US Department
of Transportation
**Research and
Special Programs
Administration**

Federal Aviation
Administration
**Free Flight Phase 1
Program Office**

DOT-VNTSC-FAA-99-12

Relative Accuracy of CDM and ETMS in Predicting Airport Arrival Demand

by

Eugene Gilbo
Lawrence McCabe

September 1999

Prepared for
Free Flight Phase 1 Program Office
Suite 360
1500 K Street, N.W.
Washington, DC 20005

Prepared by
Volpe National Transportation Systems Center
Automation Applications Division, DTS-56
55 Broadway
Cambridge, MA 02142-1093

Table of Contents

1 Introduction and Overview of Results

Safety and efficiency in national airspace usage are the fundamental goals of the Federal Aviation Administration (FAA). In order to achieve these objectives, a number of complex systems have been developed. One of these systems, the Enhanced Traffic Management System (ETMS) is used to predict and manage the efficient flow of aircraft through controlled airspace. Of particular interest to this study, ETMS predicts the flow of aircraft to airports many hours into the future. These predictions are based upon data from several sources including:

- flight schedule data received from the Official Airline Guide (OAG),
- flight plans filed by the airlines,
- radar data from FAA operational facilities such as Terminal Radar Approach Control (TRACON) installations and Air Route Traffic Control Centers (ARTCCs),
- flight substitution messages from the airlines, and
- data from other governmental sources.

Under ETMS, flight plan data received from the airlines are typically provided less than 2 hours prior to departure. Until flight plan data is received, ETMS relies on OAG schedule data that is updated weekly. This leaves a significant period of time during which the airlines may have updated flight information not available to the FAA. The Collaborative Decision Making (CDM) program was formally initiated in 1995 in order to improve air traffic flow management through increased sharing of information and decision making among the FAA and the airlines. Under CDM, as the airlines change their flight schedules, they send flight creation (FC), modification (FM), and cancellation (FX) messages to the CDM system, starting 15 hours prior to departure. These messages provide the CDM system with updates to the OAG data prior to the issuance of flight plans.

ETMS, as used in this report, refers to the operational ETMS. CDM refers to the version of ETMS that has been modified to incorporate, among other things, the data from the CDM messages into its databases.

CDM and ETMS use identical algorithms for predicting flight profiles; therefore differences in the accuracy with which each system can predict arrival demand at airports are due to the CDM messages sent in by the airlines that are incorporated into CDM but not ETMS. The data used in the analysis reported here were extracted simultaneously from CDM and ETMS to insure comparability. Data collection time periods were selected to coincide with Ground Delay Programs because it is during these intervals that accurate predictions are most critical.

This report documents the results of an analysis comparing the airport arrival demand prediction errors in the ETMS and CDM traffic management systems. The metrics used here, namely the average absolute errors, are suitable for quantifying the improvement provided by CDM over ETMS.

The overall findings indicate that the most significant improvement in the prediction of airport arrival demand provided by CDM occurred for the four hours following the completion of a GDP. For these times, CDM estimates were up to 60% better than those provided by ETMS. CDM provided a 10% to 20% improvement over ETMS for the time period covered by the GDP. The most noticeable improvements in prediction occurred when the predictions were made 3 to 8 hours in advance. The magnitude of these improvements depends upon the percentage of FX cancelled flights in the total traffic demand.

2 Objective

This study has two primary objectives:

1. Quantify the difference in the accuracy of airport arrival demand predictions performed by ETMS and CDM.
2. Identify and explain the differences between ETMS and CDM predictions.

3 Methodology

When traffic demand is forecast to exceed capacity for an extended period of time, a Ground Delay Program (GDP) is initiated. The GDP frequently occurs because weather conditions are expected to deteriorate and to produce reduced airport capacity. Because the time of most critical need for prediction accuracy occurs when capacity is lowered and demand-to-capacity ratios are high, it is appropriate to analyze the data collected when GDPs are in effect.

The GDPs from February to June 1999 were used as a basis for comparing the accuracy of ETMS and CDM in predicting airport arrival demand.

3.1 ETMS and CDM Data

Both ETMS and CDM have Traffic Databases (TDBs) which store predicted demand for monitored elements (airports, sectors, fixes) of the National Airspace System. These databases provided the data for this analysis, namely the predicted arrival demand at selected airports for each fifteen-minute interval during the 24 hours beginning at the start of each GDP of interest. Each TDB also maintains an historical record of the actual number of arrivals and the specified capacities during these intervals. Traffic management specialists input these capacities to the system to reflect the changes in arrival traffic handling capability at an airport when a GDP is in effect. The capacity values were used in the analysis of conditions affecting the magnitude of improvement of arrival traffic prediction due to CDM.

The actual demand from historic data was subtracted from the predicted value to determine the prediction error for each fifteen-minute interval. Appendix A contains a detailed description of these TDBs and how they were accessed to provide the required data.

3.2 External Data

Two data sources external to the ETMS and CDM TDBs were also used. There are the Metron-prepared GDP summaries and the airline cancellation (FX) messages available at the Volpe Center hubsite. The GDP summaries provide a clear chronology of GDP events. Beginning and end times from these summaries were used in this analysis to establish the GDP and post GDP periods for examination. The Airport Acceptance Rate (AAR) values contained in these summaries were also used to check on the validity of the capacities read from the TDB data.

The counts of airline cancellation (FX) messages were used to estimate when CDM data were most likely to have had an impact on arrival demand predictions. The analysis used the FX count for an airport on the specified date when the GDP occurred for correlation with observed improvements.

3.3 Data Analysis Time Frames

Figure 1 depicts the relevant time frames used in the analysis. Three analysis periods were selected for each GDP: (1) the 24 hours starting when the GDP started; (2) during the GDP; and (3) the four hours following the GDP.

The ETMS and CDM systems were used to predict the number of arrivals at a selected airport during each fifteen-minute period in the twenty-four hours beginning at the start of a GDP. A prediction was made every quarter hour starting ten hours before the event interval. Thus there are 40 predictions of each event. The term event is used in this report to mean the arrival demand for a specified 15-minute period. In a twenty-four-hour period there are 96 event intervals.

The 40 predictions for a given event were numbered 1 through 40, with 1 corresponding to the prediction made one quarter hour before the end of the event interval. These numbers are referred to as the "look-ahead" index numbers, and represent the number of quarter hours between the prediction time and the time of completion of the event interval. Thus, look-ahead index number 4 corresponds to the prediction made 4 quarter-hours (i.e., one hour) before the end of the event interval and look-ahead index number 40 refers to the prediction made 40 quarter hours (ten hours) before the end of the event interval.

If the event is the cumulative count of arrivals at an airport between 1200:01 and 1215, the prediction made at 1200 is at look-ahead index number1, the prediction made at 1145 is at look-ahead index number 2, and the prediction made at 0215 is at look-ahead index number 40. The actual number of arrivals is subtracted from each prediction to produce the error for that look-ahead index. It is these errors that were analyzed in this study. Appendix A contains a discussion of data reduction.

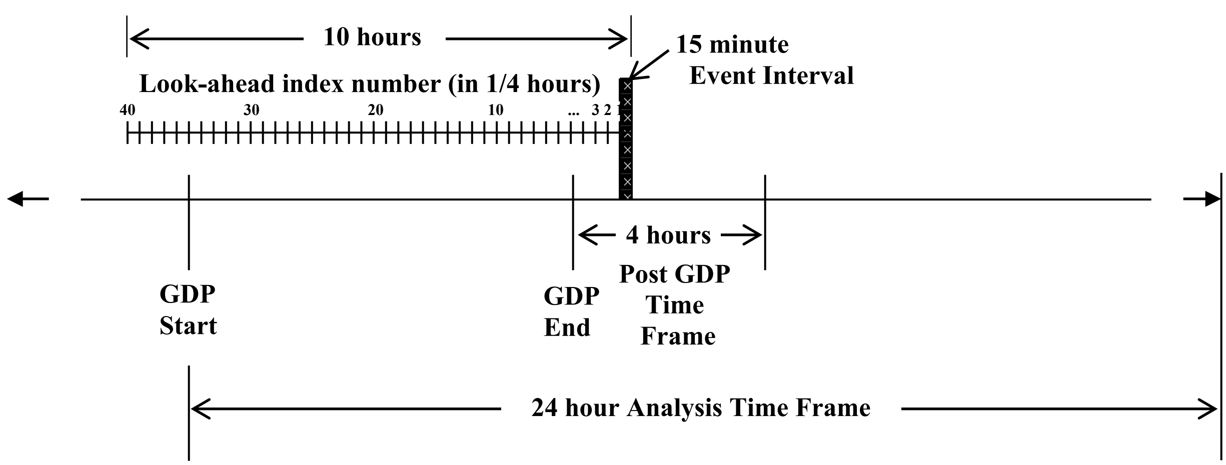

FIGURE 1. Analysis and Look-ahead Times

3.4 Prediction Error Analysis

The following metrics were used in this analysis:

- Average absolute error as a function of look-ahead index
- Average Normalized Integrated Predictive Error.

3.4.1 Average Absolute Error

As mentioned above, three analysis periods were selected for each GDP: (1) the 24 hours starting when the GDP started; (2) during the GDP; and (3) the four hours following the GDP. For each of these time periods, averages of absolute errors were calculated for each quarter hour of look-ahead time prior to the event. Thus, for example, if the first event was the cumulative total of flights arriving between 1300:01Z and 1315Z, the absolute prediction error resulting from the prediction made three hours before the event, at 1015Z, would be averaged with the absolute prediction error occurring at 1030Z for the next event. This next event is the cumulative total for flights arriving between 1315:01Z and 1330Z. Both absolute prediction errors correspond to the twelfth look-ahead index number for their respective events.

