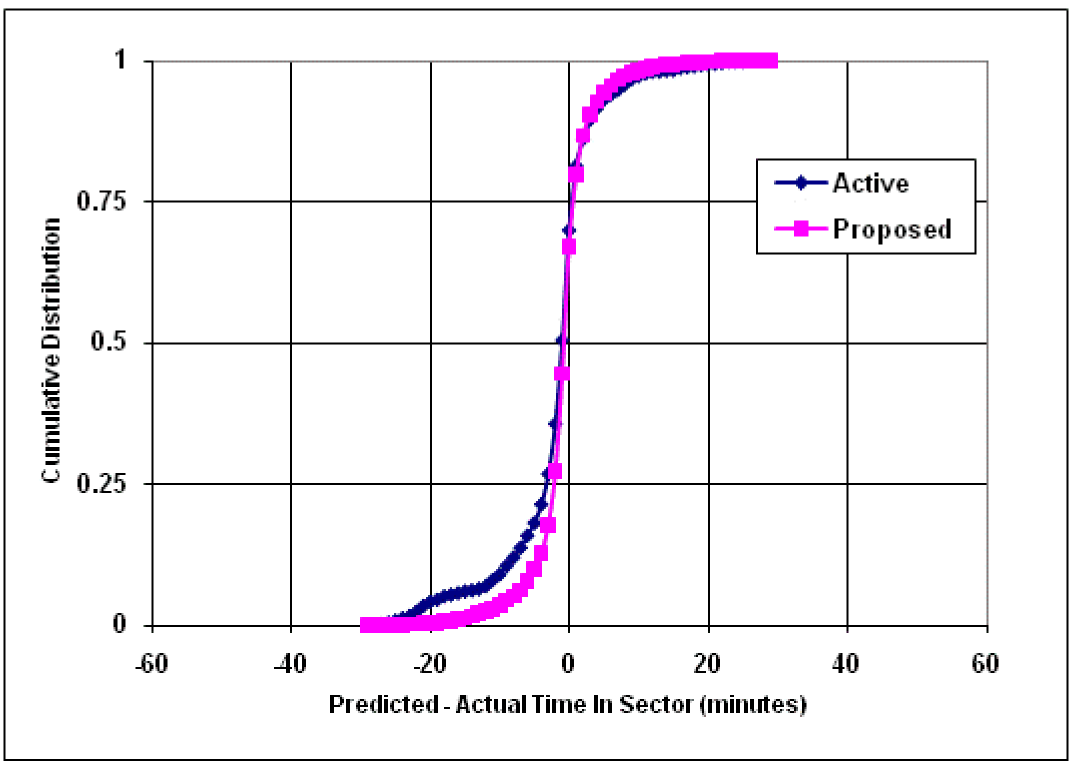

Figure 4-8 Cumulative Distribution: Time in Sector (April 2009)

Figure 4-9 is a bubble chart of the time-in-sector predictions for a 1.5 hour look ahead time. The x-axis has the forecast time-in-sector and the y-axis the actual time-in-sector. The area of each bubble is proportional to the number of observations for a given x-y pair. If predictions were perfectly accurate, all of the bubbles would lie on the 45-degree line.

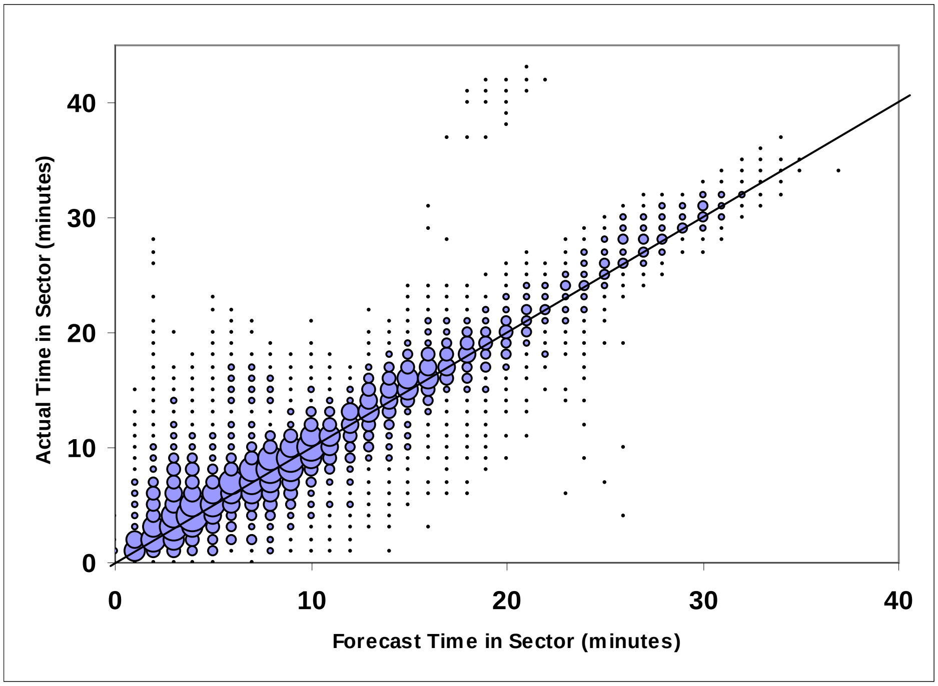

Figure 4-9 Forecast and Actual Time in Sector

4.3 Probability that a Flight will be in a Sector

TFMS predicts that a flight will be in a sector during a particular minute. The accuracy of this prediction depends on the in-sector time as well as the errors in prediction of sector entry and exit. Using empirical data on entry and exit predictions, we calculate the probabilities that a flight will actually be in a sector during a particular minute, given that it was predicted to be in that sector during that minute. Table 4-3 shows these probabilities, based on the April 2009 data. When in-sector times are higher, the probabilities are higher (for example, compare ZMP20, as sector with long in-sector time and ZOB57, a sector with short in-sector time). The probabilities are also higher for active flights than for proposed ones, because their ETA predictions are more accurate.

Table 4-3 Probability that a Flight will be in a Sector (April 2009)

	Active	Proposed
ZBW02	0.67	0.43
ZBW09	0.58	0.46
ZBW17	0.62	0.32
ZBW20	0.50	0.26
ZBW46	0.67	0.22
ZID82	0.66	0.34
ZID83	0.52	0.30
ZID86	0.64	0.30
ZLC06	0.76	0.53
ZLC16	0.83	0.50
ZMP20	0.84	0.64
ZOB57	0.51	0.32
ZOB67	0.58	0.40
ZOB77	0.48	0.28
ZSE14	0.45	0.60
ZTL43	0.50	0.26
Overall	0.70	0.41

5. Probabilistic Prediction of Sector Demand Counts

5.1 Introduction

Our previous report [4] focused on the translation in uncertainty in predictions of individual flight times of arrival at an airport to probabilistic predictions of aggregate 15-minute demand counts for airports.

In this section, we perform a similar translation for sectors. An important difference between airport and sector counts is that airport demand is measured by number of flights arriving or departing per 15-minutes or per hour, while sector demand is measured by number of flights present in the sector in one minute. The number of flights in the sector during a specific time interval (one-minute interval in TFMS) includes the flights that entered the sector during this interval and the flights that entered the sector earlier and are still in the sector. This makes the calculation aggregate demand counts for sectors more complex than for airports.

The steps taken in this section are as follows:

1. Translate a flight's time predictions and associated prediction errors into the probability for the flight to enter a sector during a particular minute.
2. Develop a probabilistic characterization for the number of flights entering a sector, based on the probabilities from step 1.
3. Develop probabilistic count predictions for the number of flights present in a sector during a particular minute (a one-minute traffic demand counts for a sector).

5.2 Probability for a Flight to Enter a Sector During a One-minute Interval

Appendix A presents a detailed analytical approach for determining probabilities for the flights to enter a sector during a one-minute interval depending on the flights' estimated sector entry times (ETAs) and probability distributions of errors in predicting sector entry times. Status of individual flights (active or proposed) is taken into account by using differences in accuracy of predicting sector entry time for active and proposed flights. As it was shown in Section 4, the prediction errors for active flights are smaller than for proposed ones.

This section presents examples of the flight probability to enter a sector during a one-minute interval.

Following the notation of Appendix A:

i is a one-minute interval of interest that is between the beginning of minute i and the beginning of minute (i+1),

k is a one-minute interval for a flight's estimated sector entry time (ETA),

$F(x)$ is a cumulative distribution function (CDF) of prediction error for sector entry time,

$F^{(a)}(x)$ is a cumulative distribution function of prediction of sector entry time for active flight,

$F^{(g)}(x)$ is a cumulative distribution function of prediction of sector entry time for proposed flights (still on the ground),

$P_{i,k}$ is a probability for a flight deterministically predicted to enter a sector during a one-minute interval k to enter a sector during a one-minute interval i,

$P^{(a)}_{i,k}$ is a probability for an active flight deterministically predicted to enter a sector during a one-minute interval k to enter a sector during a one-minute interval i.

$P_{i,k}^{(g)}$ is a probability for a proposed flight deterministically predicted to enter a sector during a one-minute interval *k* to enter a sector during a one-minute interval *i*.

Formulas (A4) – (A6) of Appendix A determine probabilities, and, respectively:

$$P_{i,k} \approx 0.5\,[F(i-k+1) - F(i-k-1)],$$

$$P_{i,k}^{(a)} \approx 0.5\,[F^{(a)}(i-k+1) - F^{(a)}(i-k-1)],$$

$$P_{i,k}^{(g)} \approx 0.5\,[F^{(g)}(i-k+1) - F^{(g)}(i-k-1)].$$

These formulas permit the calculation of probabilities for various intervals of interest, including a series of consecutive intervals, e.g., i, i+1, i+2, etc.

For the sake of simplicity, in the examples presented in this report, we assume that the prediction errors are normally distributed.

Table 5-1 shows several values of probabilities for different *k*, surrounding the interval of interest *i*. This table illustrates the probabilities for a flight to cross a sector boundary at an interval of interest *i* if it deterministically predicted to enter a sector earlier ($k < i$) or later ($k > i$) or on time ($k = i$). It also illustrates the relative significance of the probabilities depending on the time difference $|i - k|$.

The probabilities have been calculated for the Gaussian distribution $F(x)$ with zero mean and standard deviations of σ = 4 minutes (which might correspond to the accuracy of predictions for active flights), and σ = 15 minutes (which might correspond to the accuracy of predictions for proposed flights).

Table 5-1 Probabilities for Flights to Enter a Sector

Probability	σ = 4	σ = 15
$P_{i,i}$	0.099	0.027
$P_{i,i+1} = P_{i,i-1}$	0.096	0.027
$P_{i,i+2} = P_{i,i-2}$	0.087	0.026
$P_{i,i+3} = P_{i,i-3}$	0.075	0.026
$P_{i,i+4} = P_{i,i-4}$	0.060	0.026
$P_{i,i+5} = P_{i,i-5}$	0.046	0.025
$P_{i,i+6} = P_{i,i-6}$	0.033	0.025
$P_{i,i+7} = P_{i,i-7}$	0.022	0.024
$P_{i,i+8} = P_{i,i-8}$	0.014	0.023
$P_{i,i+9} = P_{i,i-9}$	0.008	0.022
$P_{i,i+10} = P_{i,i-10}$	0.005	0.021
$P_{i,i+15} = P_{i,i-15}$	0.000	0.016
$P_{i,i+20} = P_{i,i-20}$	0.000	0.009
$P_{i,i+25} = P_{i,i-25}$	0.000	0.007
$P_{i,i+30} = P_{i,i-30}$	0.000	0.004
$P_{i,i+35} = P_{i,i-35}$	0.000	0.002
$P_{i,i+40} = P_{i,i-40}$	0.000	0.001

The two values of σ were selected to provide a symmetric, approximate representation of the two cases with active and proposed flights from Section 4. The distribution with σ = 4 represents the case with active flights. Its probabilities are close to those shown in the central part of the active flight plot in Figure 4-1. The distribution with σ = 15 represents the case with proposed flights.

These probabilities tend to be small (less than 0.10), since they correspond to single minutes. Note that as the difference $|i - k|$ between intervals i and k increases, $P_{i,k}$ becomes smaller. However, for the flights with ETAs close to interval i, the decrease in probabilities to arrive during this interval is not significant; at a certain distance from i, the probabilities start decreasing sharply.

For example, $P_{i,i}$ determines the probability for an active flight to enter a sector during a one-minute interval i (say, i = 1200 interval) if it is deterministically predicted to enter a sector during same interval i. The probability is small and is close to 0.1.

If a flight is deterministically predicted to enter a sector two minutes earlier ($k = i$-2) or two minute later ($k = i$+2) than the interval of interest i, the probability for the flight to enter a sector during interval i is smaller: $P_{i,\, i-2} = P_{i,\, i+2} \approx 0.087$.

For the active flights (σ = 4 minutes) with ETAs at least nine minutes from the interval of interest i (earlier or later) the probabilities to enter a sector during interval i are negligibly small (less than 0.01). Figure 5-1 illustrates the probabilities shown in Table 5-1.

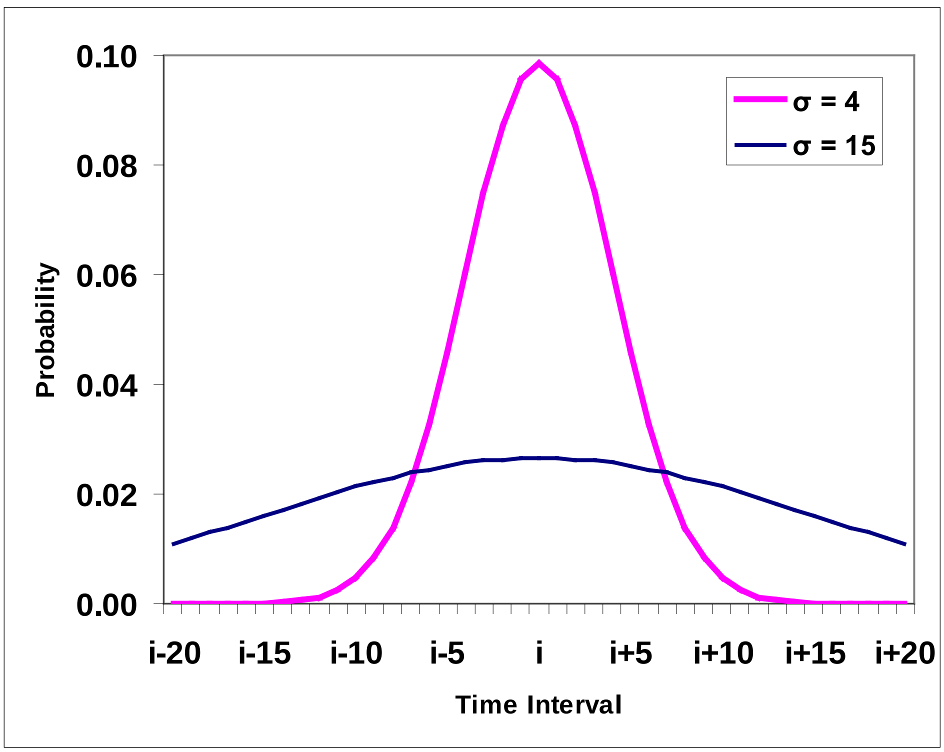

Figure 5-1 Probabilities for a Flight to Enter a Sector

It is important to note that when predictions of sector entry times are less accurate, the probability for a flight to enter a sector in becomes smaller, even in the minute that it was forecast to enter the sector (approximately 0.027 when σ = 15).

