Weather-Weighted Periodic Auto Regressive Models for Sector Demand Prediction

Neil Y. Chen*
University of California, Santa Cruz, Moffett Field, CA 94035-1000
Banavar Sridhar†
NASA Ames Research Center, Moffett Field, CA 94035-1000

This paper describes a class of weather-weighted periodic auto-regressive sector demand prediction models. The periodic auto-regressive model captures both the mid-term (30 minutes to 2 hours) trend based on the historical data, and the short-term (less than 30 minutes) transient response based on recent observations. For severe weather days, the model uses the three-dimensional weather information, considering both storm locations and echo tops, to form a weather factor to adjust the predictions. Unlike the traditional trajectory-based sector demand prediction methods, which predict the behavior of the National Airspace System adequately for short durations of up to 20 minutes and are vulnerable to the weather uncertainties, this class of models provides reliable short to mid-term sector demand predictions which account for the weather uncertainty.

I. Introduction

Demand for air transportation has grown rapidly in recent years and is expected to grow in the future. In order to ensure smooth air traffic flow and safety in the presence of disruptions caused by convective weather, innovative modeling and design methods are needed in traffic flow management (TFM). One of the main functions of TFM is to predict and resolve demand-capacity imbalances at the sector level to avoid congestion. Thus an accurate sector prediction model that can account for traffic flow uncertainty and weather impact is an essential component of TFM.

Efforts have been made in the past few years to perform sector demand predictions. Traditionally, models used in air traffic control and flow management are based on simulating the trajectories of individual aircraft. Deterministic forecasting of sector demand is routinely done within the Enhanced Traffic Management System (ETMS), which relies on the computation of each aircraft’s entry and exit times at each sector along the path of flight. Gilbo\(^1\) proposed a regression model for improving aggregate traffic demand prediction in ETMS, acknowledging the uncertainty in the predictions. A more recent TFM simulation tool, the Future ATM Concepts Evaluation Tool (FACET),\(^2\) was used to propagate the trajectories of the proposed flights forward in time and use them to count the number of aircraft in each sector for demand forecasting and establish confidence bounds on the forecasts.\(^3\) These trajectory-based models predict the behavior of the National Airspace System (NAS) adequately for short durations of up to 20 minutes and their accuracy of predictions is impacted by weather and trajectory prediction uncertainties.\(^4\)–\(^6\)

The objective of this paper is to develop an empirical sector prediction model that accounts for traffic flow uncertainty and weather impact on the prediction for both short-term (less than 30 minutes) and mid-term (30 minutes to 2 hours) predictions. Unlike the traditional methods that use trajectory prediction to perform the sector demand prediction, the periodic autoregressive (PAR) model and its variants\(^7\)–\(^8\) were used to build the prediction model. The class of PAR models consider both the historical traffic flows to capture the mid-term trend, and the flows in the near past to capture the transient response. In addition, a weather component was embedded in the model to reflect weather impact on sector demand.

\*Research Scientist, U.C. Santa Cruz, MS 210-8.
†Senior Scientist for Air Transportation Systems, Aviation Systems Division, Fellow.
The remainder of the paper is organized as follows. Section II provides the sector demand data used in the research and the description of the weather-free sector demand prediction model. Next, a weather factor is introduced and the prediction model that considers weather is described in Section III. The results and performance of the models are demonstrated in Section IV. Finally, a summary and conclusions are presented in Section V.

II. Data and Model

II.A. Sector Demand Data

The air traffic demand data used in this paper are provided by the recorded Aircraft Situation Display to Industry (ASDI) data generated by the Federal Aviation Administration’s Enhanced Traffic Management System (ETMS). The ASDI data provide the locations of all aircraft at one-minute intervals. The sector demand, defined as the number of aircraft in each sector at a given time, can be computed using the ASDI data. In this research, the recorded ASDI data were processed using FACET to obtain the sector demand.

