540 THE UTILIZATION OF CURRENT FORECAST PRODUCTS IN A PROBABILISTIC AIRPORT CAPACITY MODEL

Jonathan Cunningham*, Lara Cook, and Chris Provan

Mosaic ATM, Inc., Leesburg, VA

1. INTRODUCTION

According to FAA statistics, approximately 70% of all air traffic delays and flight cancellations in the National Airspace System (NAS) are caused by severe weather. While enhancements in current forecast products and techniques have significantly improved in recent years, additional improvements are needed in both the tactical (1-2 hours) and strategic (greater than 3 hours) time horizons. This deficiency was apparent while developing a weather translation model that integrates a weather forecast and its inherent uncertainty to produce a probabilistic prediction of the weather's impact on the NAS. The research presented in this paper concentrates on predicting the airport arrival rate (AAR) for the purpose of planning Ground Delay Programs, or GDPs. A GDP is a traffic management initiative (TMI) where aircraft are delayed at their departure airport in order to manage demand at an airport whose capacity is constrained due to airport construction, equipment failures, or drastic increases in airport demand, or weather. The primary cause of GDP issuance is weather, which accounted for over 90% of the GDPs between 2008 and 2010 (Chan et al. 2011). Over these three years, the most frequently cited weather factors were ceilings, wind, and thunderstorms. However, the uncertainty of the airport capacity predictions in these GDPs is directly related to the uncertainty of the forecast product utilized (Provan et al. 2011). Especially considering that air traffic managers are required to predict the capacity at a given airport up to 10 hours or more into the future, the uncertainty is quite significant. To account for this uncertainty, the Weather Translation Model for GDP Planning (WTMG) was developed to translate the weather forecasts into probabilistic AAR predictions. To make these predictions, the model finds correlations between historical weather forecast and airport capacity data and uses these correlations to project the future airport capacities based on recently issued forecasts. The following sections will give a general overview of the model itself, a review of the weather forecast products used and the necessary steps taken to incorporate them into the model, a brief summary of our results, and then conclusions and opportunities for future work and improvements.

* *Corresponding author address:* Jonathan Cunningham, Mosaic ATM, Inc., 801 Sycolin Road, SE, Suite 306, Leesburg, VA 20175;email:jcunningham@mosaicatm.com

2. MODEL OVERVIEW

WTMG is a two-part, self-training statistical model that operates on a strategic time scale in order to provide air traffic managers with enough time to plan GDPs. The model operates in two modes: static and dynamic. The static WTMG independently generates each future hour's probabilistic AAR prediction based on the forecast and airport capacity information available at the time of prediction. In this mode, each hour's AAR prediction is made independent of the preceding hours in the forecast horizon. The dynamic mode of WTMG generates sample AARs for each future hour's probabilistic AAR prediction based on the available weather information at the time of the prediction and the previous hour's sample AARs. In contrast to the static mode, this mode of the model accounts for the dependence of the AAR for a given hour on the AAR of the previous hour by using that AAR to condition the error distribution that is used to create the capacity samples. The following subsections provide greater information about each part of the model.

2.1. Prediction Model

The first part of the model, known as the Prediction Model, uses historical weather data (observation and forecast) and observed hourly AARs to make deterministic AAR predictions. The data used in our research came from two airports, Newark Liberty International Airport (EWR) and O'Hare International Airport (ORD) between 2008 and 2010. The various data sources from these airports are organized chronologically based on the observation time of the recorded METARs. Given a forecast lead time of x hours, a data point is created by matching the actual hourly airport capacity observations (outcomes) and the observed weather conditions to the valid weather forecasts (indicators) issued x hours in the past. Additionally, depending on the mode in which the model is operating, the AAR at the time of the prediction (static) or the AAR from the previous hour (dynamic), known within the model as the "lead AAR", is added to the array of indicators. For example, given a forecast lead time of 6 hours, the actual AAR and the weather observation time of 12 UTC is matched to the forecast information from 06 UTC. If the WTMG is operating in static mode, the lead AAR would be the AAR observed at the time of the prediction (06 UTC), while if it is operating in dynamic mode, the lead AAR would be the AAR observed at 11 UTC.