Another important observation from Table 5-1 is that, since the standard deviations of prediction errors are much larger than one-minute, the probabilities for the flights to enter a sector do not vary much from one minute to the next. In particular, for σ = 4 min, the probabilities for flights with ETAs in intervals *i*, *i*+1 and *i* -1 to arrive to a sector in interval *i* are, respectively equal to 0.099, 0.096 and 0.096, i.e., they are practically the same. For σ = 15, the corresponding probabilities a practically the same for the flights with ETAs in *thirteen (!)* intervals around the interval of interest *i* (including interval *i*).

5.3 Number of Flights Entering a Sector During a One-minute Interval

Appendix B describes a detailed analytical approach to probabilistic prediction of aggregate one-minute sector entry counts based on probabilities for individual flights to enter a sector and on deterministic predictions of sector entry counts at various one-minute intervals. In this analysis, probabilistic independence among the flights is assumed.

The main result of taking into consideration characteristics of uncertainty in flights' sector entry time predictions is that the flights with ETAs in several adjacent one-minute intervals to the interval of interest will be considered in calculating the aggregate demand in the interval of interest. For example, if d_k flights deterministically predicted to enter a sector during one-minute interval *k*, there is the probability $P_{i,k}$ for each of the d_k flights to enter the sector during the one-minute interval of interest *i*. The probability distribution of number of flights from d_k that can be counted in interval *i* is a binomial distribution (see formula (B1) in Appendix B). The mean number of flights from d_k contributing into the one-minute counts for interval *i* is equal to (see formula (B2))

$$\overline{d}_{i,k} = P_{i,k}\, d_k ,$$

The standard deviation of the number of flights from interval *k* contributing to interval *i*, according to (B4), is equal to

$$\sigma_{i,k} = \sqrt{P_{i,k}(1-P_{i,k})\, d_k}$$

There is an interesting result that illustrates how deterministic and probabilistic predictions are different. Assume that d_i flights are deterministically predicted to enter a sector in a one-minute interval *i*. There is a probability $P_{i,i}$ for each flight to be counted in the interval *i*.

Question 1: What is the probability that the deterministic prediction is correct? In other words, what is the probability that ALL deterministically predicted d_i flights would be counted as one-minute demands for entering a sector in interval *i*.

According to (B1) with $x = d_i$, the probability is equal to

$$P_i^{(i)}(d_i) = P_{i,i}^{d_i}$$

This probability for deterministic prediction to be correct is equal to the probability for a flight to enter a sector during interval *i* in power d_i. For $P_{i,i} << 1$ and $d_i >> 1$, the probability can be negligibly small.

For example, for $P_{i,i} = 0.1$ and $d_i = 10$, the probability for all ten flights to enter a sector during a one-minute interval i is equal to 10^{-10} (practically equal to zero).

Question 2: What is an average number of flights from deterministic prediction of d_i flights would be counted in the same interval i?

According to (B2), the average number is equal to

$$\overline{d}_{i,i} = P_{i,i}\, d_i ,$$

and for $P_{i,i} << 1$ it is a small fraction of deterministically predicted number of flights. In the above example, the expected number is equal to $\bar{d}_{i,i} = 1$, i.e., it is only one tenth of the deterministically predicted $d_i = 10$ flights.

Question 3: Given the results of questions 1 and 2, is there any value to deterministic predictions at all?

Yes, because even although deterministic predictions don't do very well with particular flights, the errors in predicting *aggregate sector counts* tend to cancel out. In particular, in spite of a small probability for ALL deterministically predicted flights to be counted in a specific one-minute interval, a stochastic mechanism of aggregating flights for demand predictions produces reasonable numbers by aggregating fractions of demands from some other neighboring intervals. For example, of the 10 flights predicted to enter the sector at time *i*, only one of those flights might actually enter the sector at time *i*. However, there may be several other flights, predicted to enter at other times (close to *i*), that could enter the sector at time *i*.

The advantage of probabilistic approach is that the probabilistic predictions rely not only on deterministic prediction for the interval of interest but explicitly take into account deterministic predictions for several adjacent one-minute intervals that precede and follow the interval of interest.

When the uncertainty in flights' sector-entering times is taken into account, the total random number of flights predicted for a specific one-minute interval *i* is equal to the sum of the random numbers of flights from several adjacent intervals *k* that can be counted in the interval of interest *i*. The number of adjacent intervals that should be taken into account depends on relative values of probabilities $P_{i,k}$. The probabilities $P_{i,k}$ decrease with increasing time distance $|i - k|$, and the maximum distance $\beta = \max |i - k|$ can be found, beyond which the probabilities $P_{i,k}$ become so small that the demands from those intervals *k* do not significantly contribute to the counts for interval *i* (see formula (B10)).

The expected number of probabilistically predicted one-minute counts $\bar{\tilde{d}}_i$ and the standard deviation of those counts are equal to, respectively (see equation (B11) and B(13) in Appendix B):

$$\bar{\tilde{d}}_i = \sum_{k=i-\beta}^{i+\beta} \bar{d}_{i,k} = \sum_{k=i-\beta}^{i+\beta} P_{i,k} d_k \qquad (5.1)$$

$$\sigma_i = \sqrt{\sum_{k=i-\beta}^{i+\beta} P_{i,k}(1 - P_{i,k}) d_k}\,, \qquad (5.2)$$

where d_k is a deterministic prediction of sector entry counts for a one-minute interval *k*.

The probabilistic prediction of one-minute flight counts entering a sector can be represented by the expected number $\bar{\tilde{d}}_i$ and the uncertainty area around the expected number: $\bar{\tilde{d}}_i \pm j\,\sigma_i$, where *j* determines a size of the uncertainty area restricted by certain percentiles. For $j = 1$, $\bar{\tilde{d}}_i + \sigma_i$ corresponds to the 84th percentile, and $\bar{\tilde{d}}_i - \sigma_i$ corresponds to the 16th percentile. For $j = 2$, the percentiles are 2.3% and 97.7%.

EXAMPLES

Consider an example of probabilistic prediction of number of flights entering a sector during a one-minute interval when the errors in predicting sector entry times for individual flights are normally distributed with zero average and with standard deviation of 4 minutes (σ = 4 min)

The probabilities $P_{i,k}$ in this case are shown in Section 5.2 for $k = i$-10, i-9, i-8, …, i-1, i, i+1, …, i+8, i+9, i+10, i.e., for ten one-minute intervals from both sides of the interval i of interest. The values of probabilities $P_{i,k}$ for $k \le i - 9$ and $k \ge i + 9$, become too small and can be neglected. As a result, the number of intervals β preceding and following the interval of interest i where deterministically predicted number of flights d_k entering a sector contribute to flight count predictions for interval i is equal to $\beta = 8$.

According to (5.1) and (5.2), the expected number of flights predicted to enter a sector during a one-minute interval i and standard deviation of the number can be found from the following equations.

$$\overline{\tilde{d}}_i = \sum_{k=i-8}^{i+8} P_{i,k} d_k \tag{5.3}$$

$$\sigma_i = \sqrt{\sum_{k=i-8}^{i+8} P_{i,k}(1 - P_{i,k}) d_k}\ , \tag{5.4}$$

Three cases are considered. In all three cases, the deterministic demand for the interval of interest (k=i) is 8. In Case 1, the other demands are smaller than 8. In Case 2, the other demands are similar to the demand in the interval of interest. In Case 3, the other demands are larger. Table 5-2 presents numerical values of deterministic demands d_k and probabilities $P_{i,k}$ that are to be used in equations (5.3) and (5.4).

Table 5-2 Deterministic demands d_k and probabilities $P_{i,k}$ for the intervals surrounding interval i = 1200

k	k	Case 1	Case 2	Case 3	$P_{i,k}$	$P_{i,k}$ (1 - $P_{i,k}$)
1148	i-12	6	12	15	0.001	0.001
1149	i-11	3	6	10	0.002	0.002
1150	i-10	2	4	8	0.005	0.005
1151	i - 9	3	6	11	0.008	0.008
1152	i - 8	4	6	11	0.014	0.014
1153	i - 7	2	8	13	0.022	0.022
1154	i - 6	3	7	12	0.033	0.032
1155	i - 5	2	8	13	0.046	0.044
1156	i - 4	3	7	12	0.060	0.056
1157	i - 3	2	8	13	0.075	0.069
1158	i - 2	1	9	14	0.087	0.079
1159	i - 1	2	8	13	0.096	0.087
1200	***i***	**8**	**8**	**8**	**0.099**	**0.089**
1201	i + 1	1	9	14	0.096	0.087
1202	i + 2	2	8	13	0.087	0.079
1203	i + 3	1	9	14	0.075	0.069
1204	i + 4	4	6	11	0.060	0.056
1205	i + 5	1	9	14	0.046	0.044
1206	i + 6	5	5	10	0.033	0.032
1207	i + 7	3	7	12	0.022	0.022
1208	i + 8	5	5	10	0.014	0.014
1209	i + 9	4	7	12	0.008	0.008

Figure 5-2 shows the deterministically predicted one-minute sector entry counts from 1150 to 1210 for Case 1. We need to probabilistically predict the number of flights entering a sector during a one-minute interval i = 1200 (from 1200 to 1201). According to (5.3) and (5.4), the expected number and standard deviation of one-minute counts for interval i = 1200 can be determined through the weighted sums of seventeen deterministic predictions d_k that include predictions for eight minutes preceding interval i=1200 and for eight minutes following the interval, as well as the deterministic prediction for interval i=1200. Altogether, seventeen one-minute demand counts d_k from $k = i - 8 = 1152$ to $k = i + 8 = 1208$ participate in determining mean and standard deviation of one-minute counts from interval i = 1200. Those deterministic demands are shown in Figure 5-2 by the green bars. Deterministic demands outside the interval are shown by blue bars.

We will compare the deterministic prediction of d_i = 8 flights entering a sector during a one-minute interval 1200 with the probabilistic prediction.

After using the numerical values from Table 5-2 in equations (5.3) and (5.4), the expected number $\bar{\tilde{d}}_i$ and standard deviation σ_i of number of flights entering a sector during a one-minute interval i are respectively equal to:

$\bar{\tilde{d}}_i = 2.7$ and $\sigma_i = 1.6$

Figure 5-2 shows the deterministic prediction, the expected counts and the uncertainty area around the expected value $\bar{\tilde{d}}_i = 2.7$ restricted by the 84th percentile $\bar{\tilde{d}}_i + \sigma_i = 4$ and the 16th percentile $\bar{\tilde{d}}_i - \sigma_i = 1$. Hence, with the 0.68 probability the number of flights entering the sector during a one-minute interval i = 1200 is between 1.1 and 4.2 flights with the expected number of 2.7 flights.

It is important to notice that the probability of having the deterministically predicted 8 flights entering a sector during a one-minute interval i is equal to

$(P_{i,i})^8 = 0.099^8 = 9.2 * 10^{-9}$,

and it is negligibly small. The weighted sum of deterministically predicted demands at several neighboring one minute intervals made it possible to improve the probabilistic property of one-minute demand prediction by providing a justifiable, probabilistic range of uncertainty.

The deterministic prediction $d_i = 8$, which is a single peak, appeared to be highly unreliable: it is far outside the 68% uncertainty range [1.1, 4.2] and is much higher than the expected number 2.7.

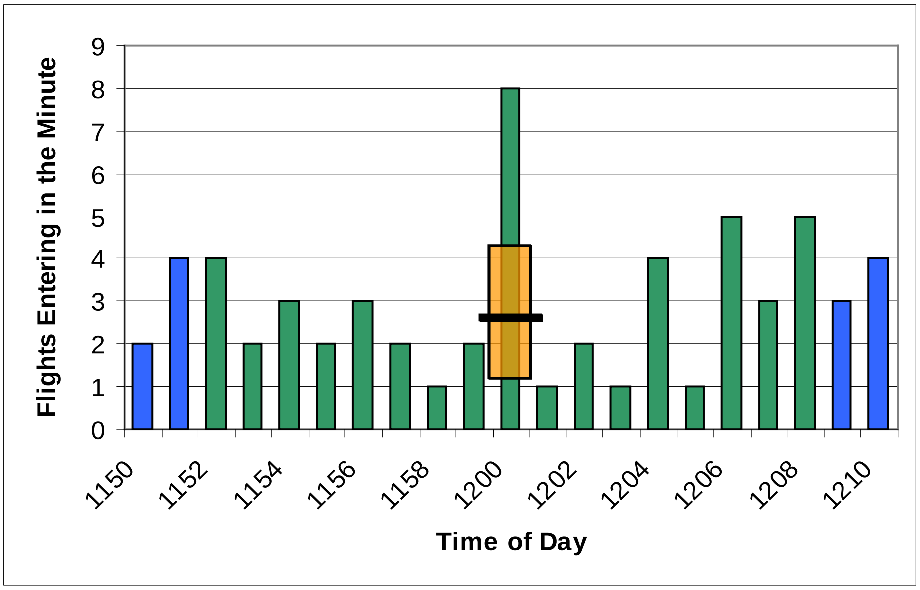

Figure 5-2 Sector Entry Counts, Case 1

For Case 2, a similar calculation is performed. Here the probabilistic demand is similar to the deterministic demand (Figure 5-3).

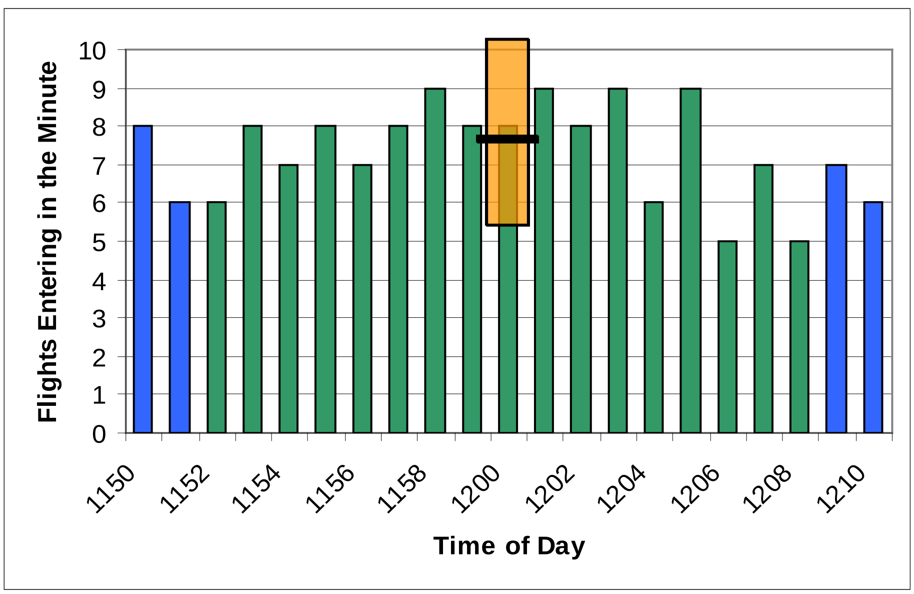

Figure 5-3 Sector Entry Counts, Case 2

Finally, in Case 3, the probabilistic demand is higher than the deterministic demand (Figure 5-4).