Since traffic flow management decisions are made by comparing the peak number of aircraft in a sector during a fifteen-minute interval with the sector’s Monitor Alert Parameter (MAP) value, the 15-minute peak sector demand, defined as the maximum sector demand every 15 minutes, was used to build the models. Figure 1 shows the sector demand at every minute and the 15-minute peak sector demand at sector ZID93 on September 3, 2007. Note that in the paper, a day is defined as a 24-hour interval starting at 4:00 AM local time since most of the aircraft departing on the previous day would have landed before 4:00 AM. The black line in Fig. 1 represents the sector demand in an one-minute intervals and the blue dots in Fig. 1 represent the 15-minute peak sector demand during a day, denoted as $d_k$, where $k = 1 \ldots 96$. The mid-term trend of sector demand on different days can be observed in Fig. 2, which shows the variation of 15-minute peak sector demand in September 2007. In this figure, each horizontal strip represents one day of 15-minute peak sector demand, and each vertical strip represents the peak sector demand at the same time of day during the entire month. As shown in the figure, the horizontal strips on 9/1, 9/8, 9/15, 9/22 and 9/29, which are Saturdays, have lower demands than the others. The blue vertical regions on the left and right show the off-peak traffic in the early morning and the late night. A vertical light blue region at around 12 o’clock divides the sector demand into morning rush left of the region and the afternoon peak right of it. The sector demand prediction model presented in the next section captures these variations in the demand.

II.B. Demand Prediction Model

Auto-regressive models have been used for short-term hourly air traffic delay prediction.\textsuperscript{9,10} This research extends the delay prediction approach to sector demand prediction.

For a given day, a 24-hour period, starting at 4:00 AM local time, is divided into 96 fifteen-minute intervals. Given the observed 15-minute peak sector demands for $n$ days, the sector demand data matrix is defined as

$$ D = \begin{bmatrix} d_{1,1} & d_{2,1} & \cdots & d_{96,1} \\ \vdots & \vdots & \ddots & \vdots \\ d_{1,n} & d_{2,n} & \cdots & d_{96,n} \end{bmatrix}, \quad (1) $$

where $d_{i,j}$ represents the 15-minute peak sector demand at the $i^{th}$ time step on day $j$. For September 2007, $D$ has a dimension of 30 by 96, and Fig. 2 shows the image of the matrix $D$. Assuming $d_k$ as the $k^{th}$ column of $D$, the $p$-step-ahead sector demand model at time step $k$ in the form of a linear regression model is described as

$$ d_{k+p} = \alpha_{k,p} d_k + \beta_{k,p} + e_k, \quad (2) $$

where $\alpha_{k,p}$ and $\beta_{k,p}$ are coefficients that map the sector demand at the $k^{th}$ time step to the $(k + p)^{th}$ time step, and $e_k$ is the error of the model. The least-square solution of $\alpha_{k,p}$ and $\beta_{k,p}$ that minimizes $e_k^T e_k$ in Eq. (2) can be written explicitly as

$$ \hat{\alpha}_{k,p} = \frac{\sum_{i=1}^{n} (d_{k,i} - \bar{d}_k)(d_{k+p,i} - \bar{d}_{k+p})}{\sum_{i=1}^{n} (d_{k,i} - \bar{d}_k)^2}, \quad (3) $$

$$ \hat{\beta}_{k,p} = \bar{d}_{k+p} - \alpha_{k,p} \bar{d}_k, \quad (4) $$

American Institute of Aeronautics and Astronautics
where $\bar{d}_k$ is the mean of $d_k$, and $\bar{d}_{k+p}$ is the mean of $d_{k+p}$.

On a day $m$ other than the $n$ days in the data set, the $p$-step prediction of the sector demand at the $k^{th}$ time step, $\hat{d}_{k+p,m}$, based on the observed sector demand, $d_{k,m}$, can then be expressed as

$$\hat{d}_{k+p,m} = \alpha_{k,p} d_{k,m} + \beta_{k,p}.$$  (5)

In the model, $\alpha_{k,p}$ and $\beta_{k,p}$, identified from the historical data, capture the periodic features during a day, and the observed sector demand $d_{k,m}$ provides the transient information. The model in Eq. (2) and Eq. (5) is referred to as the periodic auto-regressive (PAR) sector demand prediction model.

As an example, peak sector demand data in August 2007 were used to construct the data matrix in Eq. (1), and Eq. (2) was used to identify the model parameters $\alpha_{k,p}$ and $\beta_{k,p}$, where $k = 1 \ldots 96$ and $p = 1 \ldots 8$ for 1-step- to 8-step-ahead predictions. The 15-minute-ahead peak sector demand on September 3, 2007 was predicted using Eq. (5) with $p = 1$. The result is shown in Fig. 3a. The root-mean-squared (RMS) error between the actual peak sector demand and the 15-minute demand prediction is 1.93. For the 2-hour prediction, the prediction model is solved for $p = 8$, and the estimates in Eq. (5) are generated. The result is shown in Fig. 3b. The RMS error is 2.15.