Some of the predictions were performed when forecast conditions did not predict any capacity reduction or any need to reduce demand as subsequently materialized. Because predictions made for the start of each analysis period includes those that were performed far in advance of the conditions, the potential advantage of CDM may be lessened because neither string is operating under the supposition that anything other than normal conditions will prevail.

3.4.2 Average Normalized Integrated Predictive Error

The second metric used in this analysis is an adaptation of the Average Integrated Predictive Error (IPE).[1] IPE is a metric designed to express the cumulative error of a stream of predictions made over time for a single event. The IPE is applied to all predictions made for a single event. The IPE is duration and magnitude based. That is, if predictions are made at discrete points in time, the absolute error at each prediction time is considered to remain constant until the next prediction occurs. A new prediction error with a different duration then ensues. For each of these error levels, the product of their respective magnitude and duration is calculated. The sum of these products is the IPE.

When comparing IPEs resulting from two different time bases, it is useful to divide each IPE by the respective time base to obtain a normalized IPE. For the ETMS and CDM data, the time base used for each IPE was forty quarters. The normalized IPE is used in this study. For each of the time frames (24 hour, during the GDP, and post GDP), the average normalized IPE was calculated. Similar IPE calculations were performed for time periods of high and low capacity as well as high and low demand-to-capacity ratios. Percent improvements were obtained by comparing the average normalized IPEs for ETMS and CDM under the same conditions.

3.4.3 Analysis Charts

A series of charts, using the metrics and time frames described above, were generated for each GDP in order to identify and evaluate any observable trends. Samples of these charts are presented and discussed in Appendix B. The charts include plots of average absolute errors for CDM and ETMS and percentage improvements due to CDM as a function of look-ahead time. Only the charts for the 24 hour period beginning at GDP inception are included, since they are representative of the trends displayed by the other time periods analyzed (GDP and post GDP). The summary metrics for each of the time periods analyzed are given in Tables 1 and A.1. Similar charts were generated for the average normalized IPE for CDM and ETMS and percent improvement in average normalized IPE due to CDM. These charts, however, are not included because they provide essentially the same results as those for the average absolute errors.

[1] For a formal definition of IPE, see Ball, M, et al, *Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment*, NEXTOR Report Number RR-99-3, August 1998.

4 Discussion of Results

4.1 Individual GDP Analysis Charts

The charts generally indicate that CDM produced smaller absolute errors than ETMS produced. Many of the comparisons show that the magnitude of the arrival demand prediction errors provided by CDM and ETMS are comparable when look-ahead times are as large as 8 to 10 hours. CDM tends to provide decreasing errors as the look-ahead index numbers decrease. The most noticeable advantages of CDM occur within 3 to 8 hours of prediction look-ahead time. All of the charts display a reduction in error as the look-ahead value is reduced from forty to one (10 hours to 15 minutes). The average absolute error for CDM, based on all of the presented charts, is 3.4 flights per 15 minutes when the prediction is made at 40 quarter-hours (10 hours) of look-ahead. This error diminishes to 1.7 flights per 15 minutes when the prediction is made for look-ahead interval one (i.e., within the 15-minute interval starting at the current time).

4.2 Summary Tables and Charts

Table 1 provides a summary of improvements in predicted airport arrival demand error due to CDM. The results for eight GDP programs occurring in 1999 are shown and identified by location and date. The Actual 24-hour Arrival Demand in the second column is the sum of actual counts, from ETMS data, of flights arriving during the 24-hour period starting at each GDP inception. The FX Count is the total number of flight cancellations resulting from FX messages received through CDM for the specified airport during the 24-hour period starting at GDP inception. The FX Percent is the ratio, expressed as a percentage, of the FX count to the virtual demand. Virtual demand is the sum of the actual arrival demand plus FX cancellations. The remaining columns provide the percentages of improvement over the ETMS system realized by the operation of CDM. All columns except those under the During GDP and Post GDP headings contain data summarized from the 24-hour operation starting at GDP initialization.

The Average Percentage Improvement was calculated from the percentage reduction in arrival demand prediction error due to CDM. At each look-ahead value, the errors were averaged over the 24-hour period for each system. Charts in Appendix B show the size of the average absolute errors for ETMS and CDM at each look-ahead value. The percentage reduction in the average arrival demand prediction error due to CDM at each look-ahead was then calculated and is also displayed in Appendix B. The average of these percentage reductions over the 40 look-ahead values is shown in Table 1 in the column labeled Average Percentage Improvement.

The next column shows the corresponding improvement in IPE for each GDP. Improvements in both metrics were nearly identical and ranged from 6 to 31 percent. Both of these results are shown plotted with the respective FX percent for the GDP in Figure 2. The averaging obtained from the look-ahead improvements shows a slightly better fit to the data than does the averaging obtained from the normalized IPE. For this reason, and since the results for the two metrics were almost identical, the remainder of the columns in Table 1 present only the results based on the Average Absolute Error metric. Table B.1 (in Appendix B) presents the summary results for the Averaged Normalized IPE metric as well.

Both metrics show a positive trend to the relationship between the improvement percentages and the FX percentage. When the FX percentage exceeds 15, significant improvements in CDM percentages are likely to occur. Additional data points in Figure 2, taken from Table 2 for GDPs not used in this analysis because of low improvement numbers, show a lower and trendless percentage improvement for FX percentages below 15.

TABLE 1. Summary of Improvements Due to CDM

Location & Date of GDP	Actual 24 Hour Arrival Demand[2]	FX Count[3]	FX Percent[4]	Average % Improvement (24 hr)	% Improvement of IPE (24 hr)	% Predicted Error Improvement for Capacity Level			% Pred. Error Improvement for Dem/Cap Level		% Pred. Error Improvement	
						High	Medium	Low	High	Low	During GDP	Post GDP
BOS 2/25	301	273	48	30	31	41		25	21	32	18	55
BOS 3/7	474	122	20	15	14	18		10	11	15	6	27
BOS 3/15	504	161	24	16	17	22		8	9	20	12	9
LGA 3/4	432	117	21	22	23	*	*	*	10	27	6	59
ORD 3/6	889	249	22	15	16	20	14	10	12	19	10	21
SFO 3/15	568	37	6	6	7	6	4	2	6	6	9	8
SFO 3/16	539	62	10	9	9	4		10	5	12	6	16
STL 3/8	585	133	19	15	16	24		8	8	23	7	48

TABLE 2. List of Additional GDP Occurrences

Airport	Date	Actual 24 Hour Demand[2]	# of FXs in 24-hr period	FX Percent[4]	% IPE Improvement
BOS	3/8/99	603	78	11	5
	3/12/99	615	72	10	3
EWR	3/14/99	528	54	9	-2
	3/15/99	607	85	12	3
PHL	3/4/99	673	44	6	5
	3/14/99	506	89	15	4
SFO	3/14/99	540	42	7	5
	3/24/99	551	67	11	4

[2] Actual arrivals during 24 hour period starting at GDP inception.
[3] Total number of FX cancellations in 24-hour period.
[4] Number of FX Cancellations / (Actual 24-hour demand + number of FX Cancellations)
[*] No capacity variation during 24-hour period.

The next set of columns in Table 1 contains the Percentage of Predicted Error Improvement for periods of high, intermediate and low capacity levels. These percentages and those of the remaining columns were calculated from the average percentage reductions in error over the forty look-ahead intervals.

Figure 3 shows the corresponding data points and trend lines for improvements under high and low capacity levels. Capacity levels generally indicate the deterioration or improvement of weather conditions at the airport. These levels may vary during the GDP.

The improvements provided by CDM compared to ETMS are greater for periods of high capacity than for low capacity. The reduced accuracy of predictions for periods of low capacity occurs partially because of the timing of predictions. The characteristic pattern of a GDP shows a high airport capacity at the beginning of the GDP that is quickly reduced to a low capacity. Some of the predictions previously made for these low capacity intervals were produced when the airport capacity was high and relatively few FX cancellations were needed. Thus, the improvements of CDM over ETMS are small and the errors from both systems are large because the demand cannot be met. (Refer to Table A.2 for a summary of the magnitude of the prediction errors from both the ETMS and CDM systems.)

Following the GDP, the high arrival capacity is restored. The predictions produced for some of this later time period were performed when the capacity was lowered and FX cancellations were in effect. The cancellations for this later time period showed their effect by limiting the amount of traffic for this later period and producing small error sizes. Thus, the improvements provided by CDM for periods of high capacity are substantially larger than are those for periods of low capacity.

A related, but not identical, improvement grouping is shown in the next column. These results show the percentage improvements for high and low demand-to-capacity ratios. The table values are plotted in Figure 4. Although the results are compatible with those in the previous figure, there are some differences. The high and low demand-to-capacity ratio trend lines of Figure 4 are consistent with the respective low and high capacity trend lines of Figure 3. The corresponding slopes differ, however, due to the effect of demand. Some of the low demand-to-capacity ratios occur during intervals of high capacity and some during low capacity intervals. Similarly, some of the high demand-to-capacity occurrences are at times of both capacities. This mixing tends to modify the percentage improvements for this group of data.

The final two columns of Table 1 show the percentage improvements of CDM over ETMS for the GDP and post-GDP periods. Both sets of data are plotted in Figure 5. The values plotted in Figure 2 for the 24-hour period are added to this chart to provide perspective to the results. The GDP period shows relatively weak improvement when compared with the post-GDP interval. Improvement values for the latter period are scattered over a wide range and attain a value as large as 59 percent.