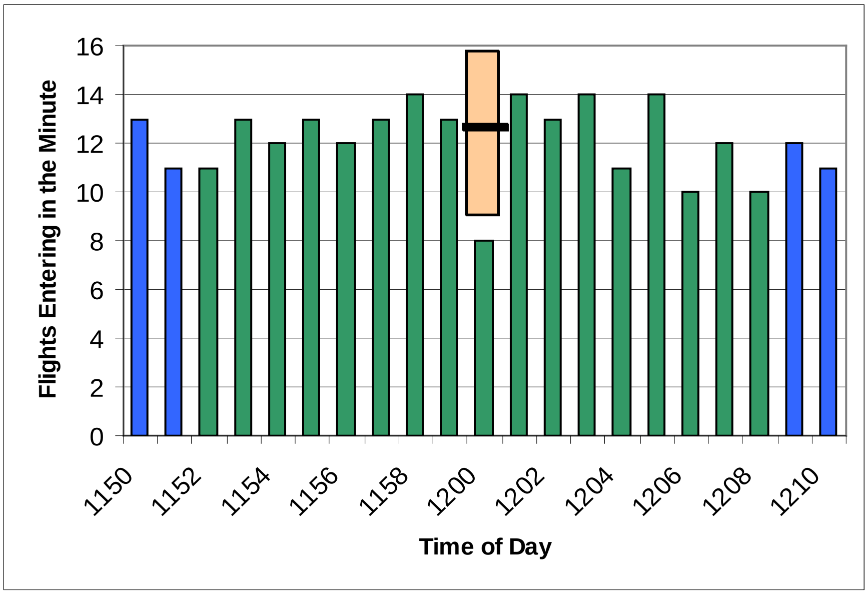

Figure 5-4 Sector Demand Counts, Case 3

These examples illustrate the strong dependency of probabilistic predictions for sector entry counts on the deterministic predictions in several adjacent intervals and on the accuracy of sector entry time predictions for individual flights.

5.4 Flights in a Sector During a One-minute Interval

In the previous section, we developed a probabilistic representation of the number of flights ***entering*** a sector at a particular 1-minute time interval i. To develop a representation of the number of flights ***in*** a sector at interval i, we need to consider an additional factor: the time-in-sector for individual flights.

Appendix C presents the mathematical expressions and analytical approach needed for probabilistic predictions of one-minute demand counts of the flights in a sector. The analytical formulas presented in the Appendix allow for determining probability distributions of one-minute demand predictions in a sector, calculating average and standard deviation of predicted number of flights in a sector at any one-minute interval by using deterministic predictions of sector demand and characteristics of uncertainty in predicting times of entering a sector for individual flights. The analytical results consider important factors for traffic demand predictions, such as status of individual flights (active or proposed) and difference in times in sector for various flights. What is important is that characteristics of probabilistic predictions of one-minute traffic demand in a sector depends heavily on probabilistic predictions of one-minute sector entry counts, considered in the previous section.

The starting point for calculating one-minute sector demand is constructing the relationship between number of flights in a sector and the number of flights entering the sector.

Let τ be the amount of time that a flight spends in a sector. If the flight's predicted entry time is k, its predicted sector exit time is $k+\tau$. In this analysis, τ is assumed to be known and non-probabilistic.[2] For this section of this report, τ is assumed to be the same for every flight. In the next section (5.5), this assumption is removed.

[2] Recall from Section 4 of this report that the error in time-in-sector is substantially lower than the error in sector entry time.

The flight will be in the sector during minute i if the flight enters at or before minute i, and exits after minute i. In inequalities, this is

$k \le i$ and

$k + \tau > i$.

Hence, if d_k is a number of flights deterministically predicted to enter a sector during a one-minute interval k and time in sector for each flight is τ, then the deterministic prediction of number of flights in a sector $D_{i,\tau}$ during one-minute interval i is

$$D_{i,\tau} = d_i + d_{i-1} + d_{i-2} + \dots + d_{i-(\tau-2)} + d_{i-(\tau-1)} = \sum_{j=i-\tau+1}^{i} d_j \,, \qquad (5.5)$$

For example, if $k = 1158$, $i = 1200$ and $\tau = 5$ minutes, the flight will be in the sector at 1200 if the actual entry time is between 1156 and 1200. If the entry time is before 1156, the flight is too early, and will have left the sector before 1200. If it is after 1200, the flight is too late. This flight will be in the sector at 1200 if it enters the sector at any of the following times: 1156, 1157, 1158, 1159, and 1200.

The numbers of predicted flights entering a sector are not deterministic because of random errors in predicting times of entering the sector for individual flights. This case was considered in the previous section and in Appendix B.

In this case, the expected number of flights $\overline{\tilde{D}}_{i,\tau}$ in a sector in a one-minute interval i is, thus the sum of the expected numbers of flights entering the sector over a series of times (see formula (C5) in Appendix C):

$$\overline{\tilde{D}}_{i,\tau} = \sum_{j=i-\tau+1}^{i} \quad \sum_{k=j-\beta}^{j+\beta} P_{j,k} d_k \qquad (5.6)$$

where $P_{j,k}$ is a probability for a flight to enter a sector during interval j if it is predicted to enter the sector during interval k.

The formula (5.6) can be modified and presented in a more convenient form so that the expected number of flights in a sector $\overline{\tilde{D}}_{i,\tau}$ is equal to (see formula (C6) in Appendix C)

$$\overline{\tilde{D}}_{i,\tau} = \sum_{k=i-\tau-\beta+1}^{i+\beta} P_{i,k,\tau}\, d_k \qquad (5.7)$$

where $P_{i,k,\tau}$ can be interpreted as a probability for a flight to be in a sector during minute i if it is deterministically predicted to enter a sector during minute k.

The standard deviation of one-minute sector demand counts is (see (C9) in Appendix C)

$$\sigma(\tilde{D}_{i,\tau}) = \sqrt{\sum_{k=i-\beta-\tau+1}^{i+\beta} P_{i,k,\tau}(1-P_{i,k,\tau})\, d_k} \;. \qquad (5.8)$$

The formula for calculating probabilities $P_{i,k,\tau}$ is given in Appendix C (see formula (C7)).

The analytical expression for probabilities $P_{i,k,\tau}$, selects the probabilities $P_{i,k,\tau}$ for certain combinations of i and k minutes only, for which the flights deterministically predicted to arrive in a sector at minute k satisfy specific constraints to be in a sector during minute i. Looking at the formula, it is hard to quickly understand which combinations of i and k are relevant and which are not.

There is another way to present the computation of probabilities $P_{i,k,\tau}$ that visualizes the computational procedure so that its structure can be seen more clearly. It can be done by placing in a grid the terms contributing to $P_{i,k,\tau}$.

In the grid (Table 5-3) each row corresponds to specific k (predicted entry time), and each column corresponds to specific one-minute entry time that contributes to the interval of interest *i*. There are τ columns. The blue area of the grid covers the combinations of *k* and *i* relevant for calculating the probabilities $P_{i,k,\tau}$ via equation (C7) in Appendix C.

The sum of terms in each row gives the probability $P_{i,k,\tau}$ for the corresponding *k*. These probabilities for the whole range of *k* can be then used in formulas (5.7) and (5.8) to calculate the expected number and standard deviation of predicted number of flights in a sector during a one-minute interval of interest. Note that the term $P_{i,k,\tau} d_k$ in formula (5.7) represents the expected number of flights from deterministic prediction d_k to be in a sector at one-minute interval *i*.

Table 5-3 Terms of Equation 5.6

	Entry times for flights that contribute to the interval of interest i				
k↓	**i**	**i-1**	**…**	**i-τ+2**	**i-τ+1**
i- β–τ +1					$P_{i-\tau+1,i-\beta-\tau+1}$
i- β – τ+2				$P_{i-\tau+2,i-\beta-\tau+2}$	$P_{i-\tau+1,i-\beta-\tau+2}$
…			…	…	…
i- β -1		$P_{i-1,i-\beta-1}$	…	$P_{i-\tau+2,i-\beta-1}$	$P_{i-\tau+1,i-\beta-1}$
i- β	$P_{i,i-\beta}$	$P_{i-1,i-\beta}$	…	$P_{i-\tau+2,i-\beta}$	$P_{i-\tau+1,i-\beta}$
…	…	…	…	…	…
i-1	$P_{i,i-1}$	$P_{i-1,i-1}$	…	$P_{i-\tau+2,i-1}$	$P_{i-\tau+1,i-1}$
i	$P_{i,i}$	$P_{i-1,i}$	…	$P_{i-\tau+2,i}$	$P_{i-\tau+1,i}$
...	...	...	...	...	...
i+ β–τ +1	$P_{i,,i+\beta-\tau+1}$				$P_{i-\tau+1,i+\beta-\tau+1}$
i+ β–τ +2	$P_{i,i+1}$	$P_{i-1,i+1}$	…	$P_{i-\tau+2,i+1}$	
…	…	…	…		
i+β-1	$P_{i,i+\beta-1}$	$P_{i-1,i+\beta-1}$			
i+β	$P_{i,i+\beta}$				

Some Numerical Examples

The numerical examples below illustrate the calculation of expected number and standard deviation for probabilistic predictions of one-minute demand counts in a sector during a specific one-minute. The examples also illustrate that the outcome of computations is hard to predict because it depends on both relative values of deterministic sector entry counts at different one-minute intervals (demand profile for sector entries) and accuracy of entry time predictions. Comparison of deterministic and probabilistic predictions of one-minute sector demand counts is illustrated in three cases when the deterministic prediction is within the range of uncertainty, and when it is outside (above and below) the uncertainty range.

In these examples, the time of interest i is 1200, time in sector τ is 5 minutes, and β is 8 minutes. With $i = 1200$ and $\tau = 5$, a flight can be in the sector at 1200 if it enters the sector at any of the following times: $(i - \tau + 1) = 1156$, 1157, 1158, 1159, and $i = 1200$.

Example 1: Probabilities taken from Table 5-2

Table 5-4, which is similar in structure to Table 5-3, shows the probabilities that contribute to $P_{i,k,\tau}$. In this table, the relevant values of k range from $i - \beta - \tau + 1$ to $i + \beta$. With $i = 1200$, $\beta = 8$ and $\tau = 5$, this yields a range for k from 1148 to 1208. The probabilities are taken from Table 5-2.

Table 5-4 Components of $P_{i,k,\tau}$

	Sector Entry Times					
k	**1200 i**	**1159**	**1158**	**1157**	**1156 $i\!-\!\tau\!+\!1$**	$P_{i,k,\tau}$
1148 **(i-β-τ+1)**	0.001	0.002	0.005	0.008	0.014	0.014
1149	0.002	0.005	0.008	0.014	0.022	0.036
1150	0.005	0.008	0.014	0.022	0.033	0.069
1151	0.008	0.014	0.022	0.033	0.046	0.115
1152	0.014	0.022	0.033	0.046	0.060	0.175
1153	0.022	0.033	0.046	0.060	0.075	0.236
1154	0.033	0.046	0.060	0.075	0.087	0.301
1155	0.046	0.060	0.075	0.087	0.096	0.364
1156	0.060	0.075	0.087	0.096	0.099	0.417
1157	0.075	0.087	0.096	0.099	0.096	0.452
1158	0.087	0.096	0.099	0.096	0.087	0.465
1159	0.096	0.099	0.096	0.087	0.075	0.452
1200	0.099	0.096	0.087	0.075	0.060	0.417
1201	0.096	0.087	0.075	0.060	0.046	0.364
1202	0.087	0.075	0.060	0.046	0.033	0.301
1203	0.075	0.060	0.046	0.033	0.022	0.236
1204	0.060	0.046	0.033	0.022	0.014	0.175
1205	0.046	0.033	0.022	0.014	0.008	0.115
1206	0.033	0.022	0.014	0.008	0.005	0.069
1207	0.022	0.014	0.008	0.005	0.002	0.036
1208 **(i+β)**	0.014	0.008	0.005	0.002	0.001	0.014

The $P_{i,k,\tau}$ from Table 5-4 are then combined with the deterministic predictions d_k and equations (5.7) and (5.8) to calculate the mean and variance of the probabilistic prediction (see Table 5-5).

Table 5-5 Predictions for Several Times *k*

k	d_k	$P_{i,k,\tau}$	$d_k\ P_{i,k,\tau}$	$d_k\ P_{i,k,\tau}(1-\ P_{i,k,\tau})$
1148	6	0.014	0.084	0.083
1149	3	0.036	0.108	0.104
1150	2	0.069	0.138	0.128
1151	4	0.115	0.460	0.407
1152	4	0.175	0.700	0.578
1153	2	0.236	0.472	0.361
1154	3	0.301	0.903	0.631
1155	2	0.364	0.728	0.463
1156	3	0.417	1.251	0.729
1157	2	0.452	0.904	0.495
1158	0	0.465	0.000	0.000
1159	2	0.452	0.904	0.495
1200	**8**	0.417	3.336	1.945
1201	0	0.364	0.000	0.000
1202	2	0.301	0.602	0.421
1203	1	0.236	0.236	0.180
1204	4	0.175	0.700	0.578
1205	1	0.115	0.115	0.102
1206	5	0.069	0.345	0.321
1207	3	0.036	0.108	0.104
1208	5	0.014	0.070	0.069
			SUM = Mean: 12.164	SUM = Variance: 8.194

The expected (mean) value is 12.16. The standard deviation is 2.86 (the square root of 8.19). Using ± one standard deviation as the uncertainty range, the range becomes 9.3 to 15.02.

Further Examples

Three examples are presented (Table 5-6). These examples have deterministic predictions of the number of flights entering a sector, for *k* ranging from 1152 to 1208. In all three examples, the number of flights predicted to enter the sector for the minutes 1156 through 1200 is the same. This means that under the assumption that time-in-sector is a constant 5 minutes, the deterministic prediction of number of flights in a sector in 1200 one-minute interval is equal to the sum of sector entry counts from 1156 to 1200, and, according to Table 5-6 it is equal to 15 flights. This number is used in all three examples.

Table 5-6 Three Examples

	Deterministic Prediction of Sector Entries			Deterministic Prediction of Sector Occupancy		
Time	Base	1-Peak	1-Valley	Base	1-Peak	1-Valley
1152	4	2	8	19	9	34
1153	2	0	6	15	6	30
1154	3	1	6	15	6	30
1155	2	0	4	15	5	30
1156	3	3	3	14	6	25
1157	2	2	2	12	6	19
1158	0	0	0	10	6	15
1159	2	2	2	9	7	11
1200	8	8	8	15	15	15
1201	0	0	0	12	12	12
1202	2	1	4	12	11	14
1203	1	0	2	13	11	16
1204	4	3	8	15	12	22
1205	1	0	2	8	4	16
1206	5	2	8	13	6	24
1207	3	3	6	14	8	26
1208	5	2	8	18	10	32

The differences between the examples are as follows:

- In the Base example, the number of flights in the sector at 1200 (15 flights) is approximately equal to the average number of flights in the sector at other times.
- In the 1-Peak example, the number of flights in the sector at 1200 is unusually high.
- In the 1-Valley example, the number of flights in the sector at 1200 is unusually low.