It is noticed that the PAR model yields larger error as the prediction interval increases. This suggests that using a single observation $d_{k,m}$ in Eq. (5) contains less information about $d_{k+p,m}$ when $p$ is large. An alternate method to perform the demand prediction is to use the cumulative sum of the past sector demands as an observation, since the sum includes more information than a single observation and has less noise compared with the single peak sector demand. Following the definition of the sector demand matrix $D$ in
Eq. (1), where $d_k$ is the $k^{th}$ column of $D$, the cumulative $p$-step-ahead sector demand model at time step $k$ can be described in terms of the cumulative sum of $q$ past sector demands as

$$
\mathbf{d}_{k+p} = \alpha_{k,p} \sum_{i=k-q+1}^{k} \mathbf{d}_i + \beta_{k,p} + \mathbf{e}_k,
$$

(6)

where $\alpha_{k,p}$ and $\beta_{k,p}$ are the coefficients that map the cumulative sector demand at the $k^{th}$ time step to the sector demand at the $(k+p)^{th}$ time step. Once the least-square solution of coefficients $\hat{\alpha}_{k,p}$ and $\hat{\beta}_{k,p}$ are identified, the $p$-step prediction of the sector demand at the $k^{th}$ time step for a day $m$, $\hat{d}_{k+p,m}$, based on the observed cumulative sector demand, $\sum_{i=k-q+1}^{k} d_{i,m}$, can be expressed as

$$
\hat{d}_{k+p,m} = \hat{\alpha}_{k,p} \sum_{i=k-q+1}^{k} d_{i,m} + \hat{\beta}_{k,p}.
$$

(7)

The model in Eq. (6) and Eq. (7) is referred to as the cumulative periodic auto-regressive (CPAR) sector demand prediction model. During the analysis, it is noticed that the CPAR model using the sum of the all demands in the past ($q = k$) works best overall. For the example used in the PAR model, the CPAR model with $q = 8$ has a RMS error of 1.68 for the 15-minute prediction, and 2.00 for the 2-hour prediction, compared with 1.93 and 2.15 respectively for the PAR model. It appears that the CPAR model performs a little better than the PAR model. More analysis is done in Section IV to evaluate this property.

### III. Weather Factor

Weather has a big influence on air traffic sector demand and the uncertainty in weather may cause error in the predictions. If a severe storm blocks a sector or regions near it, both the sector capacity and demand may drop dramatically. A weather factor that discounts the weather-free sector demand prediction is derived in this section.

In order to model the weather impact on sector demand prediction, an accurate weather forecast product with high update rate is required. In addition, to capture the impact on all low, high, and super high sectors, the storm echo tops information is useful. The weather data used in this paper was provided by the Corridor Integrated Weather System (CIWS). CIWS, developed and operated by MIT Lincoln Laboratory, provides both accurate precipitation and echo tops data and is updated every 5 minutes. In addition, CIWS provides convective forecasts at 5-minute intervals up to 2 hours in the future.

The weather factor used to discount the sector demand prediction was chosen to be the sector weather index, defined as the percentage of area covered by the storm with precipitation vertically integrated liquid...
(VI) level three and above. Only storms with the echo tops above the lower boundary of the sector are considered. The sector weather index at time $k$ is formulated as

$$w_k = \frac{A_k^w}{A},$$

where $A$ is the area of the sector and $A_k^w$ is the area of the sector covered by storms with the echo tops at or above the lower bound of the sector at time $k$. Note that if time $k$ is a future time, the weather forecast is used to determined $A_k^w$. It is possible to use other definitions of a sector weather index.\(^{12,13}\)

Figure 4a shows a snap shot of the CIWS data for the high altitude sectors at Indianapolis center (ZID) on a severe weather day. The red spots indicate the storms with VI level 3 and above, and the echo tops at 35,000 ft. As shown in this figure, most of the sector ZID93 is covered by the storm. The sector weather index for ZID93 on August 16, 2007 is shown in the red line in Fig. 4b. Also shown in the figure is the actual sector demand on the same day in blue line. Notice the sector weather index is greater than 30% from 18:00 to 20:00 Eastern Daylight Time (EDT), and clearly the sector demand drops during the same period.