Using this data, the prediction model applies a bootstrap regression tree methodology, implemented in MATLAB's *TreeBagger* function. This function first separates the data points to two sets: a training set used to teach the model and a test set used to assess the accuracy of the model. Using the training set, the *TreeBagger* function resamples the data to create a user-defined number of bootstrap samples in order to increase the robustness of the predictions by simulating the effect of having a larger set of data. A regression tree is then built from each bootstrap sample to attempt to correlate the lead AAR and weather forecast data (indicators) with the actual, observed AARs (outcomes). Once a regression tree has been built for each bootstrap sample from the training set of data, the test set of indicators is used to derive a set of deterministic predictions by taking the mean of all of the predictions from each individual tree. This process is shown in the figure below.

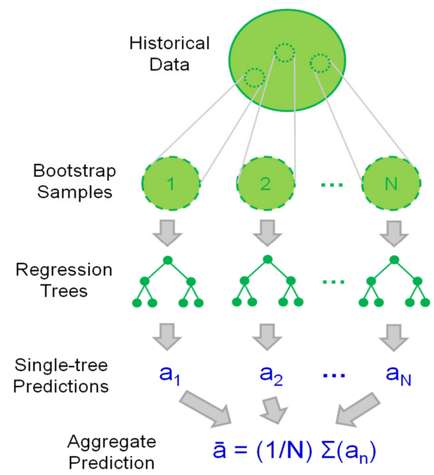

Figure 1. Graphical overview of the Prediction Model within the Weather Translation Model for GDP Planning (WTMG)

2.2. Sampling Model

The second part of the model, known as the Sampling Model, uses the deterministic predictions made by the prediction model to build a set of probabilistic capacity scenarios that account for the uncertainty in the predictions. This is accomplished by first building an empirical error distribution around each deterministic AAR prediction. This is done by grouping all of the actual AAR values (from the original data set) for each unique predicted AAR. For example, all of the actual AAR data points that resulted in a predicted AAR of 38 are grouped together to create a single error distribution around a predicted AAR of 38.

The generation of the sample AAR scenarios depends on the mode of WTMG: static or dynamic. The static mode of the model (see Figure 2A) generates each future hour's probabilistic AAR prediction based on the current AAR of the airport and the forecast information available at the time of the prediction. The dynamic mode of WTMG (see Figure 2B) generates each future hour's probabilistic AAR prediction based on the previous hour's sample AAR and the forecast information valid at the time