The probabilistic predictions are computed for each of the examples, as follows:

Table 5-7 Probabilistic Predictions for Time 1200

	Base	Low	High
Expected Value $\bar{\tilde{D}}_{1200,\tau}$	12.57	8.73	18.75
Standard Deviation $\sigma(\tilde{D}_{1200,\tau})$	2.93	2.37	3.66
$\bar{\tilde{D}}_{1200,\tau} - \sigma(\tilde{D}_{1200,\tau})$	9.65	6.36	15.09
$\bar{\tilde{D}}_{1200,\tau} + \sigma(\tilde{D}_{1200,\tau})$	15.50	11.09	22.41

In the base example (Figure 5-5), the deterministic prediction of 15 is within the range of the probabilistic prediction (9.6 to 15.5) between the16th and 84th percentiles. In the one-peak example (Figure 5-6), the deterministic prediction is higher than the range of probabilistic predictions (6.4 to 11.1). In the one-valley example (Figure 5-7), the deterministic prediction is lower than the range of probabilistic predictions (15.1 to 22.4).

In these three figures, the orange bars are the deterministic predictions, the purple diamond is the mean of the probabilistic prediction, and the black line represents the range of the probabilistic prediction. In each of the figures, the deterministic prediction for 1200 is 15.

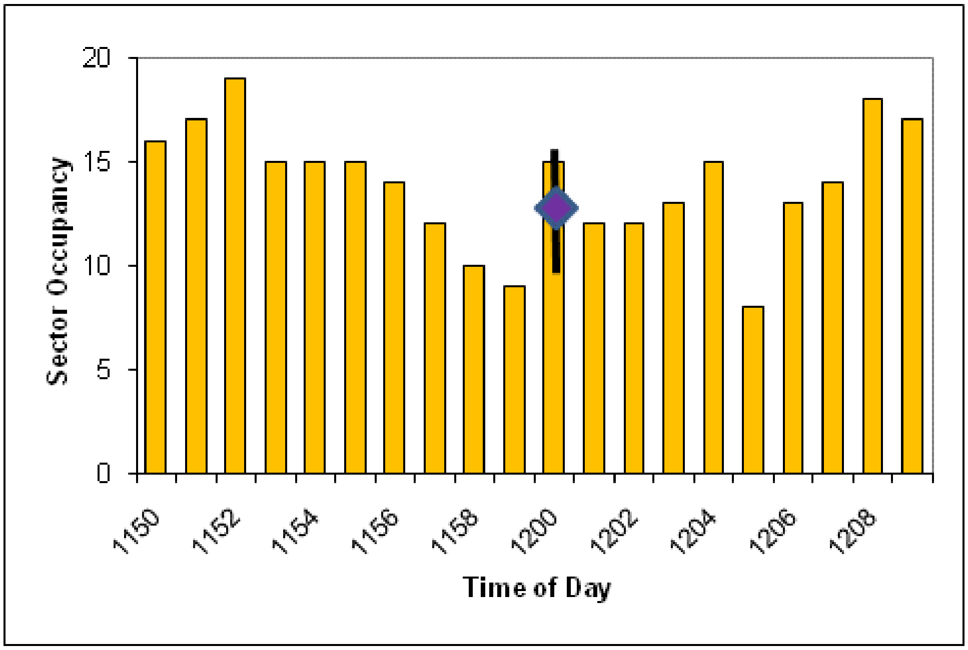

Figure 5-5 Predictions for Base Case

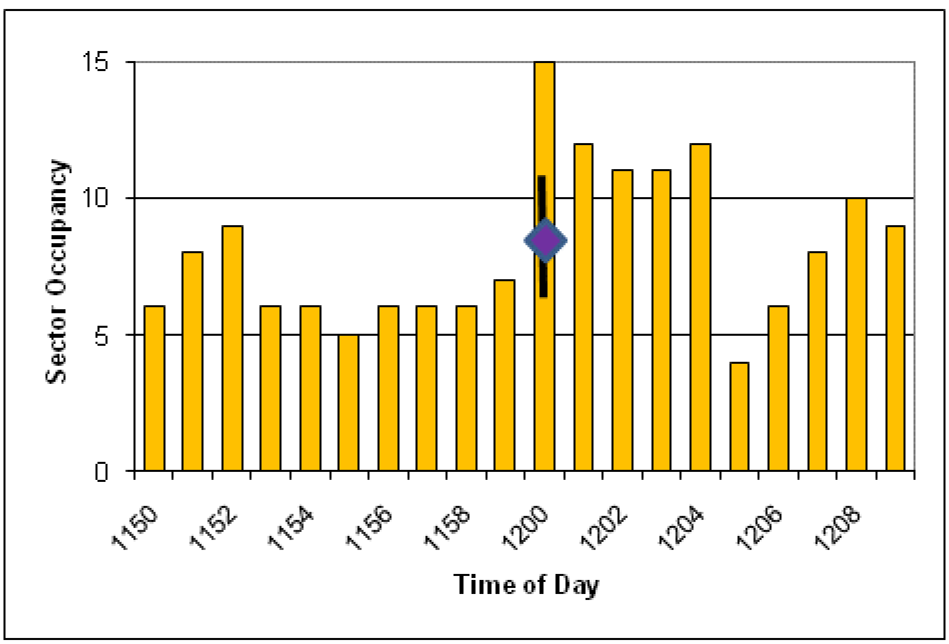

Figure 5-6 Predictions for 1-Peak Case

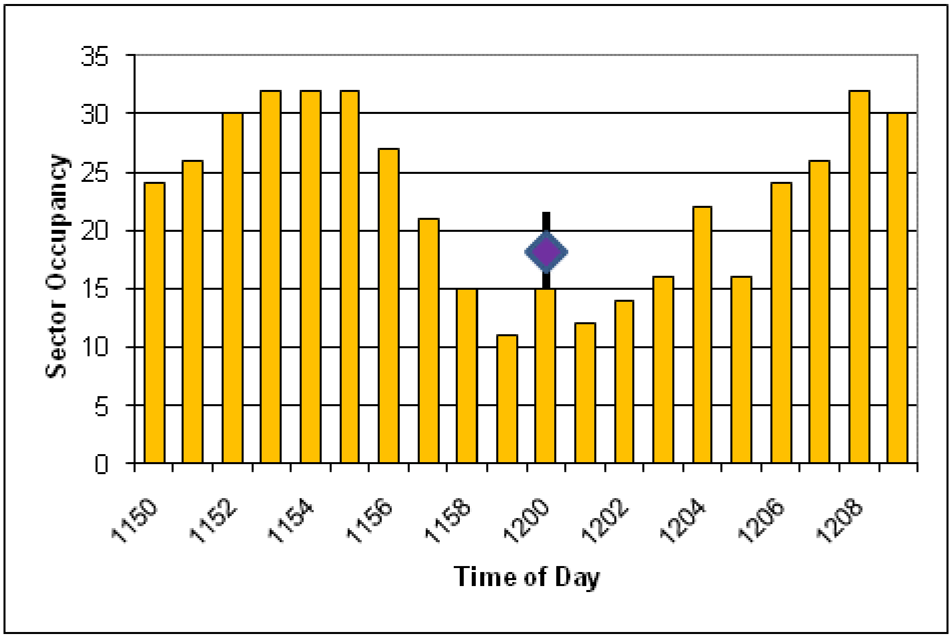

Figure 5-7 Predictions for 1-Valley Case

5.5 Probabilistic Demand Prediction for Flights with Different Times-in-Sector

Using the analytical results from Appendix C and the concepts from the previous sections, a probabilistic minute-by-minute model of the number of flights in a sector can be constructed. Recall that the predictions of in-sector time are substantially more accurate than those of sector entry time. The in-sector times are therefore treated as known and constant. Then, the probabilistic prediction is simply a matter of adding up all of the predictions across the various in-sector times (τ).

Suppose our time of interest is *i*, a flight is predicted to enter a sector at time t, and the time-in-sector is τ (this means that the predicted exit time is $t+ \tau$). Then, the probability that the flight will be in the sector at time *i* is simply the sum of the individual minute-by-minute probabilities for each minute ranging from i-τ+1 to *i*, when the flight might enter the sector. By combining these probabilities for individual flights, both an expected value and standard deviation for number of flights in a sector can be developed. The next three figures illustrate some results. In all cases, the times of interest i range from 900 to 929 (1500Z – 1529Z) and the look-ahead time is approximately 90 minutes. In these figures

- The light orange bars are the deterministic predictions.
- The black lines are the probabilistic predictions. The diamond in the middle of the line is the mean value of prediction; the bottom of the line is the mean minus one standard deviation; and the top of the line is the mean plus one standard deviation.