CDM compression procedures performed during the GDP produce this post-GDP enhancement. As a result of compression, flights are moved into open slots created by FX cancellations. The cancelled flight slots are moved to a later time outside the GDP interval. The results of this compression are sent to ETMS and CDM as new controlled departure times. During the GDP, this process makes the two systems converge. After the GDP, however, FX cancellations will have removed those flights moved to post GDP times from CDM predictions. ETMS does not have this cancellation information and these flights will appear erroneously in ETMS post GDP predictions.

5 Conclusions

- CDM substantially improves the prediction of airport arrival traffic demand during extensive Ground Delay Programs when significant numbers of airline FX cancellation messages are received.
- The most significant improvements are observed during post GDP hours. CDM improved the accuracy of predictions in the range of 40% to 60% for these intervals.
- During the GDP prediction improvements were in the range of 10% to 20%.
- The most noticeable improvements in prediction occurred between 3 to 8 hours prior to the event.
- CDM exhibited larger improvements with higher capacity levels and under lower demand to capacity ratios.
- The percentage improvements in prediction were nearly identical for the prediction look-ahead metric as for the average IPE metric.
- FX messages are the predominant factor in producing CDM improvements. When there is little or no FX messaging to CDM the improvement is minimal.

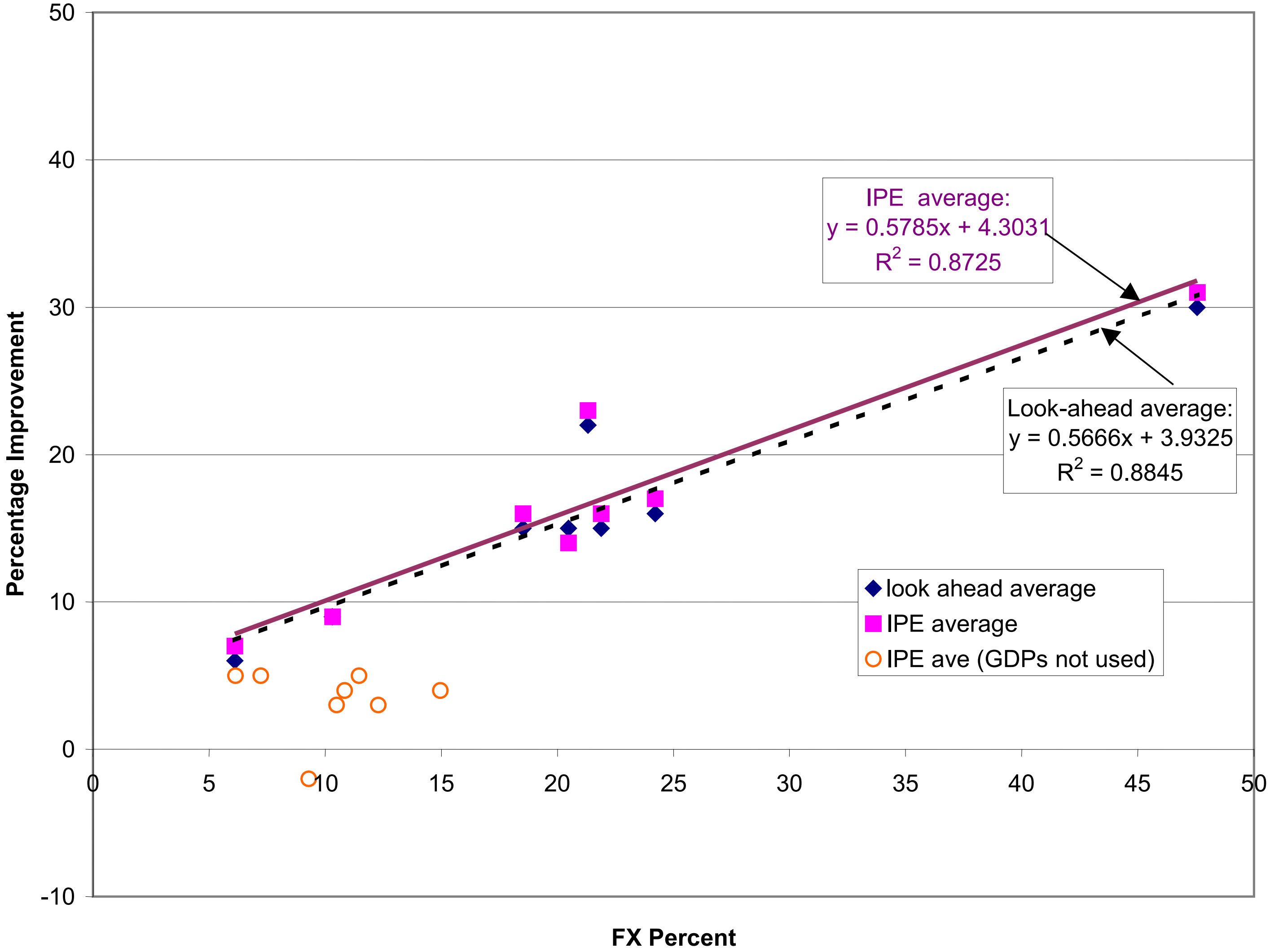

FIGURE 2. Percent Improvement of Prediction for 24-Hour Period for Several GDPs – Comparison of Look Ahead Averaging and IPE.

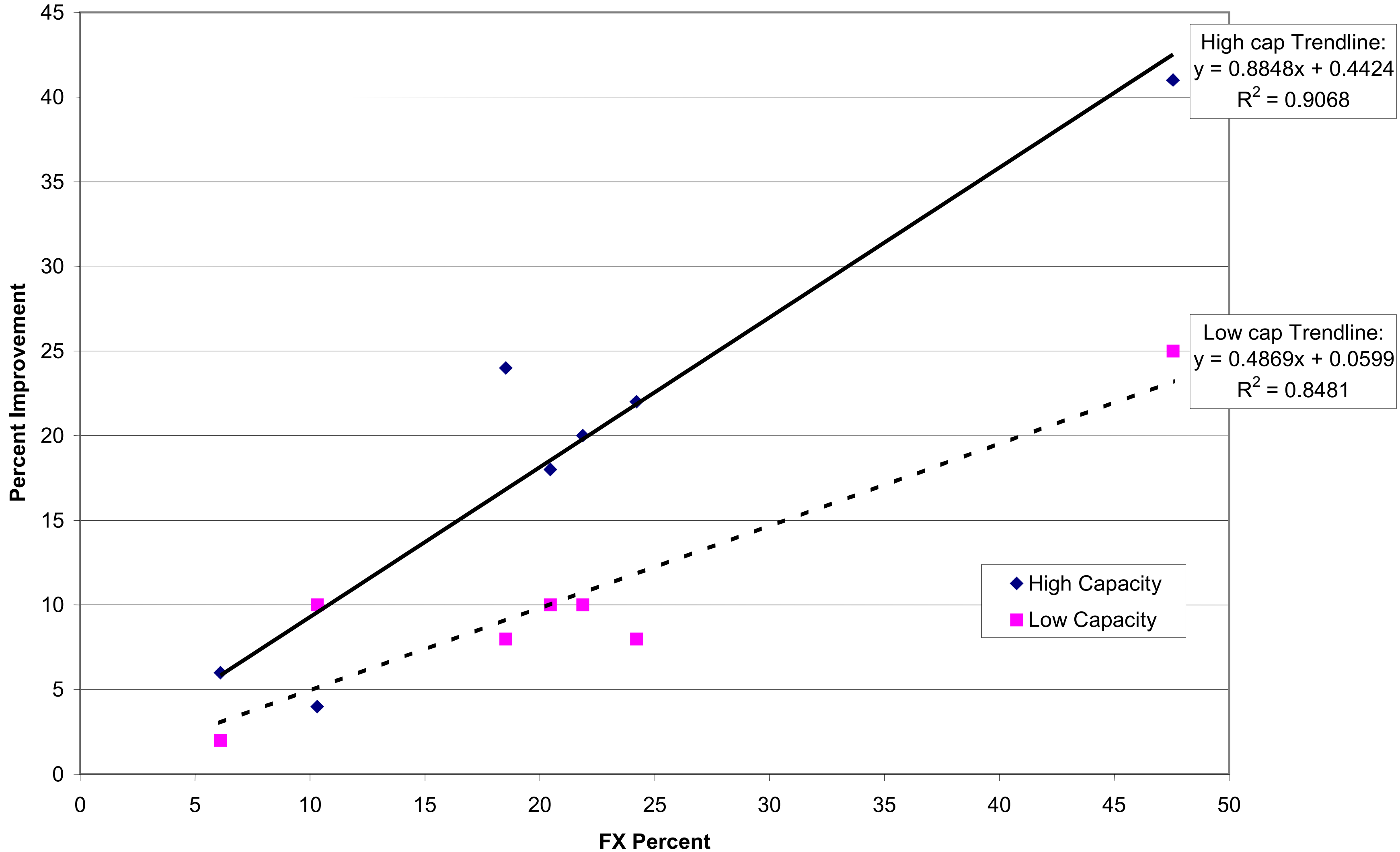

FIGURE 3. Prediction Error Percent Improvement as a Function of FX Percent (High and Low Capacity levels)