Traffic reduction due to weather impact can be modeled in many different ways.\(^{15}\) In this research, the weather-free prediction was first estimated, then the sector weather index was used to adjust the prediction. Assume that the sector demand starts to decay when the sector weather factor exceeds $w_{low}$, and reaches 0 when the weather factor reaches $w_{high}$. The sector demand reduction rate is modeled as the power law distribution, $1 - (w_k - w_{low})/(w_{high} - w_{low})^\gamma$, where $\gamma$ is the power of the distribution. To reflect the thresholds, the sector weather index in Eq. (8) is redefined as

$$w_k = \begin{cases} 
  w_{low} & \text{if } A_k^w / A \leq w_{low} \\
  A_k^w / A & \text{if } w_{low} < A_k^w / A < w_{high} \\
  w_{high} & \text{if } w_{high} \leq A_k^w / A 
\end{cases}.$$

In order to adjust the weather impact on the sector demand prediction model, the weather forecast is used to compute the predicted sector weather index. Assume at time $k$, the predicted sector weather index at time $k + p$ is $w_{k+p}$, the PAR sector demand prediction model in Eq. (5) can be rewritten as

$$\hat{d}_{k+p,m} = \left(1 - \left(\frac{w_{k+p} - w_{low}}{w_{high} - w_{low}}\right)^\gamma\right)\left(\hat{\alpha}_{k,p} d_{k,m} + \hat{\beta}_{k,p}\right),$$

or the CPAR sector demand prediction model in Eq. (7) can be rewritten as

$$\hat{d}_{k+p,m} = \left(1 - \left(\frac{w_{k+p} - w_{low}}{w_{high} - w_{low}}\right)^\gamma\right)\left(\hat{\alpha}_{k,p} \sum_{i=k-q+1}^{k} d_{i,m} + \hat{\beta}_{k,p}\right).$$
Using the echo tops information provides a more representative weather index, especially for the high sectors. If there are storms with low echo tops located at some high sectors, the weather might have minimal impact on the sector demand. The sector demand and weather index for sector ZID93 on two different days is shown in Fig. 5. Both days have severe storms, but one has high echo tops while the other has low echo tops. The sector demands on severe weather days were compared with the average sector demand on the rest of the days in the same month. In Fig. 5a, the sector demand on August 16, 2007 is lower than the average between 18:00 and 20:00 (EDT) because of the high weather index during the period, as indicated in Fig. 5c. The blue line in Fig. 5c shows the weather index considering the area covered by storms without the echo tops information, and the red line is the weather index considering the echo tops at 35,000 ft and above. In this case, the two lines are close. This suggests that there are severe storms in the area and most of the echo tops are higher than the lower bound of sector ZID93. On the other hand, on October 23, 2007, there is no demand reduction compared to the average of the other days in October 2007 during 18:00 and 20:00 (EDT), shown in Fig. 5b, even though there are storms in the sector during the period, as shown in Fig. 5d. The red line in Fig. 5d is substantially lower than the blue line, which means even though there are storms in the sector, most the echo tops are lower than the low boundary of the sector and have minor impact on the sector demand. In the next section, the sector weather index refers the index with the echo tops information.

IV. Results

The sector demands of 25 high and superhigh sectors in ZID were investigated in this research. The major flows of ZID include the arrivals to Philadelphia International Airport (PHL), Ronald Reagan Washington National Airport (DCA), Chicago O’Hare International Airport (ORD), Detroit Metropolitan Wayne County Airport (DTW), and Cleveland-Hopkins International Airport (CLE), the departures from ORD and DTW, the westbound traffic of airway J80 from New York Center (ZNY) and Boston Center (ZBW), and the traffic to New York Terminal Radar Approach Control (N90). The sector demands for the month of August, 2007 were used to build the sector demand prediction PAR and CPAR models, described in Eq. (1), Eq. (2), and Eq. (6). The time step used in the models is 15 minutes. Once the parameters were identified, Eq. (5) and Eq. (7) were used to perform the sector demand prediction for the month of September, 2007. Starting
from the 15-minute prediction model, up to 2-hours prediction model were built and evaluated. The results of four super-high sectors ZID91, ZID92, ZID93, and ZID94, and four high sectors ZID81, ZID82, ZID83, and ZID84 in the southwest region of ZID, as shown in Fig. 6, were presented. The behavior of both PAR and CPAR models are summarized in Table 1. Even though the performance of the two models are very close, CPAR prediction models perform equal to or better than the PAR models in all the cases with the exceptions of the 15-min prediction at ZID81 (PAR 1.95, CPAR 1.98) and at ZID82 (PAR 1.57, CPAR 1.58). Among the cases, the error in CPAR models is 2.46% smaller than the error in PAR models in average. Also notice the errors of both PAR and CPAR models are not sensitive to the look ahead time. In general, the errors are larger with longer look ahead time, but only slightly. The errors of 120-min prediction is 5.12% larger than the 15-min prediction in average for the PAR models, and 2.87% larger for the CPAR models. Consider all the high and super-high sector in ZID, the results are similar. The errors of the PAR models are between 1.57 and 2.11 in the 15-min prediction, and between 1.64 and 2.24 in the 120-min prediction, while the errors of the CPAR models are between 1.58 and 2.10 in the 15-min prediction, and between 1.61 and 2.15 in the 120-min prediction.