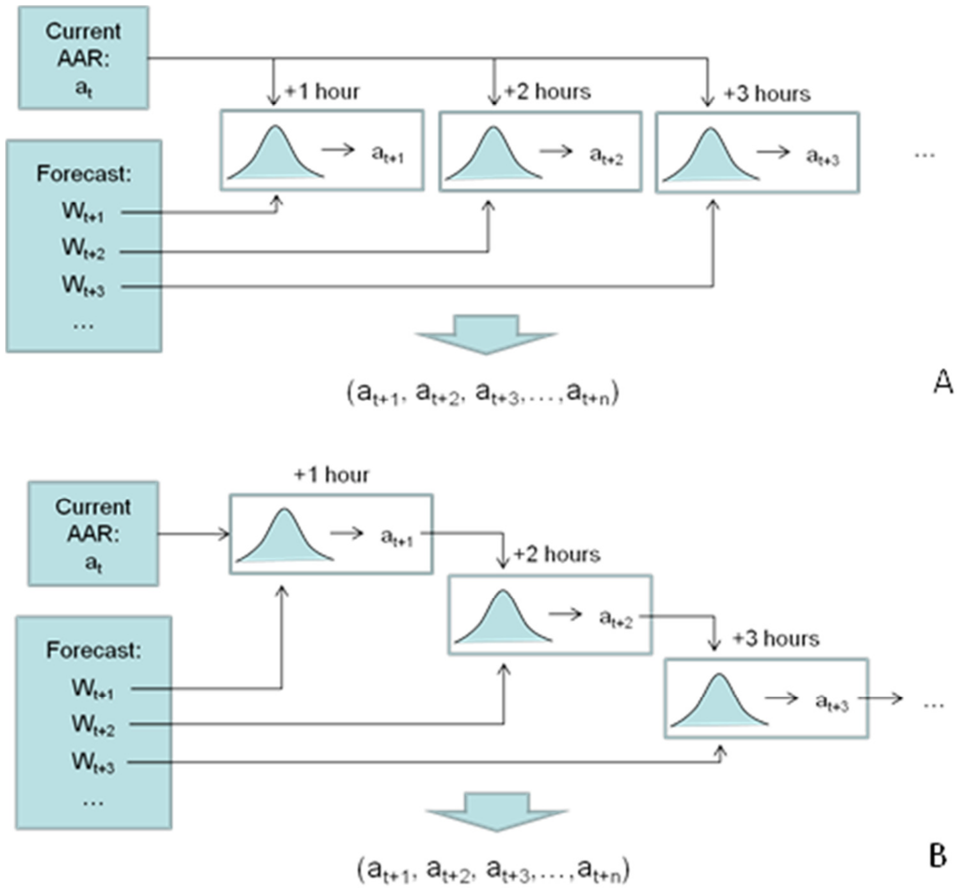

Figure 2. Static (A) and dynamic (B) sampling methodologies

of prediction. In contrast to the static mode, the dynamic mode accounts for the dependency of the AARs over time by conditioning the sampling distribution in each hour on the sample AAR selected in the previous hour in the same capacity scenario. For example, the sample AAR for hour one is selected from the distribution built around the AAR prediction made from the current capacity scenario of the airport and the valid weather forecast information. Thereafter, the resulting sample AAR from hour one is then used as the "lead AAR" for hour two to condition the ARR prediction and the resulting error distribution used to select the sample AAR. This process continues through the forecast horizon, usually 10 hours.

3. WEATHER FORECAST PRODUCTS

During this project, two forecast products distributed by the National Weather Service were utilized: the Localized Aviation Model Output Statistic (MOS) Product (LAMP) and the Terminal Aerodrome Forecast (TAF). These products were chosen based on a number of factors, most notably their public availability, popularity among aviation forecasters, controllers, and dispatchers, as well as their ability to forecast multiple sensible weather fields. As previously mentioned, data for only EWR and ORD between 2008 and 2010 was downloaded to train and test WTMG. The following subsections will give a brief overview of these weather forecast products and how they were processed to be used within the model.

3.1. Localized Aviation MOS Product (LAMP)

Developed by the National Weather Service's (NWS) Meteorological Development Laboratory (MDL), the Localized Aviation MOS Product issues hourly forecasts for 1,591 stations across the CONUS, Alaska, and Hawaii (Ghirardelli and Glahn 2010). Each forecast issued has a forecast horizon of 24 or 25 hours, depending on the run time of the product, with a temporal resolution of 1