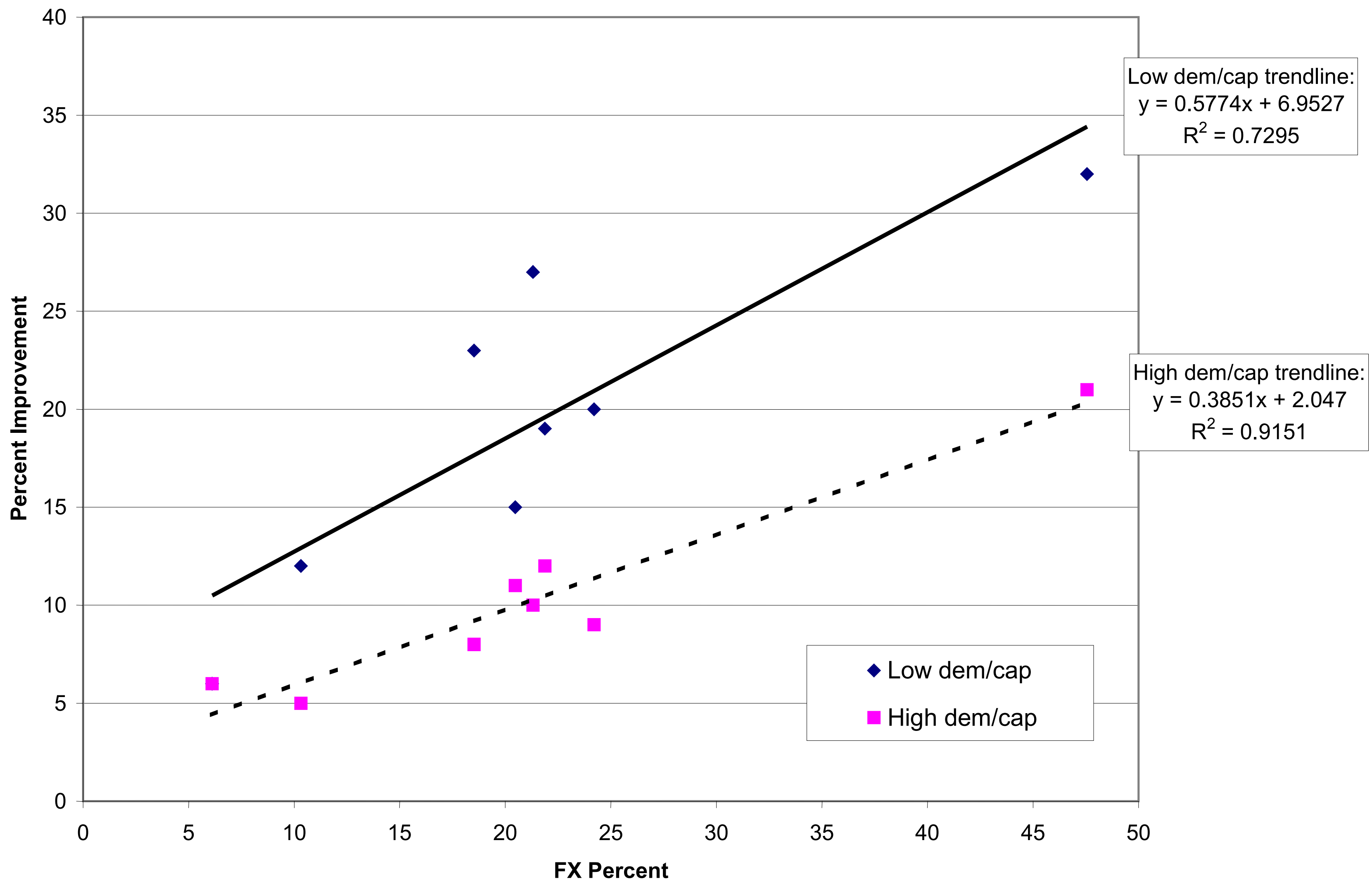

FIGURE 4. Prediction Error Percent Improvement as a Function of FX Percent (High and Low dem/cap levels)

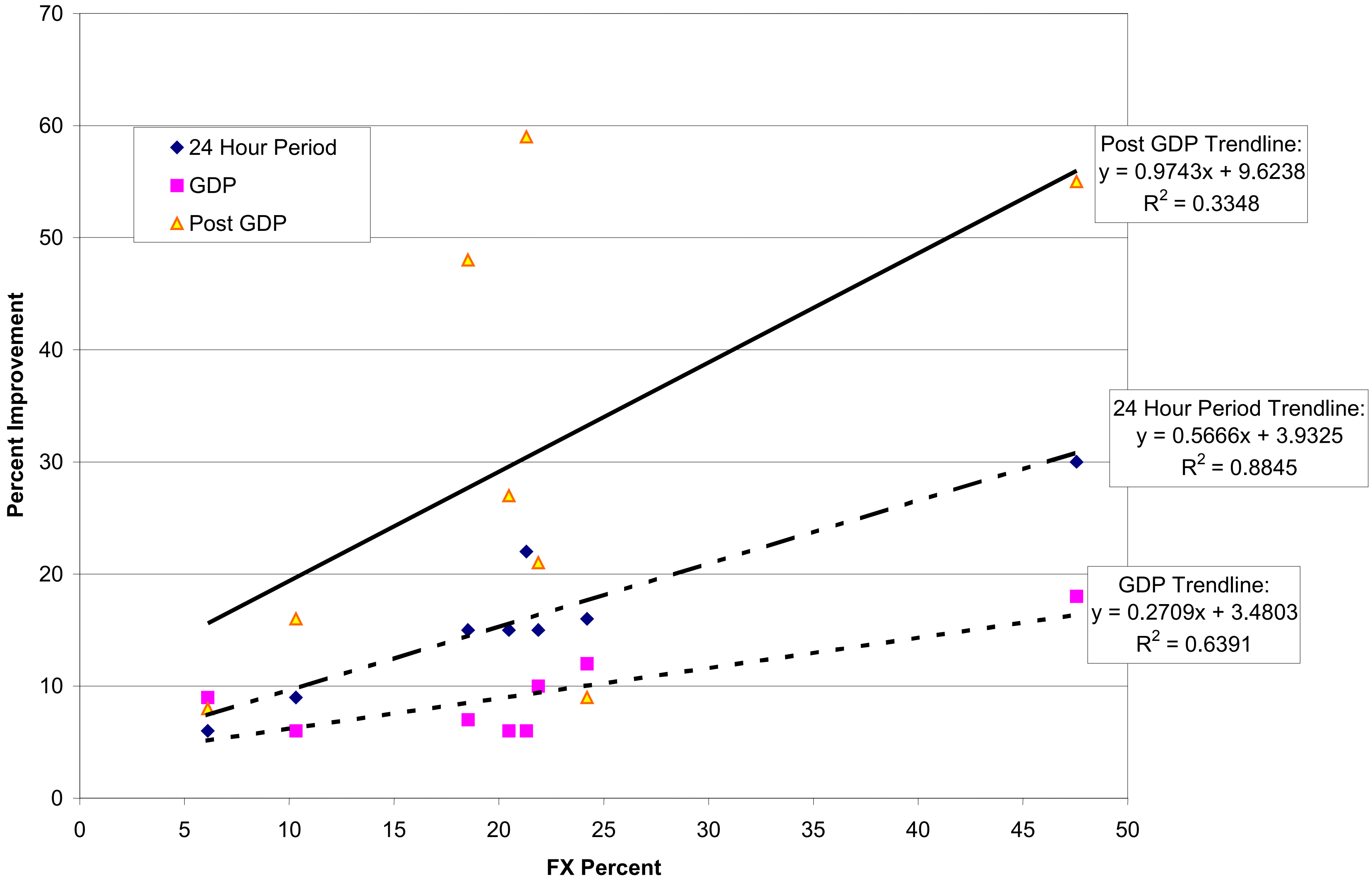

FIGURE 5. Prediction Error Percent Improvement as a Function of FX Percent (24 Hour, GDP, and Post GDP Intervals)

Appendix A. Technical Description

A.1 Traffic Data Base

Both ETMS and CDM systems store prediction data in their respective traffic database. Each system simulates scheduled, proposed and active flights operating between origin and destination airports. These simulations are based upon input data from several sources. ETMS and CDM both receive data from the Official Airline Guide, FAA host and ARTS computers, the Dynamic Oceanic Tracking System, from foreign air traffic control systems and from airlines. The CDM airline data contains significantly more content than does ETMS airline data because CDM receives early updates on planned flight operations and cancellation messages.

A series of flight events, such as departures, fix crossings, sector entries and exits, and arrivals at the destination airport are simulated by a flight data base process. Times of these events are recorded and updated upon reception of new flight data. Aggregates of flights accessing a fix, sector or airport during a specific interval form the traffic demand for that location and are stored in the Traffic Data Base [5,6] (TDB) of the respective ETMS or CDM system.

Each TDB contains demand data for monitored locations. Currently these monitored elements are: (1) all U.S. and Canadian airports with scheduled flights, all Mexican airports with international flights and several other individually specified airports; (2) all CONUS and Canadian sectors; and (3) all fixes specified by traffic management.

Data identifying each flight comprising the demand can be retained in the TDB for a 64-hour time window extending 24 hours back in the past from the current time and 40 hours into the future. Based upon OAG data, scheduled flights are simulated 12 hours before scheduled departure time. The earliest TDB data for a future time interval generally consists of these scheduled flights.

Demand data is stored in each TDB in two tables, a statistics table and a flight table. The statistics table contains the traffic demand counts for each fifteen-minute interval and the capacities applicable to the interval. The demand data is divided into four categories: active demand consisting of active and completed flight counts; scheduled demand consisting of scheduled flight counts; total demand consisting of active and scheduled flights plus filed flights; and unscheduled demand consisting of unscheduled active and filed flights.