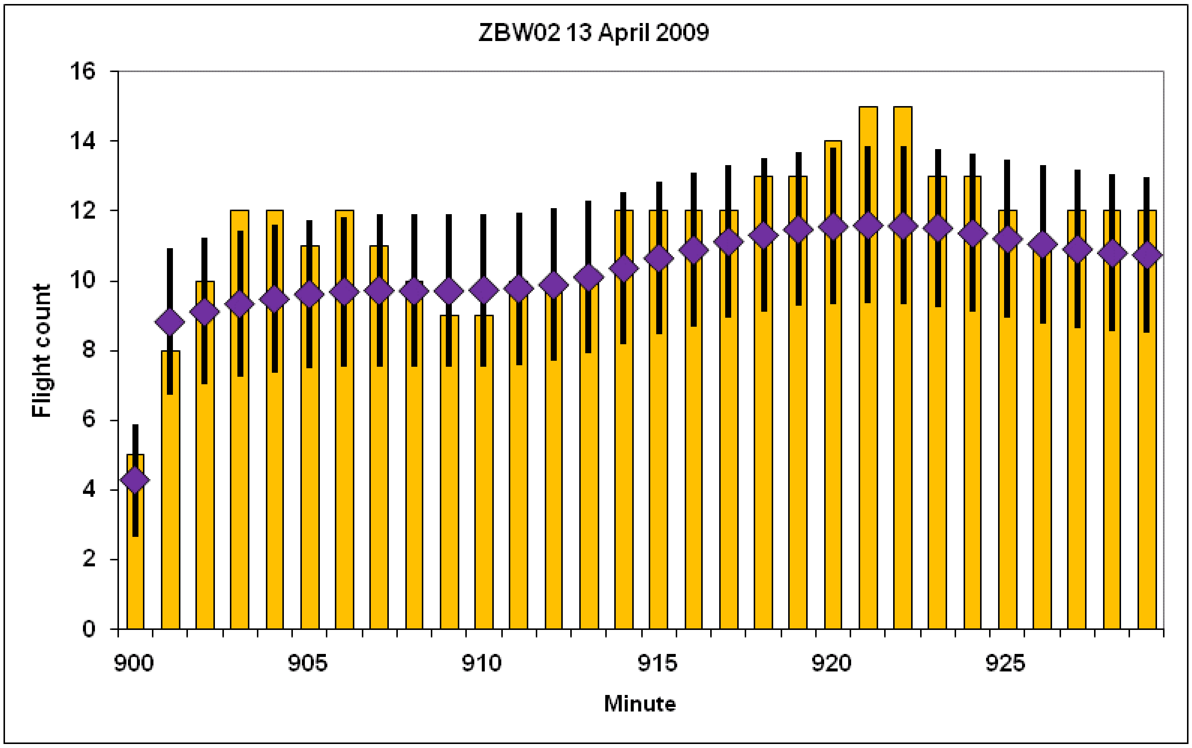

Figure 5-8 ZBW02 Probabilistic Predictions

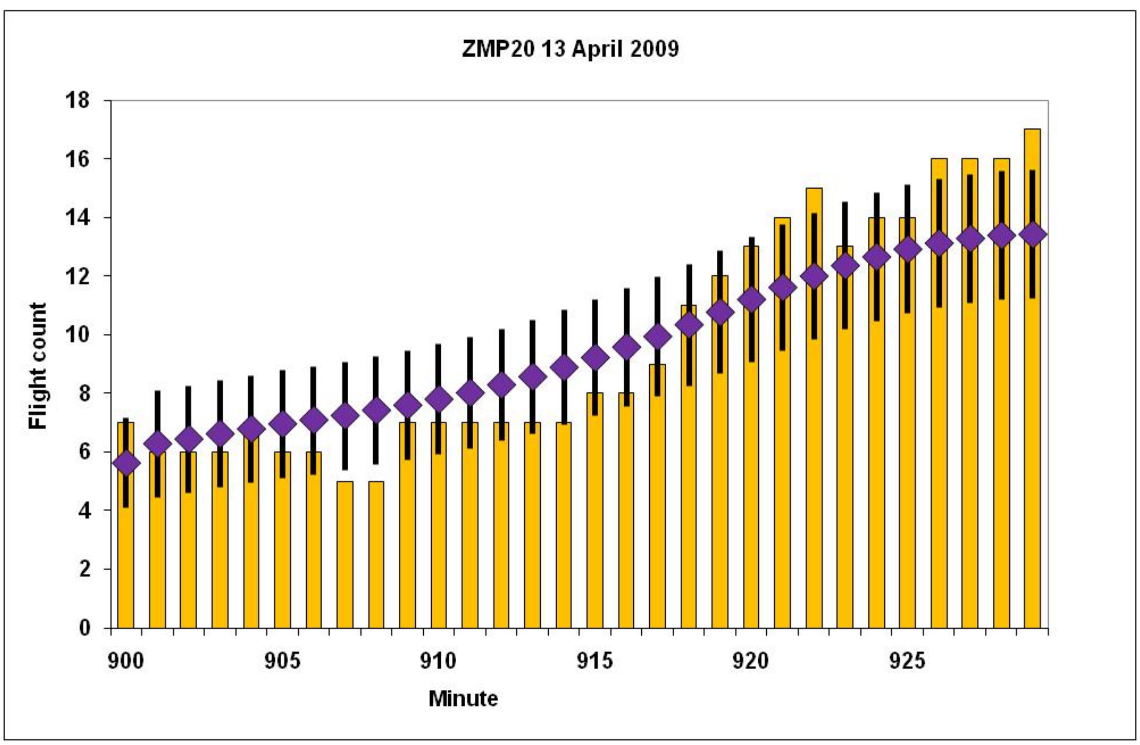

Figure 5-9 ZMP20 Probabilistic Predictions

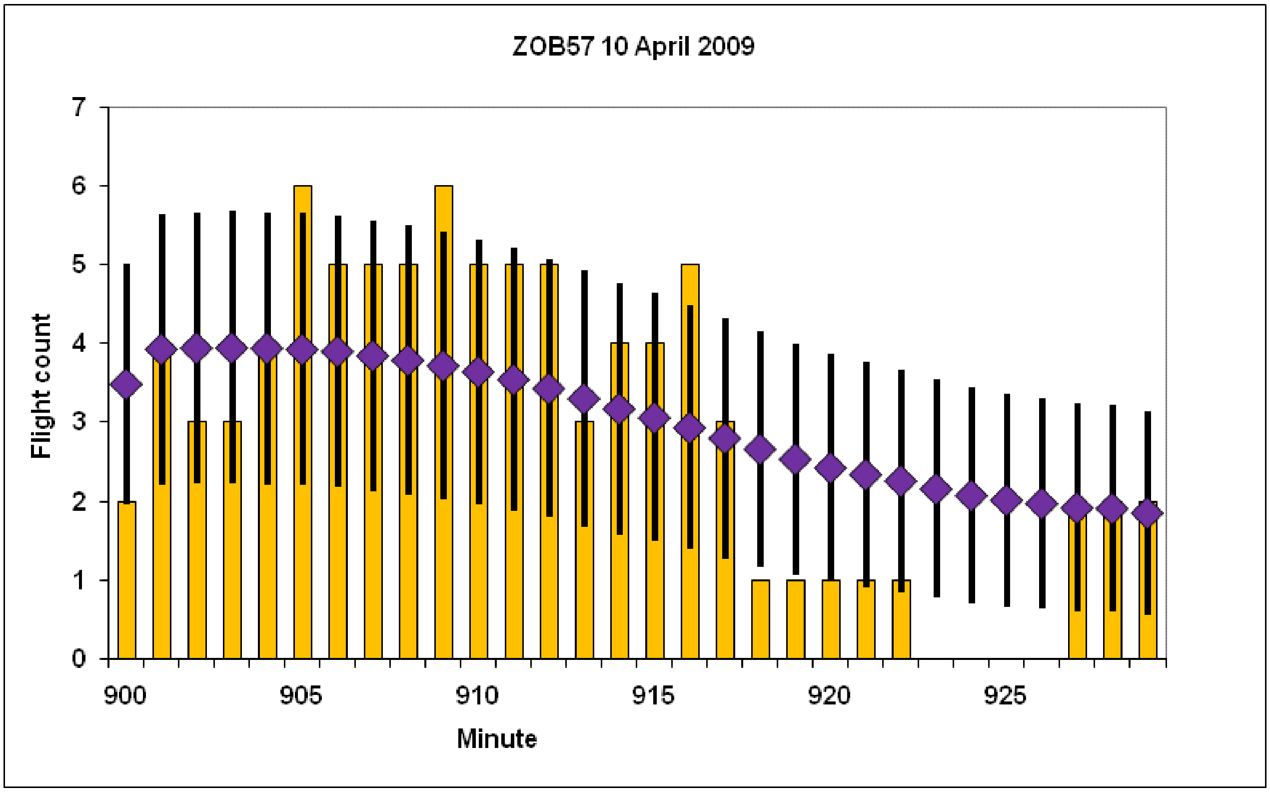

Figure 5-10 ZOB57 Probabilistic Predictions

Probabilistic predictions, similar to those shown in the three figures above, were calculated for these 30 minutes on 7 days for 16 sectors, yielding 3360 observations. The probabilistic predictions were much less volatile than the deterministic predictions. The measure used was the range of the predictions over the thirty minutes. For example, the deterministic predictions in Figure 5-10 have a range of 6 (from a minimum of 0 to a maximum of 6), while the mean probabilistic predictions (the centers of each vertical bar) have a range of 2 (from a minimum of 2 to a maximum of 4).

Using this measure, the average range of the deterministic predictions was 5.02, while the average range for the probabilistic predictions was only 2.35.

6. Conclusion

TFMS currently makes its aggregate traffic demand predictions based on deterministic projections of traffic. It neglects random errors in predictions. The purpose of this study was to

- Analyze the errors in TFMS predictions of aggregate one-minute demand counts in sector and sector entry
- Analyze the errors and characterize uncertainty in TFMS flight-by-flight predictions of sector entry times and time in sector
- Use these characteristics of uncertainty to determine the probability for an aircraft to cross a sector boundary and to be in a sector at any given time
- Develop a probabilistic model of sector demand count predictions.

Analysis of accuracy of predictions for individual flights was conducted for flight en route sector entry times and time-in-sector. Most of the analysis was performed for a look-ahead time (LAT) of 90 minutes, and the characteristics of prediction errors were separately estimated for active and for proposed flights that have not departed from origin airports. As a result of this analysis of TFMS historical data, the characteristics of prediction errors were estimated.

For sector entry times, the results can be summarized as follows:

- Predictions for active flights are significantly more accurate than for proposed flights, with a narrower range of uncertainty for the active flights. For active flights, the standard deviation of arrival time error was between five and 12 minutes, depending on the sector, while for proposed flights this standard deviation was between 16 and 21 minutes.
- For active flights, the distributions for prediction errors of airport arrival or sector entry times were either symmetric or were slightly asymmetric with a higher probability for active flights to enter a sector earlier than predicted.
- For proposed flights, the distributions were asymmetric the other way, with a higher probability for a flight to arrive later than predicted.

A separate analysis was performed to estimate the distribution of the time-in-sector for flights as well as the prediction error for time-in-sector. As expected, the flights' time-in-sector varies significantly by sector, with larger sectors requiring longer times. Typical time-in-sector ranged from 4 minutes (ZBW46) to 30 minutes (ZMP20). The analysis of accuracy of time-in-sector predictions showed that

- The TFMS prediction errors for time-in-sector were noticeably smaller than for the sector entry time. The standard deviation of time-in-sector error is typically between three and six minutes (depending on the sector's size); whereas the standard deviation of the error of sector entry time is between five and 21 minutes (depending on flight status).
- There was not much difference in accuracy of time-in-sector prediction with respect to LAT or flight status (active or proposed).
- The probability for a flight to be in the sector during the time interval that was deterministically predicted depends heavily on both the in-sector time and the flight status (active or proposed). For example, if a proposed flight is predicted (with 90-minute look-ahead time) to be in a sector during a particular minute, the probability that it will be in that sector during that minute ranges from 22 to 64% (with an average of 41%), depending on sector size. For active (airborne) flights, this range is 45 to 84% (with an average of 70%).

The results of the analysis of uncertainty in individual flight predictions were used to develop a new methodology for probabilistic predictions of aggregate, one-minute traffic demand counts for sector boundary crossing and for the number of flights in a sector. The basic results of the study and the major features of the methodology are as follows:

- Because the errors in sector entry time predictions are usually much greater than one minute, the probabilities for an individual flight with an ETA at a specific one-minute intervals to enter a sector at that interval is small (often less than 0.1). Moreover, the flights with ETAs close to (but not equal to) the one-minute interval of interest have nearly the same probabilities to enter a sector during the interval of interest. **This justifies a fundamental result of the study that, if the prediction errors of arrival time for individual flights is much greater than the time interval of interest for aggregate count predictions, then probabilistic predictions of aggregate number of flights for this interval should also consider the flights with ETAs in several neighboring intervals.**
- The proposed model for probabilistic predictions of one-minute aggregate sector entry counts provides analytical expressions for the expected number of flights along with the standard deviation of the predicted numbers. The expected number of flights predicted for a specific one-minute interval is a weighted sum of deterministic predictions of one-minute counts at several adjacent (earlier and later) intervals. The standard deviation is also expressed via a linear function of those deterministic predictions. The study determined the rules for determining the number of deterministic predictions at adjacent intervals that should be used in the probabilistic aggregate demand counts as well as the weight coefficients for both expected values and standard deviations of aggregate one-minute sector entry counts. The weight coefficients are expressed through the probabilities for individual flights to enter a sector during various one-minute intervals.
- The study also resulted in developing an analytical approach and methodology for translating characteristics of uncertainty in individual flight's predictions into probabilistic predictions for sector demand counts. It was shown that probabilistic predictions of the number of flights in a sector are expressed through probabilistic predictions of number of flights entering a sector during several consecutive one-minute intervals that, in turn, depend on probabilities for individual flights to enter a sector during various consecutive one-minute intervals. The number of intervals involved in the probabilistic predictions depends on both the accuracy of predicting times for individual flights and the predicted time-in sector for the flights.
- The characteristics of uncertainty in sector traffic demand predictions include the most likely, expected values of traffic demand counts and the range of uncertainty around those expected values. The probabilistic demand predictions along with predicted capacities then can be used together to help determine the likelihood of congestion in NAS elements.

The next step for this analysis, not undertaken in this report, is to determine exactly how these findings can be incorporated into the TFMS Monitor/Alert function, thus, leading to better decision making by traffic managers who are dealing with sector congestion.

Appendix A. Probability for a Flight to Enter a Sector During One-minute Interval

A.1 General Case

Consider a sequence of one-minute intervals during each of which the flight enters a sector. In this section, we estimate the probability for an individual flight to enter a sector during a specific one-minute interval given the ETA (Estimated Time of Arrival) for a flight to enter the sector and the probability distribution of the errors in predicting the time of sector boundary crossing.

The following notations are used in this report:

f(t) – probability density function (pdf) of error in sector entry time predictions (the error is measured as predicted minus actual sector entry time)

F(t) – cumulative probability function of error in sector entry time prediction; $F(t) = \int_{-\infty}^{t} f(y)\,dy$

i is a one-minute interval $[i, i + 1)$[2] that starts at time i and ends in time (i+1).

Note: the probability distributions of prediction errors and their parameters depend on both flights' status (active or proposed) and look-ahead time (LAT).

Consider a flight that is deterministically forecast to arrive at a sector (cross a sector boundary) at time *x*, i.e., *x* is equal to a flight's ETA. Let the ETA fall within a one-minute interval k (i.e., the ETA is within $[k, k+1)$. Note: interval *k* can be equal to *i*, or earlier or later than i. What is the probability that the flight can arrive during the one-minute interval i if it is deterministically forecasted to arrive during one-minute interval k? We denote this conditional probability as $P(\ i \mid \text{ETA in } k) = P_{i,\,k}(\text{ETA in } k)$, so that

$$P_{i,\,k}(\text{ETA in } k) = \int_{i}^{i+1} f(y - ETA)\,dy = F(i - \text{ETA} + 1) - F(i - \text{ETA});\ \ k \leq \text{ETA} < k + 1. \quad \text{(A1)}$$

According to analysis of errors in sector entry time predictions (see Tables 3-2 – 3-5), the standard deviation of errors are roughly within 4 -15 minutes? for active flights and 20+ minutes for proposed flights and the values vary by sector and LAT. As a result, the probabilities (A1) for individual flights to arrive in a sector during a one-minute interval are small even in the case of $i = k$.

Figures A1, A2 and A3 illustrate it in the cases when ETAs are within the intervals $k = i$ (interval if interest), and two earlier intervals $k = i - 1$, and $k = i - 3$, respectively. In these Figures, the probabilities $P(\ i \mid \text{ETA in } k)$ are equal to the striped areas with a one-minute width under the pdf $y = f(t - \text{ETA})$. The figures also show that the farther interval k is from the interval of interest *i*, the smaller the probability $P(\ i \mid \text{ETA in } k)$.

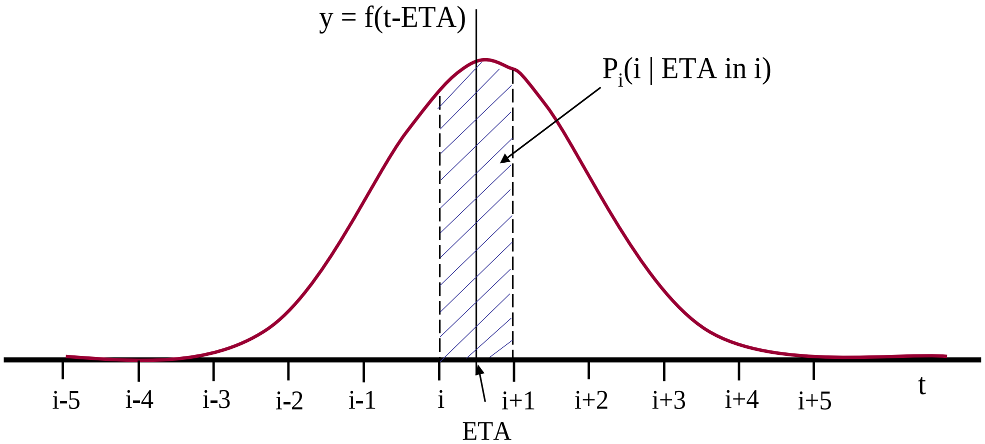

Figure A1 Probability for a flight with ETA within interval i to enter a sector within interval i

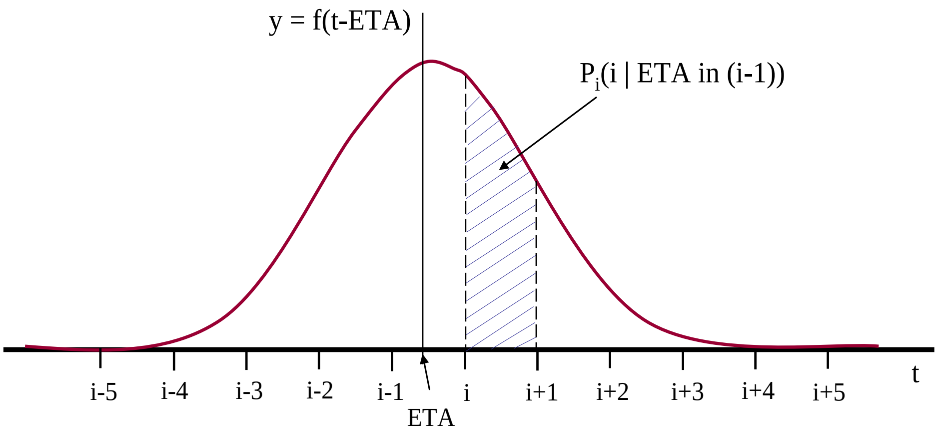

Figure A2 Probability for a flight with ETA within interval (i-1) to enter a sector within interval i

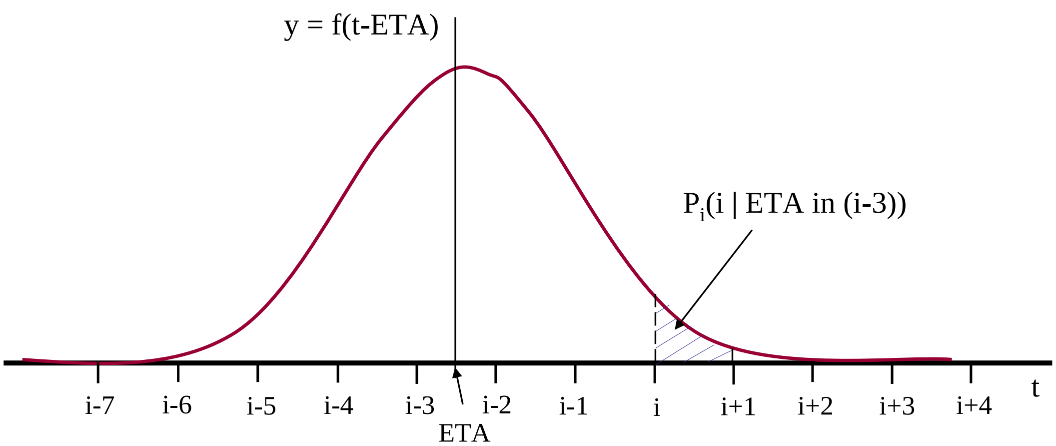

Figure A3 Probability for a flight with ETA within interval (i-3) to enter a sector within interval i

Let us consider a set of flights with ETAs within the one-minute interval k. The number of flights in this set is equal to deterministically predicted one-minute aggregate demand d_k for entering a sector within interval *k*. In this interval, the ETAs are very close and the difference between the earliest $ETA^{(E)}$ and the latest $ETA^{(L)}$ does not exceed one minute. Therefore, it is reasonable to assume that $ETA^{(E)} \approx k$, and $ETA^{(L)} \approx k+1$. The probability for the flight with the earliest ETA in interval *k* to arrive within interval *i* is

$$P_{i,k}^{(E)}\,(ETA^{(E)} = k) \;=\; \int_{i}^{i+1} f\,(y - k)\,dy \;=\; F(i - k + 1) - F(i - k). \qquad \text{(A2)}$$

The probability for the flight with the latest ETA in interval k to arrive within interval i is

$$P_{i,k}^{(L)}\,(ETA^{(L)} = k + 1) \;=\; \int_{i}^{i+1} f\,(y - (k+1))\,dy \;=\; F(i - k) - F(i - (k + 1)). \qquad \text{(A3)}$$

For the rest of the flights with ETAs within the one-minute interval *k* the probabilities to enter a sector during the one-minute interval *i* are between the probabilities $P_{i,k}^{(E)}\,(ETA^{(E)} = k)$ and $P_{i,k}^{(L)}\,(ETA^{(L)} = k + 1)$. As those probabilities are close in their values (because the standard deviation of prediction errors is at least four minutes), we can assume that the probability for each flight deterministically predicted to enter a sector during interval k to enter the sector during a one-minute interval i is close to the constant that is equal to the average of $P_{i,k}^{(E)}\,(ETA^{(E)} = k)$ and $P_{i,k}^{(E)}\,(ETA^{(E)} = k)$:

$$P_{i,k} \approx 0.5\,(\,P_{i,k}^{(E)}\,(ETA^{(E)} = k) + P_{i,k}^{(L)}\,(ETA^{(L)} = k + 1)\,),$$

where $P_{i,k}$ is a probability for a flight, deterministically predicted to enter a sector during a one-minute interval k, to enter the sector during a one-minute interval i.