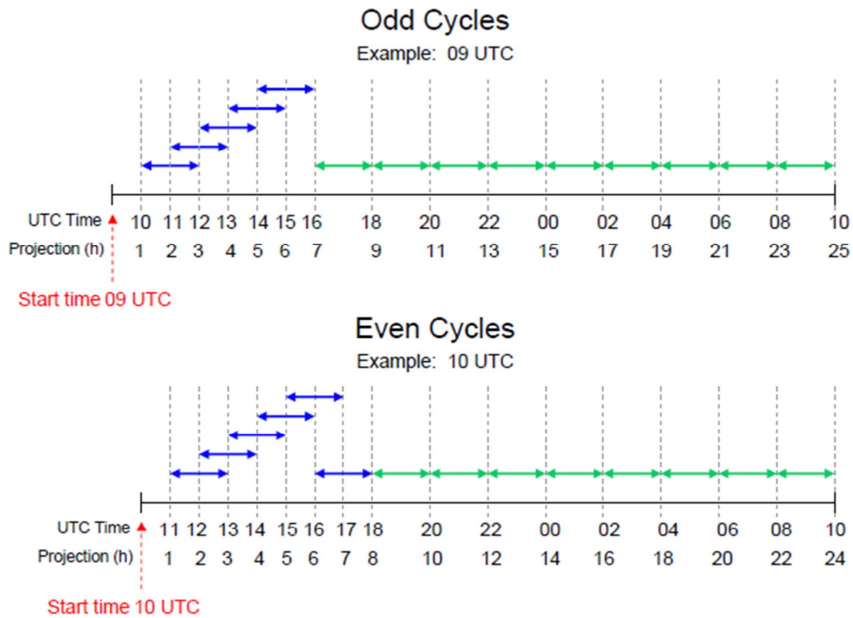

Figure 3. Schematic of LAMP Thunderstorm 2 hour valid periods (Image from *http://www.nws.noaa.gov/mdl/gfslamp/docs/Tstorm_proj_schematic.pdf)*

hour. While there are forecasts across the entire U.S., the forecasts themselves are only valid over the terminal area, or a five statute mile radius from the center of the airport runway complex. The LAMP fields used by WTMG include deterministic forecasts of temperature (TMP), dewpoint (DPT), wind direction (WDR), wind speed (WSP), wind gust (WGS), visibility (VIS), visibility conditional on precipitation occurring (CVS), ceiling height (CIG), ceiling height conditional on precipitation occurring (CCG), sky cover (CLD), obstruction to vision (OBV), and precipitation type (TYP). For precipitation, WTMG utilizes probabilistic forecasts of precipitation occurring on the hour (PPO), measurable precipitation in a six hour period (P06), the occurrence of freezing precipitation conditional on precipitation occurring (POZ), and the occurrence of snow conditional on precipitation occurring (POS).

Both the probabilistic (TP2) and the deterministic (TC2) thunderstorm forecasts with LAMP are used by WTMG. These forecasts predict the occurrence or non-occurrence of one or more cloud-to-ground lightning strikes in a 2 hour period in a 20 km grid box. The probabilistic field forecasts the probability of thunderstorms occurring the two hour period ending at an indicated time, while the deterministic field is a categorical forecast (yes or no) of thunderstorms occurring in the two hour period ending at an indicated time. The unique feature about both of these forecasts is that they are valid over a two hour period and in some cases there is an overlap of forecasts. A schematic of this valid time period is shown in Figure 3. In the figure, the blue arrow indicate overlapping two-hour valid periods out to 7 or 8 hours from the time issuance and the green arrow represent the subsequent two-hour valid periods, ending on even UTC hours. For the purposes of WTMG, a probabilistic and deterministic 1-hour thunderstorm forecast was derived by using the most recently issued forecast when there is an overlap. For example, based on the 09 UTC example in figure above, the WTMG uses the two-hour LAMP thunderstorm forecast from 12 UTC for the 11 UTC prediction time.

While the majority of the fields used from the LAMP were able to be directly imported into WTMG, a few fields required additional processing. The LAMP sky cover, obstruction to vision, and precipitation type fields were converted into numerical values as the model was unable to process categories in their original format. For example, instead of the clear (CL), few (FW), scattered (SC), broken (BK), and overcast (OV) sky cover categories used in the LAMP, they were simply converted to a 1, 2, 3, 4, or 5, respectively.

The wind fields in the LAMP also required additional processing. In order to provide WTMG with a measure of severity for these fields, they were organized into categories (see Table 1).

Table 1. Wind Speed and Gust categories

Group Number	*Wind Speed Groups (kts)*	*Wind Gust Groups (kts)*
1	0 - 5	0 - 13
2	6 - 7	14 - 19
3	8 - 10	20 - 24
4	11 - 14	>= 25
5	15 - 17	
6	>= 18	

These categories were subjectively chosen based on the distribution of all the wind speed and wind gust observations within the three years of data used by WTMG. Similar to the other processed LAMP fields above, the categories provide the model with a measure of severity for the wind fields forecasted.