Capacity values stored in each TDB consist of permanent default, or, nominal values and today values that are effective for time periods designated by FAA traffic managers. The today values can vary for each fifteen minute period and do reflect the capacity values forecast for the administration of a GDP.

The flights table contains pointers to flight lists. The information is stored by interval and monitored element. Each flight record referenced by the table contains the internal ETMS flight index, flight time at the element or sector entry and exit times, and a status flag indicating current activity. Flight records are held in blocks. These blocks are held as the linked lists with pointers in the flights table.

The TDB also maintains an alarms table that contains flags indicating overcapacity and alert color for every monitored element. When the demand exceeds the capacity in any interval, the

[5] *Enhanced Traffic Management System (ETMS) Functional Description*, Version 5.0, Volpe Center Report DTS-56-TMS-002, June 1995.
[6] *Enhanced Traffic Management System (ETMS) System Design Document,* Version 5.6, Volpe Center Report DTS-56-TMS-008, February 1997.

TDB detects this condition. For an excessive demand condition, the TDB enters sector and time data in the appropriate table and sets the color flag for the element. Monitor alert displays use this information generated within the TDB to provide graphic notification to the traffic management specialist when capacity is projected to exceed the today value.

A.2 Data Generation

The TDB counts of predicted demand at monitored locations and the capacities estimated to be in effect for each fifteen minute interval in the future are available through internal messaging within ETMS or CDM or through list requests and Capacities List (CAPL) commands in each system. Although it is theoretically possible to examine predictions as far ahead in the future as 40 hours, a practical limit of ten (10) hours provides most of the useful demand predictions for this analysis. The data for each system at and beyond this time limit is generally from the same OAG source.

A program obtaining data through messages accessed the TDB statistics tables every fifteen minutes and collected data for this analysis. At each collection time, the TDB provided arrival and departure counts for every fifteen minute look-ahead interval into the future during the next ten hours and also provided arrival and departure counts for every fifteen minute interval backward in time during the past two hours. From the future counts, the prediction data used in this analysis consisted of the total demand numbers. These are the sum of the active, scheduled and filed flight counts at each look-ahead interval. From the data representing past flights, the active demand counts were used as the actual numbers that had arrived or departed during those past intervals.

The program collects data for twelve airports. These are: BOS, EWR, LGA, ORD, ATL, DFW, STL, LAX, SFO, PIT, DTW and PHL. A file is produced each fifteen minutes containing ETMS A-string (VNTSCA) and a CDM string (CDMB) data for all twelve airports.

A.3 Data reduction

After the GDP date and time is selected from the summaries, the TDB data for that date is grouped in a directory consisting of the files previously generated for that date. The initial file contained in the directory was generated ten hours prior to the start of the GDP. This selection insures that the earliest demand prediction for the start time of the GDP period is contained in the data file. The TDB data files are collected in this directory in sequence such that every fifteen-minute period from the initial file time to 36 hours later is present.

The data extends for 36 hours because a 24-hour period starting at the time of the GDP was chosen for analysis. The first ten hours are needed to get the first prediction at the start of the GDP. The next 24 hours provides the demand prediction estimates for the GDP day and the last two hours provide the actual demand numbers applicable to the final prediction. Thus, a total of 144 TDB sequential data files are used for each GDP analysis.

Data is reduced so that the prediction values for each event time are collected as a stream of predictions starting with a look-ahead time of ten hours before the event up to the last prediction made for the event. For this analysis, arrival demand at a selected airport every fifteen minutes is the event of interest. All predictions, including the last, are for the cumulative total of flights predicted to arrive within fifteen minutes from the start of the interval. Forty predictions performed fifteen minutes apart are made for each event. The actual demand value for the event is taken from data collected two hours after the event. This delay insures that late messages coming to ETMS or CDM will have arrived and the arrival demand count will not change.

The prediction error is the predicted value minus the actual value. Actual values are taken from the ETMS data. The actual demand value for an event is subtracted from all prediction values belonging to the stream of predictions for that event time. A matrix of errors is generated with each row containing the prediction errors resulting from the stream of predictions generated for one event time. Each column contains the prediction errors occurring at the same look-ahead length from each event. For example, if the GDP start time is 1300Z, the first row would contain prediction errors made each fifteen minutes starting at 0315Z for the cumulative number of flights expected to arrive at a selected airport between 1300:01Z and 1315Z. The earliest prediction error would be contained in the fortieth element of the row and the last prediction error, made at 1300Z, would be in the first element.

When data is missing, predicted or actual data is filled in by using the previous set of estimates. Because the prediction for the fortieth quarter hour look-ahead of the missing data is not available from the previous set of estimates, this missing prediction for the tenth hour look-ahead is filled from the set of predictions made during the quarter hour following the missing set. Because of the two-hour delay in determining actual values, several opportunities are available to locate a useable actual value. The delay may be reduced to as little as one half hour, if necessary. No attempt is made to fill in data when two or more sequential predictions are missing. This results in a prediction of zero flights for the missing estimates and, consequently, larger errors for the affected string. The data reduction program produces a record displaying missing data. When too many predictions in sequence are missing for a given airport, the data is not analyzed.

The output of the data reduction program is an error matrix containing the time predicted for, the capacity at this time, the actual value and the errors for the forty predictions made over the previous ten hours. A spreadsheet produces the required averages and resulting graphs and summaries from this data. Selected graphs based on this analysis are presented in Appendix B

The arrival demand prediction errors are positively or negatively signed. When ETMS and CDM data is compared, signed error averages taken at each look-ahead interval could fortuitously indicate small values when some of the underlying errors are large. A more complete analysis requires standard deviation calculations as well as the average. Instead of signed averages and standard deviations, the use of absolute error magnitudes is appropriate when comparing differences in averages between two strings. The absolute error incorporates some of the features of the mean and of the standard deviation of signed errors. Generally, the standard deviation and absolute value mean tend to indicate nonlinear decreases as the number of look-ahead intervals are decreased. Furthermore, the absolute error averages do not mask large excursions in values and provide a direct means of comparison between ETMS and CDM. Thus, this analysis used absolute error values as a basis for comparison.

A.4 Data Summary

Table A.1 repeats the information contained in Table 1 and presents an additional set of columns with the percent improvement based on IPE values. The results based upon improvements for each look-ahead value are discussed in the body of the report.

Table A.2 contains a summary of Average Normalized IPE results for the 24-hour, GDP and post-GDP periods. IPE units are flights per 15-minutes. The percentage IPE improvements shown in Table A.1 are calculated from the ETMS and CDM values in this table (prior to their being rounded off).

TABLE A.1. Summary of Improvements Due to CDM

Location & Date of GDP	Actual 24 Hour Arrival Demand[7]	FX Count[8]	FX Percent[9]	Average % Improvement (24 hr)	% Improvement of IPE (24 hr)	% Predicted Error Improvement for Capacity Level			% Improvement of IPE for Capacity Level			% Pred. Error Improvement for Dem/Cap Level		% Improvement of IPE for Dem/Cap Level		% Pred. Error Improvement		% Improvement of IPE	
						High	Medium	Low	High	Medium	Low	High	Low	High	Low	During GDP	Post GDP	During GDP	Post GDP
BOS 2/25	301	273	48	30	31	41		25	46		25	21	32	23	33	18	55	18	57
BOS 3/7	474	122	20	15	14	18		10	19		10	11	15	11	16	6	27	6	28
BOS 3/15	504	161	24	16	17	22		8	22		8	9	20	10	20	12	9	13	11
LGA 3/4	432	117	21	22	23	*	*	*	*	*	*	10	27	14	28	6	59	7	63
ORD 3/6	889	249	22	15	16	20	14	10	19	12	9	12	19	9	19	10	21	11	20
SFO 3/15	568	37	6	6	7	6	4	2	6	4	10	6	6	7	6	9	8	12	9
SFO 3/16	539	62	10	9	9	4		10	5		11	5	12	6	14	6	16	7	18
STL 3/8	585	133	19	15	16	24		8	26		10	8	23	10	25	7	48	8	54

[7] Actual arrivals during 24 hour period starting at GDP inception.
[8] Total number of FX cancellations in 24-hour period.
[9] Number of FX Cancellations / (Actual 24-hour demand + number of FX Cancellations)
* No capacity variation during 24-hour period.

TABLE A.2. Summary of Average Normalized IPEs for Various Time Periods

Location & Date of GDP	Actual 24 Hour Arrival Demand[10]	FX Count[11]	FX Percent[12]	IPE for 24-Hour Period Beginning at GDP Inception		IPE for High Capacity Intervals During 24-Hour Period		IPE for Low Capacity Intervals During 24-Hour Period		IPE for Intervals of High Dem/Cap Ratio During 24-Hour Period		IPE for Intervals of Low Dem/Cap Ratio During 24-Hour Period		IPE During GDP		IPE During Post-GDP Period	
				ETMS	CDM	ETMS	CDM	ETMS	CDM	ETMS	CDM	ETMS	CDM	ETMS	CDM	ETMS	CDM
BOS 2/25	301	273	48	4.3	3.0	2.4	1.3	5.8	4.3	4.1	3.2	4.4	3.0	5.4	4.4	6.8	2.9
BOS 3/7	474	122	20	3.2	2.7	2.7	2.1	3.7	3.3	3.4	3.1	3.0	2.5	3.8	3.6	3.1	2.2
BOS 3/15	504	161	24	3.2	2.7	2.8	2.2	4.5	4.1	4.1	3.7	3.0	2.4	4.4	3.9	3.6	3.2
LGA 3/4	432	117	21	2.8	2.1	*	*	*	*	3.2	2.8	2.5	1.8	3.0	2.7	4.4	1.6
ORD 3/6	889	249	22	4.9	4.1	4.5	3.6	6.0	5.4	4.9	4.4	4.9	4.0	7.1	6.3	6.0	4.8
SFO 3/15	568	37	6	2.4	2.2	3.2	3.0	2.5	2.2	3.0	2.7	1.8	1.7	2.5	2.2	3.4	3.1
SFO 3/16	539	62	10	2.2	2.0	1.5	1.4	2.6	2.3	2.5	2.4	1.9	1.6	2.7	2.5	2.1	1.7
STL 3/8	585	133	19	3.4	2.8	2.8	2.1	3.8	3.4	3.9	3.5	2.8	2.1	3.9	3.5	3.7	1.7

[10] Actual arrivals during 24 hour period starting at GDP inception.
[11] Total number of FX cancellations in 24-hour period.
[12] Number of FX Cancellations / (Actual 24-hour demand + number of FX Cancellations)
* No capacity variation during 24-hour period.