After substituting $P_{i,k}^{(E)}\,(ETA^{(E)} = k)$ and $P_{i,k}^{(L)}\,(ETA^{(L)} = k + 1)$ by their expressions from (A2) and (A3), the final equation for determining the probability looks as follows:

$$P_{i,k} \approx 0.5\,[F(i - k + 1) - F(i - k - 1)]. \qquad \text{(A4)}$$

Note that the probabilities $P_{i,k}$ depend only on the distance $(i - k)$ between one-minute intervals *i* and *k*.

A.2 Active and Proposed Flights

Our previous analysis (see [1]) showed that the ETA prediction accuracy is different for active and proposed flights: the ETA predictions for active flights are more accurate. Therefore, the flight's status should be taken into account for determining the probabilities $P_{i,k}$ (A4).

Let's denote the probability density functions of errors in sector entry time predictions for active and proposed flights as $f^{(a)}(t)$ and $f^{(g)}(t)$, respectively, with the corresponding cumulative distribution functions $F^{(a)}(t)$ and $F^{(g)}(t)$.

Then, according to (A4), the probability $P_{i,k}^{(a)}$ for an active flight, deterministically predicted to enter a sector during a one-minute interval k, to enter the sector during a one-minute interval i is equal to

$$P_{i,k}^{(a)} \approx 0.5\ [F^{(a)}(i - k + 1) - F^{(a)}(i - k - 1)], \tag{A5}$$

and the probability $P_{i,k}^{(g)}$ for a flight still on the ground, deterministically predicted to enter a sector during a one-minute interval k, to enter the sector during a one-minute interval i is equal to

$$P_{i,k}^{(g)} \approx 0.5\ [F^{(g)}(i - k + 1) - F^{(g)}(i - k - 1)]. \tag{A6}$$

Appendix B. Probabilistic Characterization of Number of Flights Entering Sector During One-minute Interval

B.1 Basic Case

Let us introduce a random number of flights $d_{i,k}$ that, due to random errors in predicting flights' times of entering a sector, can enter a sector during one-minute interval i from the set of flights d_k deterministically predicted to enter the sector during one-minute interval k.

Using the probabilities $P_{i,k}$ from Appendix A, we can determine the probability distribution of a random number of flights $d_{i,k}$. Assuming the flights are independent, the probability distribution of the number of flights from the set of d_k flights each of which has a probability $P_{i,k}$ (A4) to enter a sector during interval i is a binomial distribution:

$$P_i^{(k)}(x) = C_x^{d_k} \; P_{i,k}^x \; (1 - P_{i,k})^{(d_k - x)}, \quad x = 0, 1, \ldots, d_k, \qquad \text{(B1)}$$

where

$P_i^{(k)}(x)$ is the probability for x flights from the set of d_k flights deterministically predicted to enter the sector during interval k to enter the sector during interval i;

$C_x^{d_k} = \dfrac{d_k!}{x!(d_k - x)!}$ is a number of combinations of x flights from the set of d_k flights, $0 \le x \le d_k$;

$x! = 1*2*\ldots*(x-1)*x$. Note: $0! = 1$.

A binomial distribution (B1) with parameters d_k and $P_{i,k}$ has mean $P_{i,k} d_k$ and variance $P_{i,k}(1 - P_{i,k})\, d_k$. (See Reference [19]). Furthermore, if $(1 - P_{i,k})\, d_k$ and $P_{i,k}\, d_k$ are sufficiently large (> 5), the distribution can be approximated by a normal distribution[19].

Hence, according to (B1), the mean $\overline{d}_{i,k}$, variance $\sigma_{i,k}^{\;2}$ and the corresponding standard deviation $\sigma_{i,k}$ of number of flights predicted to enter a sector during a one-minute interval i from the set of d_k flights deterministically predicted to enter the sector during interval k are equal to, respectively:

$$\overline{d_{i,k}} = P_{i,k}\, d_k, \qquad \text{(B2)}$$

$$\sigma_{i,k}^{\;2} = P_{i,k}(1 - P_{i,k})\, d_k \qquad \text{(B3)}$$

$$\sigma_{i,k} = \sqrt{P_{i,k}(1 - P_{i,k})\, d_k} \qquad \text{(B4)}$$

If $P_{i,k} << 1$, the variance (B3) is approximately equal to

$$\sigma_{i,k}^{\;2} \approx P_{i,k}\, d_k, \qquad \text{(B5)}$$

Hence, the standard deviation of the number of flights entering a sector during a one-minute interval *i* from the set of flights deterministically predicted to enter the sector during a one-minute interval k is approximately equal to

$$\sigma_{i,k} \approx \sqrt{P_{i,k}\, d_k} \qquad \text{(B6)}$$

Now we can illustrate why traffic managers do not trust predictions of a peak one-minute count of flights in a sector for monitor/alert purposes. It is because the probability for all deterministically predicted flights to enter a sector during a one-minute interval is usually very small, as shown below.

If d_i is a deterministic prediction of the number of flights entering a sector during a one-minute interval i, then, according to (B1), the probability distribution of the number of flights from the set of d_i that can enter the sector during interval i is determined as follows

$$P_i^{(i)}(x) = C_x^{d_i} \; P_{i,i}^{x} \; (1 - P_{i,i})^{(d_i - x)}, \quad x = 0, 1, \ldots, d_i, \qquad \text{(B7)}$$

Determine the probability for deterministic prediction to be correct. It is equal to the probability for all d_i flights to enter a sector during a one-minute interval *i*. This probability can be obtained from (B7) for $x = d_i$, and it is equal to

$$P_i^{(i)}(d_i) = P_{i,i}^{d_i}. \qquad \text{(B8)}$$

In other words, this probability for deterministic prediction to be correct is equal to the probability for a flight to enter a sector during interval *i* in power d_i. For $P_{i,i} << 1$ and $d_i >> 1$, the probability (B8) can be extremely small.

For example, for $P_{i,i} = 0.1$ and $d_i = 10$, the probability for all ten flights to enter a sector during a one-minute interval *i* is equal to 10^{-10}, i.e. it is negligibly small and practically equal to zero.

Moreover, according to (B2), the expected number of flights from d_i to enter a sector during interval i is equal to

$$\overline{d_{i,i}} = P_{i,i}\, d_i,$$

and for $P_{i,i} << 1$ is a small fraction of deterministically predicted number of flights. In the above example, the expected number is equal to $\overline{d_{i,i}} = 1$, i.e., it is only one tenth of the deterministically predicted $d_i = 10$ flights.

Note: in reality, however, it is not so bad because the stochastic mechanism aggregates some flights from several adjacent intervals into the interval of interest that gives a reasonable number for aggregate sector demand. In other words, although predictions for concrete flights to be in a one-minute interval are highly unreliable, and there might be a significant change in the list of concrete flights in the series of updates for a one-minute interval, predictions of aggregate number of flights might be more stable.

Formula (A4) shows that the probability $P_{i,k}$ depends on the difference $(i - k)$, i.e., on the time distance between the one-minute interval of interest *i* and interval *k*, where the flight was deterministically predicted to enter a sector. It is clear that the farther interval *k* is from interval *i* the smaller the probability for a flight predicted to enter a sector during interval *k* to enter the sector during interval *i* (e.g., see Figures 1, 2, and 3). Therefore, it is necessary to determine the maximum distance between intervals *i* and *k*, beyond which the probabilities $P_{i,k}$ are too small and should be neglected. This would allow to restrict the number of intervals *k* around the interval of interest *i* and, hence, the number of deterministic predictions d_k, which can contribute to the probabilistic prediction of number of flights entering the sector during interval i. Let's denote the maximum distance as:

$$\beta = \max | i - k |. \qquad \text{(B9)}$$

Then the total random number of flights $\tilde{d}_i$ predicted to enter a sector during a one-minute interval i is equal to the sum of random numbers of flights $d_{i,k}$ from the sets of the flights d_k deterministically predicted to enter the sector during one-minute intervals k ($k = i - \beta,\ i - \beta + 1, \ldots, i,\ i + 1, \ldots, i + \beta$):

$$\tilde{d}_i = \sum_{k=i-\beta}^{i+\beta} d_{i,k} \qquad \text{(B10)}$$

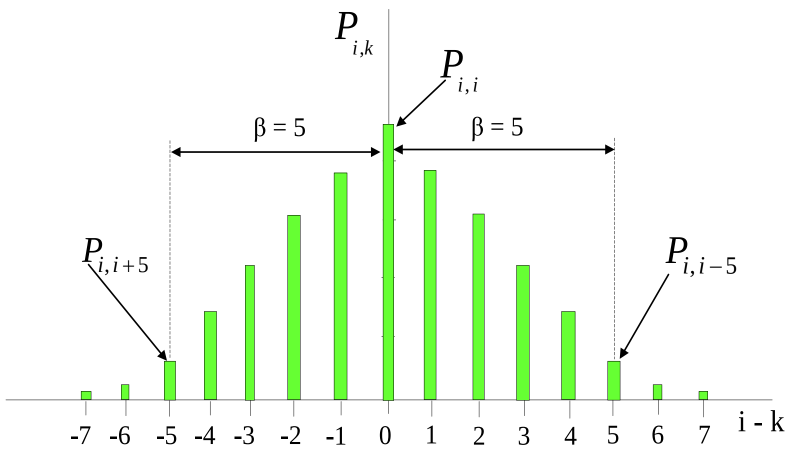

Figure B1 Probabilities $P_{i,k}$ as a function of (i – k)

Figure B1 illustrates an example of how the probabilities $P_{i,k}$ depend on $|i - k|$. In this Figure, the probabilities $P_{i,k}$ decrease sharply at $|i - k| = 6$, and can be considered negligible small for $|i - k| > 5$, so that $\beta = 5$. It means that the flights deterministically predicted to enter the sector up to five minutes earlier and up to five minutes later than the interval of interest might noticeably contribute to the predicted flight counts for the one-minute interval of interest.

Analysis of $P_{i,k}$ values for the probability distributions of prediction errors with different standard deviations σ showed that it is practically reasonable to restrict parameter β by two sigmas, i.e., $\beta = 2\sigma$. It means that the flights deterministically predicted to enter the sector up to 2σ minutes earlier and up to 2σ minutes later of the interval of interest might noticeably contribute to the predicted flight counts for the one-minute interval of interest. In other words, to probabilistically estimate the number of flights entering a sector during a one-minute interval of interest, one needs to process $2\beta + 1 = 4\sigma + 1$ deterministic predictions of number of flights to enter the sector in $(4\sigma + 1)$ consecutive one-minute intervals.

According to (B10) and (B2), the average (expected, mean) number of flights predicted to enter a sector during a one-minute interval i is equal to

$$\bar{\tilde{d}}_i = \sum_{k=i-\beta}^{i+\beta} \overline{d_{i,k}} = \sum_{k=i-\beta}^{i+\beta} P_{i,k} d_k . \qquad \text{(B11)}$$

As the sets of flights entering a sector in different one-minute intervals contain different flights, one can assume that the numbers of flights crossing the sector in different one-minute intervals are independent. Therefore, the variance of number of flights entering the sector during a one-minute interval i is equal to

$$\sigma_i^2 = \sum_{k=i-\beta}^{i+\beta} \sigma_{i,k}^2 = \sum_{k=i-\beta}^{i+\beta} P_{i,k}(1-P_{i,k})\, d_k, \qquad \text{(B12)}$$

and the standard deviation of number of flights entering the sector during a one-minute interval *i* is

$$\sigma_i = \sqrt{\sum_{k=i-\beta}^{i+\beta} P_{i,k}(1-P_{i,k})\, d_k}\ . \qquad \text{(B13)}$$

In the case of $P_{i,k} << 1$, the standard deviation (B13) can be approximately determined as

$$\sigma_i \approx \sqrt{\sum_{k=i-\beta}^{i+\beta} P_{i,k}\, d_k} \qquad \text{(B14)}$$

B.2 Active and Proposed Flights in the Predicted Demand Counts

The number of flights deterministically predicted to enter a sector during the k^{th} one-minute interval may include both active and proposed flights:

$$d_k = d_k^{(a)} + d_k^{(g)}, \qquad \text{(B15)}$$

where $d_k^{(a)}$ and $d_k^{(g)}$ are active and proposed components of the predicted number of flights entering the sector.

Hence, to determine the probability distribution of the number of flights from the set of $d_k = d_k^{(a)} + d_k^{(g)}$ flights that can enter a sector during one-minute interval *i* we will need to take into consideration the flights' status.

For the active component $d_k^{(a)}$, the probability $P_{i,k}^{(a)}$ for each flight to enter a sector during a one-minute interval *i* is given by equation (A5), so that the probability distribution $P_i^{(k),(a)}(x)$ of number of active flights x from the set of $d_k^{(a)}$ flights to enter a sector during a one-minute interval *i* is a binomial distribution:

$$P_i^{(k),(a)}(x) = C_x^{d_k^{(a)}} (P_{i,k}^{(a)})^x (1-P_{i,k}^{(a)})^{d_k^{(a)}-x},\ x = 0, 1, \ldots, d_k^{(a)}. \qquad \text{(B16)}$$

Similarly, for the proposed component $d_k^{(g)}$, the probability distribution $P_i^{(k),(g)}(x)$ of the number of proposed flights *x* from the set of $d_k^{(g)}$ flights to enter a sector during a one-minute interval is also a binomial distribution

$$P_i^{(k),(g)}(x) = C_x^{d_k^{(g)}} (P_{i,k}^{(g)})^x (1-P_{i,k}^{(g)})^{d_k^{(g)}-x},\ x = 0, 1, \ldots, d_k^{(g)}, \qquad \text{(B17)}$$

where the probability for a proposed flight to enter a sector during a one-minute interval i if it was deterministically predicted to enter a sector within interval k is given by formula (A6).