3.2. Terminal Aerodrome Forecast (TAF)

The Terminal Aerodrome Forecast (TAF) is a concise operational forecast product consisting of the expected meteorological conditions significant to aviation at an airport for a specified period of time (Caldwell 2008). Scheduled TAFs are prepared and issued by the National Weather Service every six hours, four times per day at 00, 06, 12, and 18 UTC and is valid for 24 or 30 hours. However, unscheduled TAFs or amendments are issued if and when a forecaster recognizes significant difference between forecasted and observed conditions that would lead to a change in flight category value. For the purposes of the WTMG, deterministic forecasts of wind direction, wind speed, wind gust, ceiling, visibility, and significant weather are used. Similar to the LAMP, the spatial resolution of the TAF is the terminal area, which is defined as a 5 statue mile radius of the center of the airport's runway complex.

Similar to the LAMP, a number of forecast fields within the TAF required additional processing before it was able to be fully utilized by WTMG. The ceiling and visibility fields were organized into the same categories used by the ceiling and visibility fields found in the LAMP. This conversion was done by simply comparing the forecasted field in the TAF to the category definitions of these fields in the LAMP and then assigning it the appropriate category number. The LAMP categories for ceiling and visibility are shown in Table 2.

Table 2. LAMP Ceiling and Visibility categories

Category Number	*Ceiling Groups*	*Visibility Groups*
1	< 200 ft	< 1/2 mile
2	200-400 ft	1/2- < 1 miles
3	500-900 ft	1- < 2 miles
4	1000-1900 ft	2- < 3 miles
5	2000-3000 ft	3-5 miles
6	3100-6500 ft	6 miles
7	6600-12,000 ft	> 6 miles
8	> 12,000 ft or unlimited	

Additionally, the significant weather forecast within the TAF was processed in order to summarize the forecasted weather into a singular event. To do this, a subjective order of severity was developed (see Table 3) based on the impact these weather events would have on an airport. For example, a frequently used TAF significant weather forecast in the spring and summer months is "TSRA BR," which is translated to thunderstorm with moderate rain, and mist. WTMG is unable to process each one of these weather events together, so it is summarized into a single event by choosing the event of maximum severity in the table below, which in this case is thunderstorm.

Table 3. Order of severity for TAF significant weather field

Order of Severity	
1 - Tornado	9 - Ice Pellets
2 - Thunderstorm	10 - Rain
3 - Squalls	11 - Fog
4 - Blowing Snow	12 - Drizzle
5 - Snow	13 - Mist
6 - Freezing Rain	14 - Smoke
7 - Freezing Drizzle	15 - Haze
8 - Freezing Fog	

Lastly, similar to the forecast fields in the LAMP, the wind speed and wind gust forecasts in the TAF are organized into categories. The categories used are the same ones used by the LAMP, shown in Table 1.

4. RESULTS

Four different versions of the model were built and tested based on the two different modes of the model and forecast products. The different versions are: LAMP/Static, LAMP/Dynamic, TAF/Static, and TAF/Dynamic. Each version of the model contains a unique set of regression trees for each lead time from 1 to 10 hours into the future. For each model instance the two parts of the model, prediction and sampling, were evaluated for their accuracy and robustness, respectively.

For the prediction model, the primary metric used to evaluate the predictions is the root mean squared error (RMSE) between the predicted AAR and the actual AAR. This metric was chosen as it places a greater weight on large outlier errors than a linear distance measure. The unit of RMSE is the same as the unit of the AAR, arrivals per hour.

In addition to the four instances of the model, two baseline RMSEs are used for comparison. The baseline RMSEs are the results of a simple naïve predictor that assumes a constant AAR, regardless of the weather. The "Baseline – Static" model uses the actual AAR at the time of the forecast as the predicted AAR for each hour of the forecast horizon. The "Baseline – Dynamic" model assumes that the AAR from the previous hour is always available and uses that for its prediction. Both of these baseline models represent the benefit that WTMG achieves in prediction accuracy by incorporating weather information.