### Table 1. Sector demand prediction error of the PAR and CPAR models in September 2007. The model is built using August 2007 data. The smaller errors in each case are in bold. The unit is the number of aircraft.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Average prediction RMS error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15-min</td>
</tr>
<tr>
<td>ZID81</td>
<td>17</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID82</td>
<td>16</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID83</td>
<td>16</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID84</td>
<td>16</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID91</td>
<td>19</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID92</td>
<td>17</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID93</td>
<td>19</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
<tr>
<td>ZID94</td>
<td>17</td>
</tr>
<tr>
<td>PAR</td>
<td>2.06</td>
</tr>
<tr>
<td>CPAR</td>
<td>2.04</td>
</tr>
</tbody>
</table>
The sector demand prediction for bad weather days uses the weather factor described in the previous section to adjust the weather-free prediction, formulated in Eq. (9), Eq. (10), and Eq. (11), with $w_{\text{low}} = 0$, $w_{\text{high}} = 1$, and $\gamma = 1$. The days with peak weather factors greater than 30% are considered bad weather days. For the days and sectors tested, there are four cases of severe weather periods: ZID83 on 08/16/07 between 1600-2200 (EDT), ZID93 on 08/16/07 between 1600-2200 (EDT), ZID82 on 08/21/07 between 0600-1400 (EDT), and ZID92 on 08/21/07 between 0800-1400 (EDT), shown in Fig. 7. Since all these cases happened in August 2007, the model is built using data for July 2007. Two types of weather-weighted models are built, one uses the actual weather information and the other uses the forecast weather information. Using the actual weather information to perform sector demand prediction represents the cases with perfect weather forecast. It is used to evaluate the performance of the weather-weighted prediction model and eliminate the error caused by forecast inaccuracy. The average prediction errors of the four severe weather periods in August 2007 are shown in Fig. 8. It is noticed that in all four cases, both the weather-weighted model using actual weather information (red dash line) and the model using forecast weather (green dash-dot line) produce smaller error than the weather-free model (blue solid line). The weather-weighted model using forecast weather performs as well as the model using actual weather when the prediction time is small (less than 30 minutes). However, with longer prediction time (more than 60 minutes), the performance starts to decay and the errors are closer to the weather-free model. As an example, in Fig. 8b, the weather-weighted sector demand prediction model using actual weather information improves the 15-minute prediction over the weather-free model by 36.38%, the 60-minute prediction by 42.92%, and the 120-minute prediction by 40.77%. For the weather-weighted model using forecast weather, the improvement is 34.73% for the 15-minute prediction, reduced to 27.81% for the 60-minute prediction, and down to 7.71% for the 120-minute prediction. This suggests that with longer prediction time, the forecast inaccuracy might effect the performance of the weather-weighted prediction model using forecast weather.

V. Conclusion

A class of auto-regressive models developed for sector delay estimation is used for predicting traffic demand in a sector between 15 minutes and two hours in the future. The PAR and CPAR models capture both the mid-term trend based on the historical data, and the short-term transient response based on the near past observation. For the sectors tested, the errors of CPAR models are 2.46% smaller than the PAR models. The CPAR model provides the demand predictions with an average RMS error between 1.58 and 2.10 in the
Figure 8. Average prediction errors during severe weather periods; blue solid line: weather-free model, red dash line: weather-weighted model using actual weather, green dash-dot line: weather-weighted model using forecast weather.

15-min prediction, and between 1.61 and 2.15 in the 120-min prediction. The performance of the prediction only decays slightly as the prediction interval is increased from 15-minute to 2-hour in both the PAR and CPAR models, as the error increases 5.12% in PAR models and 2.87% in the CPAR models. To improve the accuracy of sector demand prediction in the presence of severe weather, the paper introduced the concept of weather factor. For severe weather days, the model uses the three-dimensional weather information, considering both storm location and echo tops to form the weather factor and then adjusts the weather-free prediction. The weather-weighted model improves the sector demand prediction by as much as 34.73% for the 15-minute prediction, 27.81% for the 60-minute prediction, and 7.71% for the 120-minute prediction on the days and sectors tested. Unlike traditional trajectory-based sector demand prediction methods which predict the behavior of the National Airspace System adequately for short durations of up to 20 minutes and are vulnerable to weather uncertainties, the weather-weighted periodic auto-regressive models provide a reliable short- to mid-term sector demand prediction which accounts for weather uncertainty.

References