Using equations (B16) and (B17), we can determine the probability $P_i^{(k)}(x)$ of the number of flights x from the set of flights $d_k = d_k^{(a)} + d_k^{(g)}$, deterministically predicted to enter a sector during a one-minute interval *k*, to enter a sector during interval *i* as follows:

$$P_i^{(k)}(x) = \sum_{j=0}^{d_k^{(g)}} P_i^{(k),(g)}(j)\, P_i^{(k),(a)}(x-j) \text{ ; } x = 0, 1, 2, \ldots, d_k^{(a)} + d_k^{(g)} \quad \text{(B18)}$$

where $0 \le j \le x \le d_k^{(a)} + j$.

Equation (B18) is based on considering the probabilities for all possible combinations of numbers of active and proposed flights in the total deterministic predictions d_k.

After using expressions (B16) and (B17), the equation (B18) for the probability can be rewritten as follows

$$P_i^{(k)}(x) = \sum_{j=0}^{d_k^{(g)}} C_j^{d_k^{(g)}} C_{x-j}^{d_k^{(a)}} (P_{i,k}^{(g)})^j (1-P_{i,k}^{(g)})^{d_k^{(g)}-j} (P_{i,k}^{(a)})^{x-j} (1-P_{i,k}^{(a)})^{d_k^{(a)}-x+j} \text{; } 0 \le j \le x \le d_k^{(a)} + j \le d_k \text{, (B19)}$$

where $C_j^n = \dfrac{n!}{j!(n-j)!}$ is a number of combinations of j flights from the set of n flights ($0 \le j \le n$).

The mean (average, expected) number of airborne flights from $d_k^{(a)}$, deterministically predicted to enter a sector during a one-minute interval k, to enter a sector during a one-minute interval i is equal to:

$$\overline{d_{i,k}^{(a)}} = P_{i,k}^{(a)}\, d_k^{(a)} \text{ .} \quad \text{(B20)}$$

The mean number of proposed flights from $d_k^{(g)}$, deterministically predicted to enter a sector during a one-minute interval k, to enter a sector during a one-minute interval i is

$$\overline{d_{i,k}^{(g)}} = P_{i,k}^{(g)}\, d_k^{(g)} \text{ .} \quad \text{(B21)}$$

The variance $\sigma_{i,k}^{(a)2}$ of number of airborne flights from $d_k^{(a)}$ to enter a sector during a one-minute interval i is equal to:

$$\sigma_{i,k}^{(a)2} = P_{i,k}^{(a)} (1 - P_{i,k}^{(a)})\, d_k^{(a)} \quad \text{(B22)}$$

The variance $\sigma_{i,k}^{(g)2}$ of number of proposed flights from $d_k^{(g)}$ to enter a sector during a one-minute interval i is

$$\sigma_{i,k}^{(g)2} = P_{i,k}^{(g)} (1 - P_{i,k}^{(g)})\, d_k^{(g)} \text{ .} \quad \text{(B23)}$$

As was discussed above, because of errors in predicting times of sector boundary crossings, there might be a set of one-minute intervals around a one-minute interval k from where the flights have substantial probability to actually enter a sector during a one-minute interval i and should be accounted for in the total flight counts in this interval. Parameter β was introduced to represent the number of minutes from each side of interval k that can affect traffic counts in the one-minute interval i.

As the probability distributions of errors in predicting sector entry times for airborne and proposed flights are different, parameter β can be separately determined for airborne and proposed flights (see. Fig. B1)

Let $\beta = \beta_1$ for airborne flights and $\beta = \beta_2$ for proposed flights.

Then, similar to (B11), the mean number of active flights and the mean number of proposed to enter a sector during a one-minute interval i are, respectively, equal to

$$\overline{d_i^{(a)}} = \sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} d_k^{(a)} , \qquad \text{(B24)}$$

$$\overline{d_i^{(g)}} = \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} d_k^{(g)} . \qquad \text{(B25)}$$

Hence, the mean number of flights to enter a sector during interval i is equal to

$$\overline{d_i} = \overline{d_i^{(a)}} + \overline{d_i^{(g)}} = \sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} d_k^{(a)} + \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} d_k^{(g)} \qquad \text{(B26)}$$

After using (B12) for each of the subsets of airborne and proposed flights, the variance of flights entering the sector during the one-minute interval i is

$$\sigma_i^2 = \sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} (1-P_{i,k}^{(a)}) d_k^{(a)} + \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} (1-P_{i,k}^{(g)}) d_k^{(g)} . \qquad \text{(B27)}$$

If $P_{i,k}^{(a)} << 1$ and $P_{i,k}^{(g)} << 1$, then the variance (B28) is approximately equal to

$$\sigma_i^2 \approx \sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} d_k^{(a)} + \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} d_k^{(g)} . \qquad \text{(B28)}$$

According to (B27), the standard deviation of the number of flights entering a sector during one-minute interval i is

$$\sigma_i = \sqrt{\sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} (1-P_{i,k}^{(a)}) d_k^{(a)} + \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} (1-P_{i,k}^{(g)}) d_k^{(g)}} . \qquad \text{(B29)}$$

or, in the case when $P_{i,k}^{(a)} << 1$ and $P_{i,k}^{(g)} << 1$,

$$\sigma_i \approx \sqrt{\sum_{k=i-\beta_1}^{i+\beta_1} P_{i,k}^{(a)} d_k^{(a)} + \sum_{k=i-\beta_2}^{i+\beta_2} P_{i,k}^{(g)} d_k^{(g)}} . \tag{B30}$$

Thus, the expressions (B26), (B29) and (B30) make it possible to determine average and standard deviation of number of flights entering a sector during a one-minute interval of interest through the weighted sums of deterministic predictions of number of active and proposed flights for both the interval of interest and several one-minute intervals surrounding the interval of interest. The weight coefficients represent probabilities for individual flights to enter a sector during the one-minute interval of interest if they are deterministically predicted to enter a sector during some specific one minute intervals (including the interval of interest).

The probabilistic predictions of one-minute counts of the flights entering a sector include the mean value $\bar{d}_i$ and the uncertainty range between $(\bar{d}_i - j\,\sigma_i)$ and $(\bar{d}_i + j\,\sigma_i)$, where *j* can be assigned depending on the desired percentiles for uncertainty area. For $j = 1$, the uncertainty area covers plus/minus one sigma interval around the mean value between 0.16 and 0.84 percentiles.

Appendix C. Probabilistic Characterization of Number of Flights in a Sector During a One minute Interval

In order to determine whether a flight is within a sector during a certain one minute interval, one needs to know the flight's entry and exit time (the difference gives the total time for a flight to be in the sector, or the time required to traverse the sector). Aggregating all the flights that are predicted to be within the sector during a specific one-minute interval would provide the aggregate sector traffic demand for this minute.

C.1 Flights entering a sector have the same time in the sector

Suppose that the time to traverse the sector for each flight entering the sector is the same and equal to τ minutes. It means that each flight spends $\boldsymbol{\tau}$ minutes within the sector since its entry before leaving the sector. Therefore, for each one-minute interval of interest, for example, interval i, the flights will be in the sector during this interval if they entered the sector during the interval i and $(\boldsymbol{\tau} - 1)$ preceding one-minute intervals, i.e., during the one-minute intervals $i, i - 1, i - 2, \ldots, i - (\boldsymbol{\tau} - 2), i - (\boldsymbol{\tau} - 1)$.This provides the opportunity to determine the sector traffic demand count predictions at any one-minute interval through the series of one-minute aggregate number of flights *entering* the sector during several consecutive one-minute intervals. In particular, deterministically predicted traffic demand D_i in a sector for a one-minute interval i is equal to

$$D_{i,\tau} = d_i + d_{i-1} + d_{i-2} + \ldots + d_{i-(\tau-2)} + d_{i-(\tau-1)} = \sum_{j=i-\tau+1}^{i} d_j \,, \qquad \text{(C1)}$$

where d_k is deterministically predicted number of flights entering a sector during a one-minute interval k.

It is important to notice that each component d_k in (C1) consists of different flights.

The deterministically predicted number of flights in a sector during s one-minute intervals immediately following the interval i can be expressed recursively as follows

$$D_{i+s,\tau} = d_{i+s} + D_{i+s-1,\tau} - d_{i-\tau+s}, \quad s = 0, 1, 2, 3, \ldots \qquad \text{(C2)}$$

Formula (C2) shows that the one-minute sector demand for the $(i + s)$ minute is equal to the demand prediction in the previous $(i + s - 1)$ one-minute interval plus the number of flights entering a sector during the $(i + s)$ minute minus the number of flights $d_{i-\tau+s}$ that left the sector during the $(i + s)$ minute (those flights are the ones that entered the sector τ minutes prior to the $(i + s)$ minute).

Due to random errors in predicting number of flights that enter a sector, the sector demand for a one-minute interval is actually a random number that can be represented by a formula like (C1), where each component of the sum is a random number of flights entering a sector during the corresponding one-minute intervals. If $\tilde{D}_{i,\tau}$ is the random number of flights predicted to be in a sector during one-minute interval i then it can be determined as follows:

$$\tilde{D}_{i,\tau} = \tilde{d}_i + \tilde{d}_{i-1} + \tilde{d}_{i-2} + \ldots + \tilde{d}_{i-(\tau-2)} + \tilde{d}_{i-(\tau-1)} = \sum_{j=i-\tau+1}^{i} \tilde{d}_j , \qquad \text{(C3)}$$

where $\tilde{d}_j$ is a random number of flights predicted to enter a sector during a one-minute interval *j*.

Therefore, according to (C3), the average predicted traffic demand $\bar{\tilde{D}}_i$ in a sector for a one-minute interval i is equal to the sum of average numbers of flights entering the sector during τ consecutive one-minute intervals: during the interval *i* and *(τ – 1)* intervals preceding interval *i*:

$$\bar{\tilde{D}}_{i,\tau} = \sum_{j=i-\tau+1}^{i} \bar{\tilde{d}}_j , \qquad \text{(C4)}$$

where the average number of flights $\bar{\tilde{d}}_j$ predicted to cross the sector boundaries during one-minute interval *j* (for $j = i, i-1, i-2, \ldots, i-\tau+1$) are determined by formula (B11).

After substituting $\bar{\tilde{d}}_j$ in (C4) with (B11), the equation (C4) is as follows:

$$\bar{\tilde{D}}_{i,\tau} = \sum_{j=i-\tau+1}^{i} \sum_{k=j-\beta}^{j+\beta} P_{j,k} d_k . \qquad \text{(C5)}$$

After several simple transformations, equation (C5) can be rewritten as follows:

$$\bar{\tilde{D}}_{i,\tau} = \sum_{k=i-\tau-\beta+1}^{i+\beta} P_{i,k,\tau}\, d_k , \qquad \text{(C6)}$$

where $P_{i,k,\tau}$ is the probability for a flight to be in a sector during a one-minute interval *i* if it is deterministically predicted to enter a sector during one-minute interval *k* and be in a sector during time interval τ.

The probabilities $P_{i,k,\tau}$ are determined by the following summation, for various values of *k*:

$$P_{i,k,\tau} = \sum_{j=\max(k-\beta,\, i-\tau+1)}^{\min(i,\, k+\beta)} P_{j,k} , \text{ where } i-\beta-\tau+1 \le k \le i+\beta , \qquad \text{(C7)}$$

where $P_{j,k}$ is a probability for a flight to enter a sector during a one-minute interval j if its ETA to a sector is in the one-minute interval k (see formula (A4) from Appendix A). In other words, this summation states that:

- The relevant values of *j* are those ranging from $i-\tau+1$ to *i*, and
- We are only interested in those values of *j* that are within β of *k*.

If a flight deterministically predicted to enter a sector during interval k has the probability $P_{i,k,\tau}$ to be in a sector during interval *i,* then the predicted number of flights in the sector during the one-minute interval *i* is a binomially distributed random number.

The variance Var ($\widetilde{D}_{i,\tau}$) and standard deviation σ($\widetilde{D}_{i,\tau}$) of one-minute counts in a sector during minute i are, respectively, equal to

$$\mathrm{Var}\,(\widetilde{D}_{i,\tau}) = , \sum_{k=i-\beta-\tau+1}^{i+\beta} P_{i,k,\tau}(1-P_{i,k,\tau})\, d_k \tag{C8}$$

$$\sigma(\widetilde{D}_{i,\tau}) = \sqrt{\sum_{k=i-\beta-\tau+1}^{i+\beta} P_{i,k,\tau}(1-P_{i,k,\tau})\, d_k}\ . \tag{C9}$$

When the sets of deterministically predicted flights d_k contain both active and proposed flights, $d_k = d_k^{(a)} + d_k^{(g)}$, equations (C8) – (C9) can be used to determine corresponding averages and standard deviations separately for the two types of flights.

For active flights the mean, variance and standard deviation of number of flights in a sector during one-minute interval *i* are, respectively, equal to

$$\overline{\widetilde{D}}{}^{(a)}_{i,\tau} = \sum_{k=i-\tau-\beta_1+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}\, d^{(a)}_{k} , \tag{C10}$$

$$\mathrm{Var}\,(\widetilde{D}^{(a)}_{i,\tau}) = \sum_{k=i-\beta_1-\tau+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}(1-P^{(a)}_{i,k,\tau})\, d^{(a)}_{k} , \tag{C11}$$

$$\sigma^{(a)}(\widetilde{D}^{(a)}_{i,\tau}) = \sqrt{\sum_{k=i-\beta_1-\tau+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}(1-P^{(a)}_{i,k,\tau})\, d^{(a)}_{k}}\ . \tag{C12}$$

The probabilities $P^{(a)}_{i,k,\tau}$ for an active flight to be in a sector during interval *i*, if it was deterministically predicted to cross sector boundary during interval *k*, can be found from equations (C7) by substituting $P_{j,k}$, *β*, and d_k with $P^{(a)}_{j,k,}$, β_1, and $d^{(a)}_{k}$, respectively (see Appendix B).

For proposed flights the mean, variance and standard deviation of number of flights in a sector during one-minute interval i are, respectively, equal to

$$\overline{\widetilde{D}}{}^{(g)}_{i,\tau} = \sum_{k=i-\tau-\beta_2+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}\, d^{(g)}_{k} , \tag{C13}$$

$$\mathrm{Var}\,(\widetilde{D}^{(g)}_{i,\tau}) = \sum_{k=i-\beta_2-\tau+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}(1-P^{(g)}_{i,k,\tau})\, d^{(g)}_{k} , \tag{C14}$$

$$\sigma^{(g)}(\widetilde{D}^{(g)}_{i,\tau}) = \sqrt{\sum_{k=i-\beta_2-\tau+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}(1-P^{(g)}_{i,k,\tau})\, d^{(g)}_{k}}\ . \tag{C15}$$

The probabilities $P^{(g)}_{i,k,\tau}$ for a proposed flight to be in a sector during interval *i*, if it was deterministically predicted to cross sector boundary during interval *k*, can be found from equation (C7) by substituting $P_{j,k}$, β, and d_k with $P^{(g)}_{j,k}$, β_2, and $d^{(g)}_k$, respectively (see Appendix B).