APPENDIX B. Individual GDP Analysis Charts

As part of this study, the data from the eight GDPs listed in Table 1 were analyzed in various ways. Twenty charts (identified in the list below) were generated for each of the eight GDPs. Only a sampling of these charts is included in this report (see Section B.2). Section B.1 contains a discussion of the trends observed in the charts.

For the 24-hour period beginning at the start of the GDP, the following comparison charts were made:

a) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead.

b) Percentage Reduction in Arrival Demand Prediction Error due to CDM vs. Prediction Look-ahead.

c) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead for Periods of High Capacity.

d) Percentage Reduction in Arrival Demand Prediction Error due to CDM vs. Prediction Look-ahead for Periods of High Capacity.

e) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead for Periods of Low Capacity.

f) Percentage Reduction in Arrival Demand Prediction Error due to CDM vs. Prediction Look-ahead for Periods of Low Capacity.

g) Average Percent Reduction in Arrival Demand Prediction Error due to CDM as a Function of Capacity.

h) Average IPE of Arrival Demand (ETMS and CDM) as a Function of Capacity.

i) Percent Reduction in Average IPE of Arrival Demand Due to CDM as a Function of Capacity.

j) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead for Periods of High Demand to Capacity Ratio (dem/cap>= 0.7).

k) Percentage Reduction in Arrival Demand Prediction Error due to CDM vs. Prediction Look-ahead for Periods of High Demand to Capacity Ratio (dem/cap>= 0.7).

l) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead for Periods of Low Demand to Capacity Ratio (dem/cap< 0.7).

m) Percentage Reduction in Arrival Demand Prediction Error due to CDM vs. Prediction Look-ahead for Periods of Low Demand to Capacity Ratio (dem/cap< 0.7).

n) Average Percent Reduction in Arrival Demand Prediction Error due to CDM as a Function of Demand to Capacity Ratio.

o) Average IPE of Arrival Demand (ETMS and CDM) as a Function of Demand to Capacity Ratio.

p) Percent Reduction in Average IPE of Arrival Demand Due to CDM as a Function of Demand to Capacity Ratio.

Charts comparing the errors observed on ETMS and CDM during the GDP period and the four hours immediately following the GDP (post GDP period) were also produced. These include:

q) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead During the GDP.

r) Percentage Reduction in Arrival Demand Prediction Error due to CDM during the GDP.

s) Average Absolute Error of Arrival Demand Prediction (ETMS and CDM) vs. Prediction Look-ahead During the Post GDP Period.

t) Percentage Reduction in Arrival Demand Prediction Error due to CDM during the Post GDP Period.

B.1 Observations and Discussion of Charts

Figures B.2-1 through B.2-8 show average absolute errors for arrival demand predictions as a function of look-ahead value. These charts are all produced from data taken from the 24-hour period beginning at the time of each GDP start. Predictions were made every 15 minutes for the 24-hour period. As a result, 96 data points were used to calculate the average error for each look-ahead size.

The charts show the prediction error averages for both ETMS and CDM. The prediction error for each system is defined as the predicted value minus the actual value.

All of the charts display a reduction in error as the look-ahead value is reduced from forty to one (10 hours to 15 minutes). The average absolute error for CDM, based on all of the presented charts, is 3.4 flights per 15 minutes when the prediction is made at 40 quarter-hours (10 hours) of look-ahead. This error diminishes to 1.7 flights per 15 minutes when the prediction is made for look-ahead interval one (i.e., within the 15-minute interval starting at the current time).

The charts display a generally monotonic decrease in absolute prediction error for both systems from higher error values at large look-ahead values to smaller errors. This decrease generally persists until the look-ahead value is in the range of six to eight quarter hours (1½ to 2 hours). When the look-ahead time is in the range of one to six quarter hours (¼ to 1½ hours), a phenomenon known as the rolling spike[13] occurs. This rolling spike phenomenon shows a large increase in prediction errors that persists until the look-ahead value is one or two. Although absolute error values are presented here, the data shows that the sign of the errors during the rolling spike period is nearly always positive (that is the traffic demand is overestimated). The largest errors observed for these small look-ahead numbers are generally clustered within a two-hour period. During the times contributing to this phenomenon, the large and erroneous traffic volume predicted for one particular quarter-hour is pushed into the next succeeding quarter-hour when the next prediction is made 15 minutes later. According to the Rolling Spike Task Force Report, "the primary cause [of the rolling spike] is *the net redistribution of predicted demand to later time intervals as the … system is notified of arrival delays incurred both on the ground and in the air…."*

Because these charts represent absolute errors, they indicate a maximum value for the averages shown. Some of the predicted values are less than the actual demand resulting in a negative error. Charts produced showing averages of signed errors (not included in this report) show that the frequency of negatively signed errors (under-prediction of demand) increases when the look-ahead value is greater than eight (two hours or more into the future).

[13] R. Hoffman, L. Shisler, K. Howard, M. Klopfenstein, & M. Ball, "Report of the Rolling Spike Task Force," April 1, 1998 (available on http://www.isr.umd.edu/NEXTOR/publications.html)

A quick look at charts B.2-1 through B.2-8 shows that the CDM errors are consistently smaller than those of ETMS. A closer look, however, shows that the CDM predictions tend to be more accurate when the look-ahead period is roughly from three to eight hours. For look-ahead periods of less than two hours, both systems are in large part driven by radar data from active flights, and give similar predictions, except for the irregularities caused by the rolling spike phenomenon. For look-ahead periods of more than eight hours, both systems are in large part driven by OAG data, so they give comparable predictions. When the look-ahead period is in the three to eight hour range, ETMS is driven by the OAG data while CDM has the benefit of CDM message updates; consequently, this is the period when CDM tends to give better predictions.

The improvement in prediction accuracy of CDM over ETMS for each of the eight GDPs analyzed is shown in Figures B.2-9 through B.2-16. At each look-ahead value the percentage reduction in arrival demand prediction error is displayed for each GDP. The difference in the two error sizes is expressed as a percentage of the ETMS error. During the 3-hour to 8-hour prediction look-ahead times CDM typically provided an improvement of 15 to 30 percent over ETMS. CDM also shows a substantial improvement near the time of the rolling spike.

The improvements shown here were averaged for each GDP to provide the values shown in the summary Tables 1 and B.1.