As a result, the mean $\overline{\tilde{D}}_{i,\tau}$ and variance Var ($\tilde{D}_{i,\tau}$) of predicted one-minute sector demand during interval *i*, when deterministically predicted one-minute sector entry counts contain both active and proposed flights, are determined by summation of expected (mean) values and variances, respectively, for active and proposed flights:

$$\overline{\tilde{D}}_{i,\tau} = \overline{\tilde{D}}^{(a)}_{i,\tau} + \overline{\tilde{D}}^{(g)}_{i,\tau} = \sum_{k=i-\tau-\beta_1+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}\, d^{(a)}_k + \sum_{k=i-\tau-\beta_2+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}\, d^{(g)}_k \tag{C16}$$

$$\mathrm{Var}\,(\tilde{D}_{i,\tau}) = \mathrm{Var}\,(\tilde{D}^{(a)}_{i,\tau}) + \mathrm{Var}\,(\tilde{D}^{(g)}_{i,\tau}) =$$

$$\sum_{k=i-\beta_1-\tau+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}(1-P^{(a)}_{i,k,\tau})\, d^{(a)}_k + \sum_{k=i-\beta_2-\tau+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}(1-P^{(g)}_{i,k,\tau})\, d^{(g)}_k\,. \tag{C17}$$

The standard deviation σ($\tilde{D}_{i,\tau}$) of predicted one-minute demand in a sector is equal to square root of (C17):

$$\sigma(\tilde{D}_{i,\tau}) = \sqrt{\sum_{k=i-\beta_1-\tau+1}^{i+\beta_1} P^{(a)}_{i,k,\tau}(1-P^{(a)}_{i,k,\tau})\, d^{(a)}_k + \sum_{k=i-\beta_2-\tau+1}^{i+\beta_2} P^{(g)}_{i,k,\tau}(1-P^{(g)}_{i,k,\tau})\, d^{(g)}_k}\,. \tag{C18}$$

This section presented the basic results necessary for probabilistic characterization of one-minute sector traffic demand that consists of flights with the same time for traversing a sector.

These results are used in Section C.2 below that covers more realistic demand scenarios when the flights predicted to enter a sector require different times for traversing the sector.

C.2 Flights entering a sector have different times in the sector

Consider the case when flights may have different times in sector. Generally, sectors contain several routes that have different lengths so that for the flights using different routes there might be different times in the sector (to traverse the sector). This factor should be taken into consideration for predicting number of flights in the sector. There are two ways of doing it.

1. For each sector's route structure, the average time of traversing can be determined, and the routes requiring practically the same time τ to traverse can be put in a single cluster so that a sector might have several route clusters. As a result, using the flights' flight plans that contain the routes, the aggregate number of flights (flight counts) entering a sector during a specific one-minute interval can be broken down by number of flights associated with specific route clusters, and, hence, by number of flights with practically the same time in the sector.

2. Instead of associating flights with specific routes, clustering the flights by time-in-sector can be done by direct calculation of time-in-sector for each flight through the difference between predicted times of leaving and entering the sector that are estimated in the TFMS.

Suppose, a sector has m clusters of flights with average times-in-sector $\tau_1, \tau_2, \ldots, \tau_m$, respectively. Suppose also that $\tau_1 < \tau_2 \ldots < \tau_m$. Practically, the number of clusters m can be restricted by 3 or 4.

Then the deterministically predicted number of flights entering a sector during one-minute interval i can be represented by a sum

$$d_i = \sum_{j=1}^{m} d_{i,\tau_j}, \tag{C19}$$

where d_{i,τ_j} is the number of flights predicted to enter a sector during one-minute interval i with the time-in-sector τ_j ($j = 1, 2, \ldots, m$). Traffic demand d_{i,τ_j} can contain number of active flights $d^{(a)}_{i,\tau_j}$ and the number of proposed flights $d^{(g)}_{i,\tau_j}$, so that

$$d_{i,\tau_j} = d^{(a)}_{i,\tau_j} + d^{(g)}_{i,\tau_j}. \tag{C20}$$

The fraction D_{i,τ_j} of the total number of flights D_i in the sector deterministically predicted for one-minute interval i, that contains flights with the time-in-sector τ_j is

$$D_{i,\tau_j} = \sum_{q=i-\tau_j+1}^{i} d_{q,\tau_j} = \sum_{q=i-\tau_j+1}^{i} (d^{(a)}_{q,\tau_j} + d^{(g)}_{q,\tau_j}), \tag{C21}$$

Then the total number of flights D_i deterministically predicted in a sector during one-minute interval i is

$$D_i = \sum_{j=1}^{m} D_{i,\tau_j} = \sum_{j=1}^{m} \sum_{q=i-\tau_j+1}^{i} (d^{(a)}_{q,\tau_j} + d^{(g)}_{q,\tau_j}). \tag{C22}$$

As the predicted number of flights in a sector, as well as the number of flights entering a sector, are actually random numbers due to random errors in predicted times of crossing sector borders by individual flights, the random total traffic demand $\tilde{D}_i$ in a sector predicted for one-minute interval i can be represented by a formula similar to (C3) where the deterministic values are replaced by corresponding random ones:

$$\tilde{D}_i = \sum_{j=1}^{m} \tilde{D}_{i,\tau_j} = \sum_{j=1}^{m} \sum_{q=i-\tau_j+1}^{i} \tilde{d}_{q,\tau_j} = \sum_{j=1}^{m} \sum_{q=i-\tau_j+1}^{i} (\tilde{d}^{(a)}_{q,\tau_j} + \tilde{d}^{(g)}_{q,\tau_j}). \tag{C23}$$

After applying the same technique used in Case 1 to each fraction of one-minute sector demand predictions $\tilde{D}_{i,\tau_j}$ with specific time in sector τ_j and broken down by active and proposed flights components, the following expressions have been obtained for the mean and standard deviation of one-minute demand counts in a sector, respectively:

$$\bar{\tilde{D}}_i = \sum_{j=1}^{m} \Big(\sum_{k=i-\tau_j-\beta_1+1}^{i+\beta_1} P^{(a)}_{i,k,\tau_j} \, d^{(a)}_{k,\tau_j} + \sum_{k=i-\tau_j-\beta_2+1}^{i+\beta_2} P^{(g)}_{i,k,\tau_j} \, d^{(g)}_{k,\tau_j} \Big), \tag{C24}$$

$$\sigma(\tilde{D}_i) = \sqrt{\sum_{j=1}^{m} \Big[\sum_{k=i-\tau_j-\beta_1+1}^{i+\beta_1} P^{(a)}_{i,k,\tau_j} (1-P^{(a)}_{i,k,\tau_j}) d^{(a)}_{k,\tau_j} + \sum_{k=i-\tau_j-\beta_2+1}^{i+\beta_2} P^{(g)}_{i,k,\tau_j} (1-P^{(g)}_{i,k,\tau_j}) d^{(g)}_{k,\tau_j} \Big]}. \tag{C25}$$

Equations (C24) and (C25) provide the values that are necessary for probabilistic prediction of one-minute traffic demand in a sector that take into account characteristics of accuracy of predictions of flights' sector entry times depending on flight status and time in sector. The predicted number of flights in a sector at any one-minute interval can be represented by an average number $\bar{\tilde{D}}_i$ (see formula (C24)) and the area of uncertainty around the average, say, within $\pm \gamma \, \sigma(\tilde{D}_{i,})$, where the standard deviation $\sigma(\tilde{D}_{i,})$ is determined by formula (C25). The coefficient γ can be selected to determine the percentiles that limit the upper and lower limits of uncertainty area. For $\gamma = 1$, the area of uncertainty around the average number of flights (plus/minus one sigma) would be restricted by the 84th and the 16th percentiles.

C.3 Summary

How the analytical results presented in the Appendices should be used for probabilistic predictions of one-minute traffic for en route sectors

Appendices A, B and C gave a full theoretical background and analytical results for probabilistic predictions of one-minute demand counts for both sector entries and number of flights in a sector.

The whole approach for probabilistic prediction of one-minute sector demand counts is based on probability distribution functions (pdf) of errors in predicting sector entry times of individual flights for determining probabilities for flights with various estimated sector entry times (ETA) to enter a sector during a one-minute interval of interest. The analytical formulas for probabilistic predictions of sector demand counts manipulate with those probabilities.

The following is an overview on how to use the results described above for probabilistic predictions of one-minute sector demand at a specific sector and for each one-minute interval of interest.

1. For each look-ahead time (LAT), the following prediction data on individual flights is required (the data is obtained beforehand through statistical data analysis of historical data):

   - probability density functions (pdf) and corresponding cumulative distribution functions (CDF) of prediction errors for flights' sector entry times (for both active and proposed flights)
   - the values of parameters β_1 and β_2 for active and proposed flights, respectively, that are determined by the corresponding pdf of prediction errors (the rules for selecting those

parameters are described in Appendix B). The parameters determine the number of one-minute intervals preceding and following the interval of interest where deterministic predictions of sector entry counts should be taken into account for probabilistic predictions of sector entry counts for the interval of interest

- for all $|i - k| \leq \beta_1$ determine the probabilities $P_{i,k}^{(a)}$ for an active flight to enter a sector at interval *i* if it deterministically predicted to arrive at interval *k* and determine similar probabilities for proposed flights $P_{i,k}^{(g)}$ for all $|i - k| \leq \beta_2$
- for a specific sector, determine a set of typical times in sector τ, and find the maximum time τ_{max} needed for traversing the sector from all groups of flights

2. For a specific LAT
   - determine a set of consecutive deterministic predictions d_k of one-minute sector entry counts that are needed for probabilistic demand prediction for a one-minute interval of interest i. This set includes sector entry counts from intervals k that are from the range of $i - \tau_{max} - \beta + 1 \leq k \leq i + \beta$, where $\beta = \max(\beta_1, \beta_2)$. Altogether, $2\beta + \tau_{max}$ deterministic predictions d_k are needed for probabilistic prediction of sector demand counts for a one-minute interval of interest.
   - obtain deterministic predictions of one-minute sector entry counts d_k from TFMS for each consecutive one-minute interval *k* relevant to the time window of interest, and determine the active $d_k^{(a)}$ and proposed $d_k^{(g)}$ components in $d_k = d_k^{(a)} + d_k^{(g)}$.
   - at each one-minute interval *k*, determine
     - groups of flights with different times in sector τ_j (j = 1, 2, …, m)
     - the number of active and proposed flights in each group: $d_{k,\tau_j}^{(a)}$ and $d_{k,\tau_j}^{(g)}$
   - for all intervals *k* ($i - \tau_{max} - \beta + 1 \leq k \leq i + \beta$), using formula (C7), calculate the probabilities $P_{i,k,\tau_j}^{(a)}$ and $P_{i,k,\tau_j}^{(g)}$ for active and proposed flights, respectively, to be in a sector at a one-minute interval of interest if they deterministically predicted to enter the sector during interval *k*.
3. For probabilistic predictions of one-minute traffic demand in a sector for each minute of interest, use formulas (C24) and (C25) to determine the mean and standard deviation of number of flights in a sector. These values characterize the expected number of flights in a sector and the area of uncertainty around the expected number.
4. For probabilistic predictions of one-minute sector entry counts, use formulas (B26) and (B29) that determine the mean and standard deviation of number of flights entering a sector during any one-minute interval of interest. These values characterize the expected number of flights entering a sector and area of uncertainty around the expected number.

Appendix D. References

[1] Smith, Scott and Eugene Gilbo, "Analysis of Uncertainty in ETMS Aggregate Demand Predictions", Volpe National Transportation Systems Center, Report no. VNTSC-ATMS-05-05, November 2005.

[2] Gilbo, Eugene and Scott Smith, "Reducing Uncertainty in ETMS Aggregate Traffic Demand Predictions," Volpe National Transportation Systems Center, Report no. VNTSC-CE-07-01, March 2007.

[3] Gilbo, Eugene and Scott Smith "A New Model to Improve Aggregate Air Traffic Demand Predictions," Paper number AIAA 2007-6450, AIAA Guidance, Navigation, and Control Conference and Exhibit, Hilton Head, SC, 20 -23 August 2007.

[4] Gilbo, Eugene and Scott Smith, "Characterization of Uncertainty in ETMS Flight Events Predictions and its Effect on Traffic Demand Predictions," Volpe National Transportation Systems Center, Report no. VNTSC-TFM-08-08, July 2008.

[5] Gilbo, Eugene and Scott Smith "Probabilistic Prediction of Aggregate Traffic Demand Using Uncertainty in Individual Flight Predictions," Paper number AIAA 2009-6494, AIAA Guidance, Navigation, and Control Conference and Exhibit, Chicago, IL, 10 – 13 August, 2009

[6] Meyn, Larry A., "Probabilistic Methods for Air Traffic Demand Forecasting," Paper number 2002-4276, AIAA Guidance, Navigation, and Control Conference and Exhibit, Monterey, CA, 5 - 8 August, 2002.

[7] K. Tysen Mueller, John A. Sorensen and George J. Couluris, "Strategic Aircraft Trajectory Prediction Uncertainty and Statistical Sector Traffic Load Modeling," Paper number AIAA 2002-4765, AIAA Guidance, Navigation, and Control Conference and Exhibit, Monterey, CA, 5 - 8 August, 2002.

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

[9] Wanke, Craig R., Michael B. Callaham, Daniel P. Greenbaum, Anthony J. Masalonis, "Measuring Uncertainty in Airspace Demand Predictions for Traffic Flow Management Applications," AIAA Guidance, Navigation, and Control Conference and Exhibit, Paper # AIAA-2003-5708, Austin, TX, 11 – 14 August, 2003.

[10] Wanke, C., Mulgund, S., Greenbaum, D. and Song, L., "Modeling Traffic Prediction Uncertainty for Traffic Flow Management Decision Support,", AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, RI, 16 – 19 August, 2004.

[11] Wanke, C., Song, L., Zobell, S., Greenbaum, D. and Mulgund, S., "Probabilistic Congestion Management," 6th USA/Europe Air Traffic Management R&D Seminar, Baltimore, MD, June 27-30, 2005.

[12] Zobell, Stephen, Craig Wanke and Lixia Song, "Probabilistic Airspace Congestion Management," paper number 2005-7433, AIAA 5th Aviation, Technology, Integration, and Operations Conference (ATIO), Arlington, VA, 2005

[13] Hunter, George, and Kris Ramamoorthy, "A Trajectory-Based Probabilistic TFM Evaluation Tool and Experiment", ICNS Conference, Baltimore, MD, May 2006

[14] Ramamoorthy, K.; Boisvert, B.; Hunter, G., “A Real-Time Probabilistic Traffic Flow Management Evaluation Tool”, 25th Digital Avionics Systems Conference, 2006 IEEE/AIAA, pp. 1 – 13, Portland, OR, 15-19 Oct. 2006.

[15] George Hunter, Ben Boisvert, and Kris Ramamoorthy, “Advanced national Airspace Traffic Flow Management Simulation Experiments and Validation,” Proceedings of the 2007 Winter Simulation Conference, pp. 1261-1267

[16] Wanke, C., and Greenbaum, D., “Incremental, Probabilistic Decision-Making for En Route Traffic Management”, 7th USA/Europe ATM R&D Seminar, Barcelona, Spain, July 2007

[17] J. DeArmon, C. Wanke, D. Greenbaum, “Probabilistic TFM: Preliminary Benefits Analysis of an Incremental Solution Approach”, MITRE, 2007.

[18] Craig Wanke, “Continual, Probabilistic Airspace Congestion Management”, AIAA Guidance, Navigation, and Control Conference and Exhibit, Chicago, IL 10 – 13 August, 2009.

[19] R. C. Pfaffenberger and J. H. Patterson, “Statistical Methods,” Third Edition, IRWIN, 1987