B.2 Analysis Charts

Figure B.2-1 Arrival Demand Prediction Errors, BOS, 2/25/99

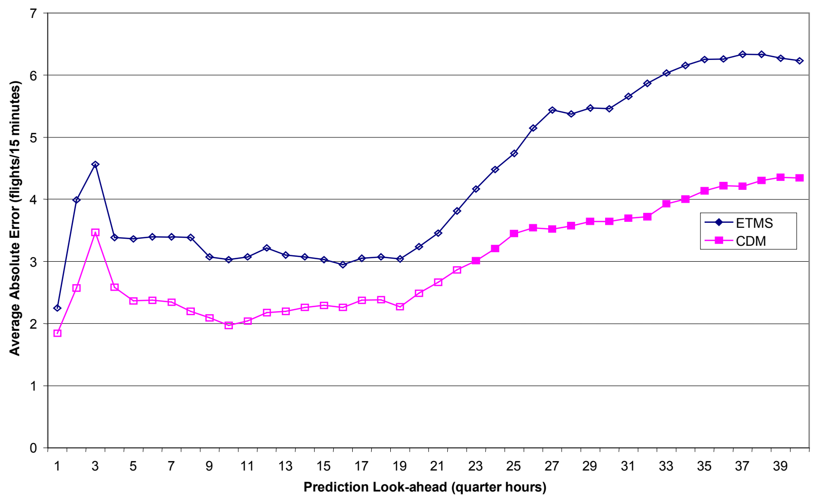

Figure B.2-2 Arrival Demand Prediction Errors, BOS, 3/7/99

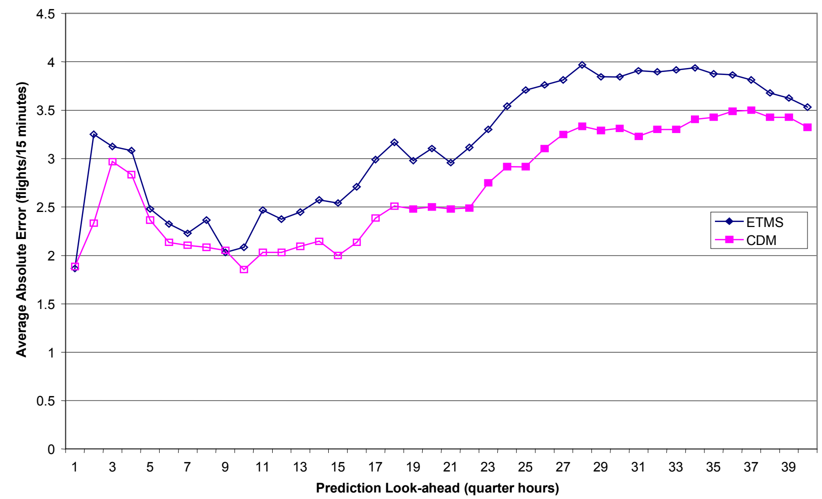

Figure B.2-3 Arrival Demand Prediction Errors, BOS, 3/15/99

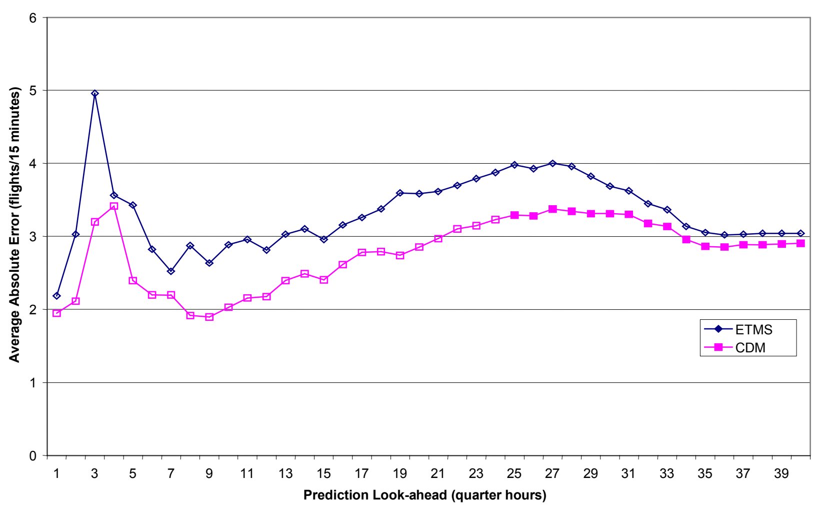

Figure B.2-4 Arrival Demand Prediction Errors, LGA, 3/4/99

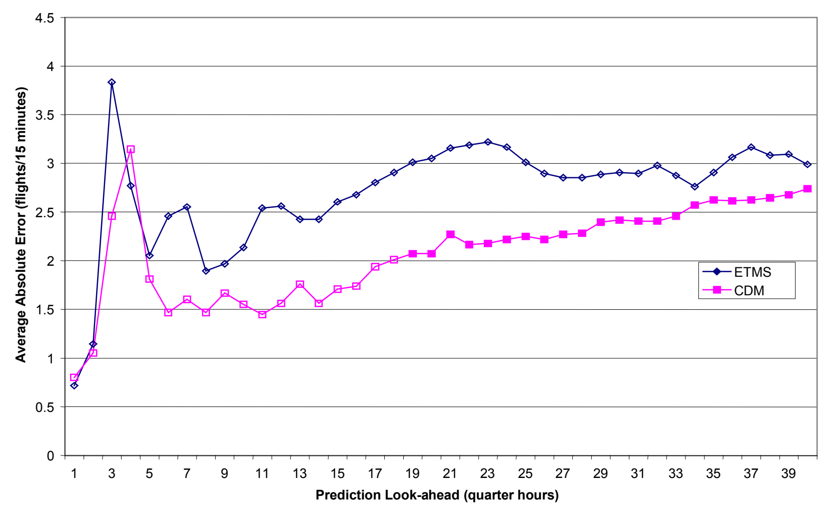

Figure B.2-5 Arrival Demand Prediction Errors, ORD, 3/6/99

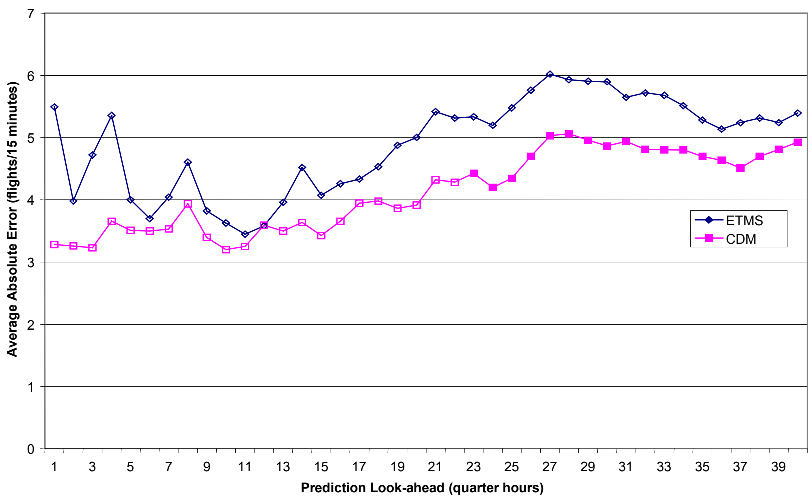

Figure B.2-6 Arrival Demand Prediction Errors, SFO, 3/15/99

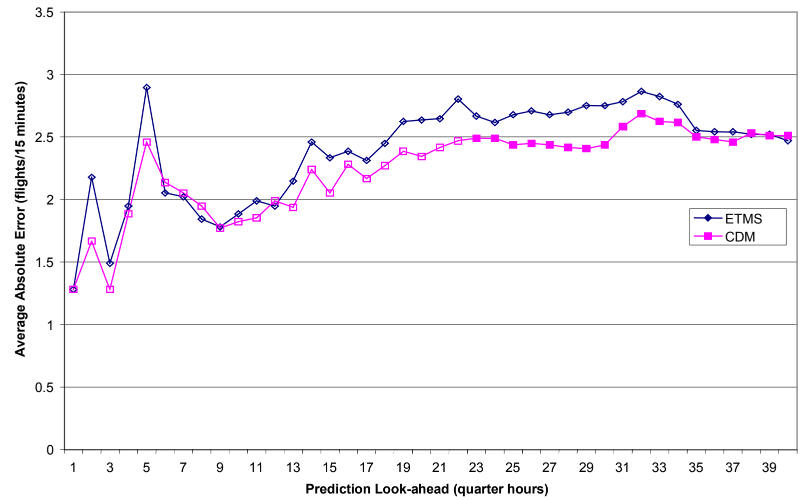

Figure B.2-7 Arrival Demand Prediction Errors, SFO, 3/16/99

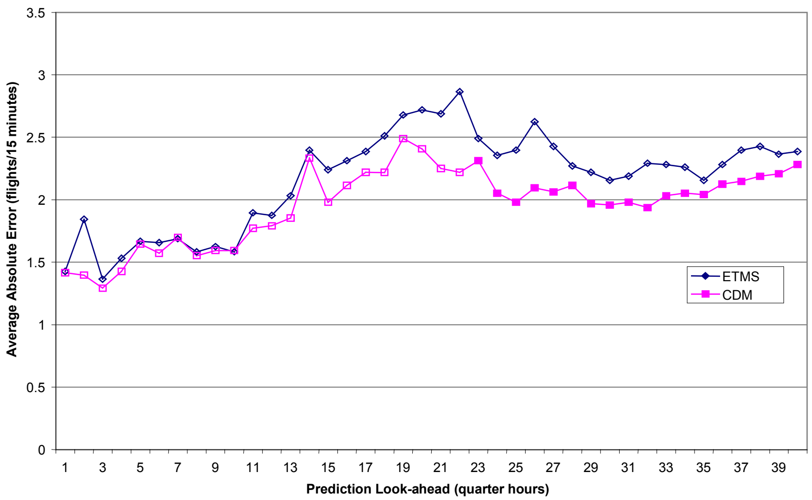

Figure B.2-8 Arrival Demand Prediction Errors, STL, 3/8/99

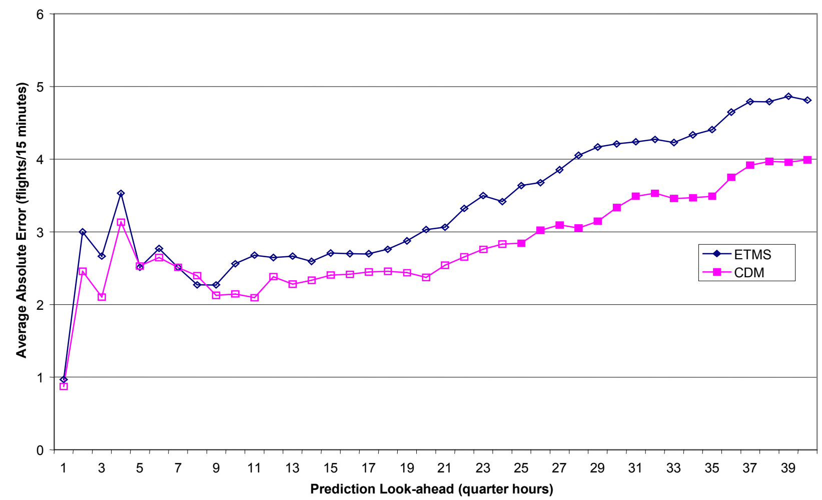

Figure B.2-9 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, BOS, 2/25/99

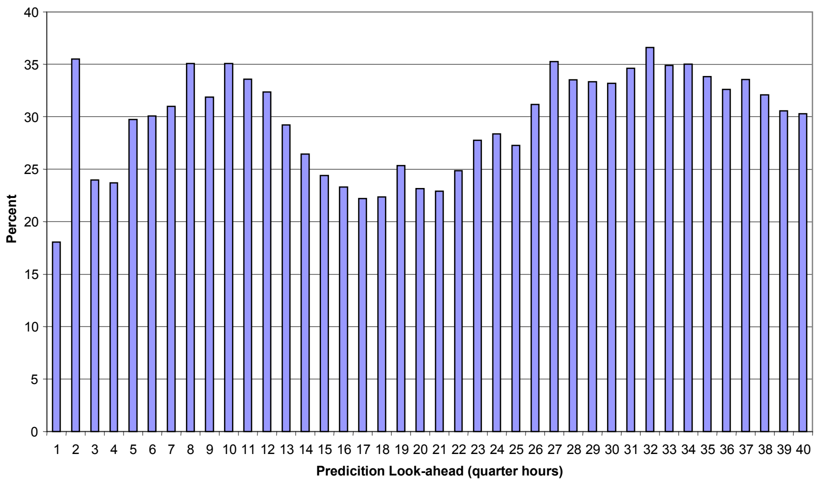

Figure B.2-10 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, BOS, 3/7/99

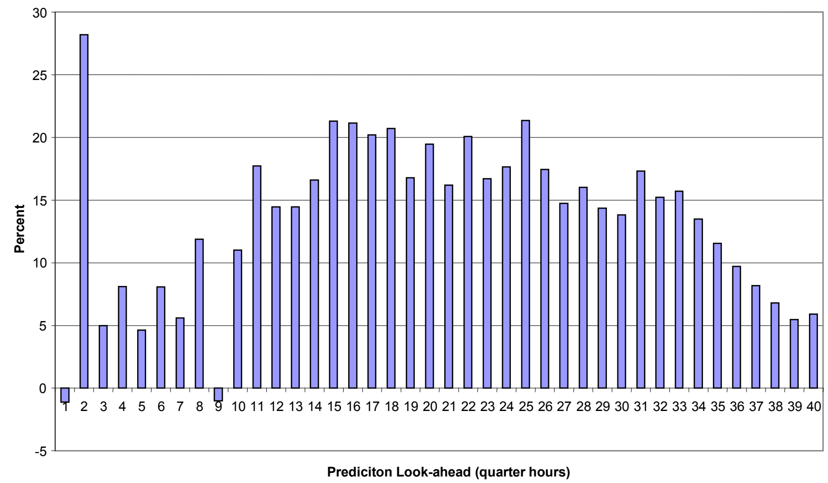

Figure B.2-11 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, BOS, 3/15/99

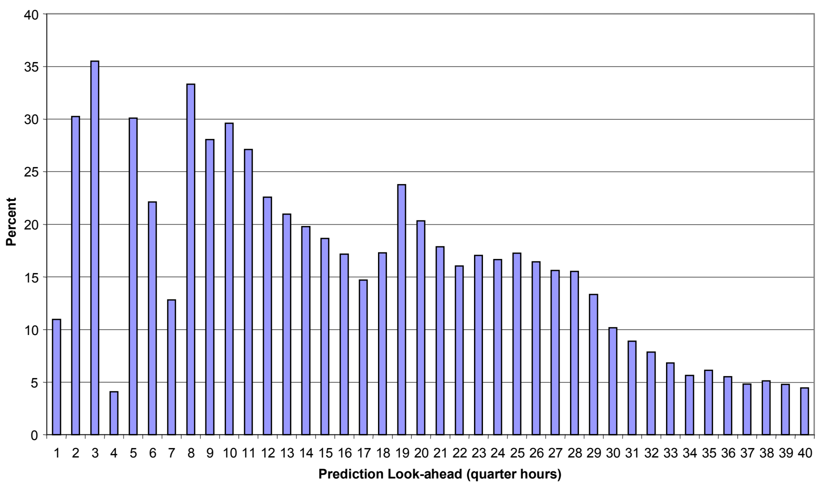

Figure B.2-12 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, LGA, 3/4/99

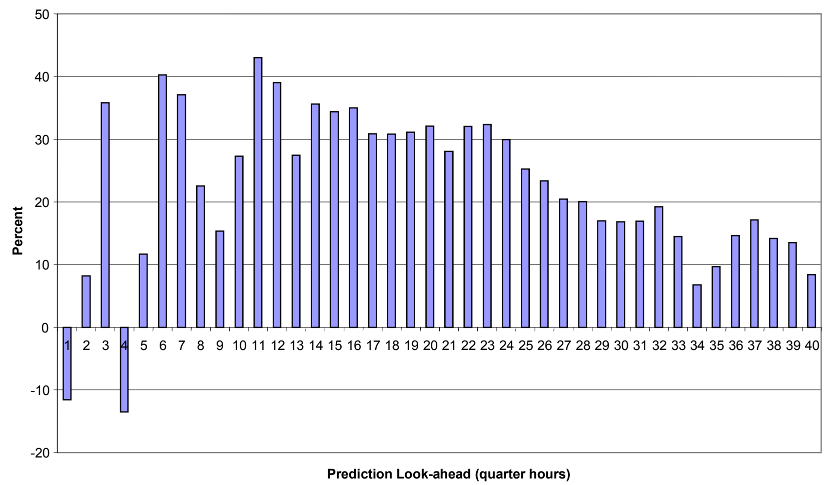

Figure B.2-13 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, ORD, 3/6/99

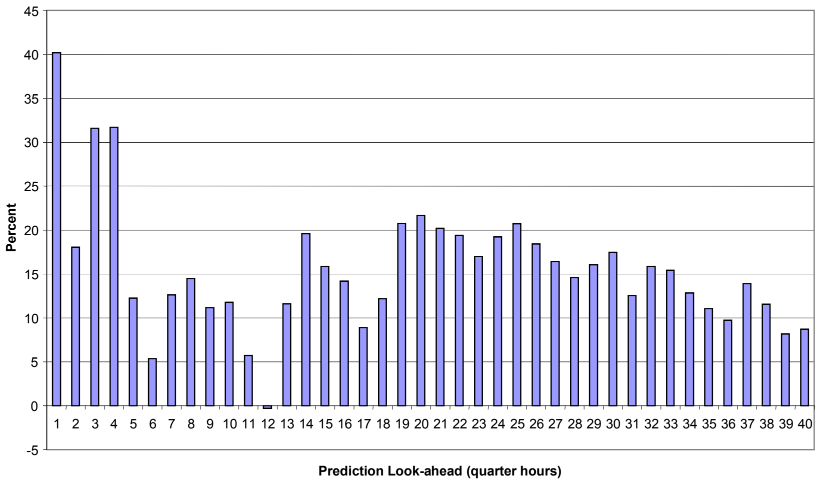

Figure B.2-14 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, SFO, 3/15/99

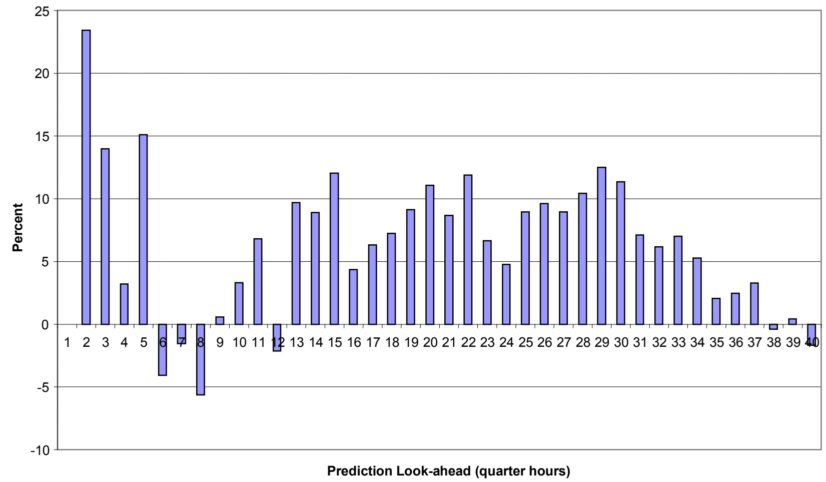

Figure B.2-15 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, SFO, 3/16/99

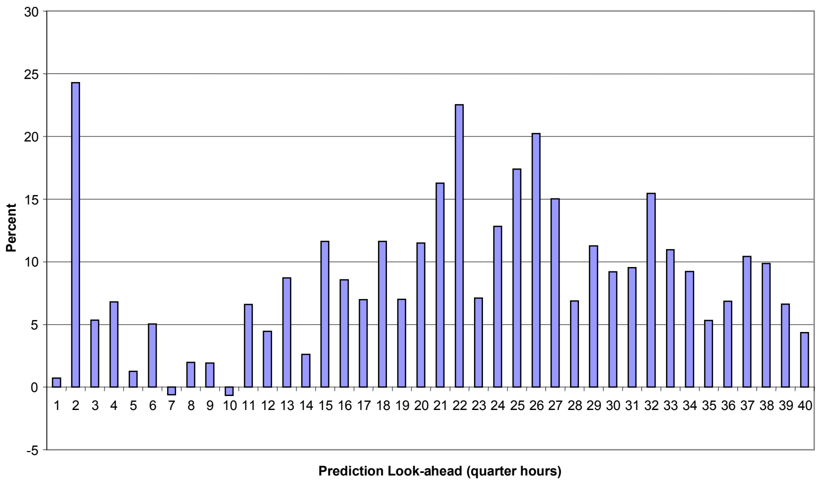

Figure B.2-16 Percentage Reduction in Arrival Demand Prediction Error Due to CDM, STL, 3/8/99