The RMSE comparisons between the four model instances and the two baseline models are shown below.

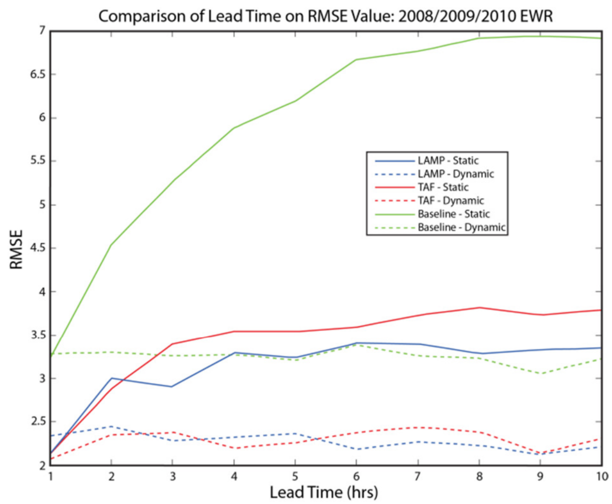

Figure 4. RMSE by lead time for EWR

In the figure, two distinct patterns between the two different modes of the model are easily discernible. First, the static WTMG and static baseline model (solid green, red, and blue lines) show an increase error at the longer lead times. This trend can be attributed to the fact that the lead AAR becomes an increasingly weaker predictor as the lead time approaches 10 hours. Second, the dynamic WTMG and dynamic baseline model (dashed green, red, and blue lines) show a relatively steady error for all 10 hours of lead time. This pattern is the result of the lead AAR not having as much of an impact on the overall error since the previous hour's AAR is always used as an indicator.

In terms of weather product comparison, the LAMP static models slightly, but consistently outperform the TAF static models for lead times above two hours. For the dynamic models, the performance of the LAMP and TAF versions are about the same as neither outdoes the other.

Perhaps the most notable pattern in the figure is the improvement in the RMSE of the WTMG predictions compared to the baseline RMSE prediction models. The static WTMG shows the greatest improvement as it produces increasingly better predictions over all 10 hours ranging from between 1 and 1.5 flights per hour at a one hour lead time to approximately 3.5 flights per hour for those lead times greater than 8 hours. On the other hand, while there is a noticeable improvement in RMSE between the dynamic WTMG and the dynamic baseline model, the overall improvement remains the same for the forecast horizon, with error reductions of approximately 40% for each lead time, relative to the baseline.

To evaluate the sampling model, WTMG was analyzed during time periods that GDPs were issued in response to weather at EWR. For each one of these times, the sampling model was run using the weather forecast and airport capacity at the time of issuance. For each GDP, 100 sample scenarios were generated with a forecast horizon of 10 hours. These predictions were then compared to the actual AARs to determine how well the sampling model captured the actual errors the AARs produced by the prediction model.

For this analysis, two methods were used: capture rate and cumulative capture rate. Given a value x between 0 and 100, an x-th percentile capture occurs in a particular instance if the actual hourly AAR falls inside the central x-th percentile of the sample AARs for that hour. The fraction of scenarios in which a capture occurs at a given lead time across all GDP instance represents the x-th percentile capture rate for that lead time. The cumulative capture rate tracks how well the uncertainty in the cumulative AAR up through a given lead time is modeled. This metric uses cumulative sample AARs and cumulative actual AARs to determine a capture at each lead time. For a perfect error distribution, the capture rate would match the percentiles. For example the 50^{th} percentile capture rate would be 0.5.

Figure 5, below, shows the hourly (left) and the cumulative (right) capture rates for the LAMP/Static version of WTMG for EWR. In the figure, four different percentiles are plotted: 30 (green), 50 (red), 80 (blue), and 90 (black). The hourly capture rate plot shows that the plotted percentiles are much higher than their target rates. This suggests that the sample distributions created by the sampling model allow for too much error around the hourly AARs produced by the prediction model. In the cumulative capture rate plot on the right, however, while the plotted

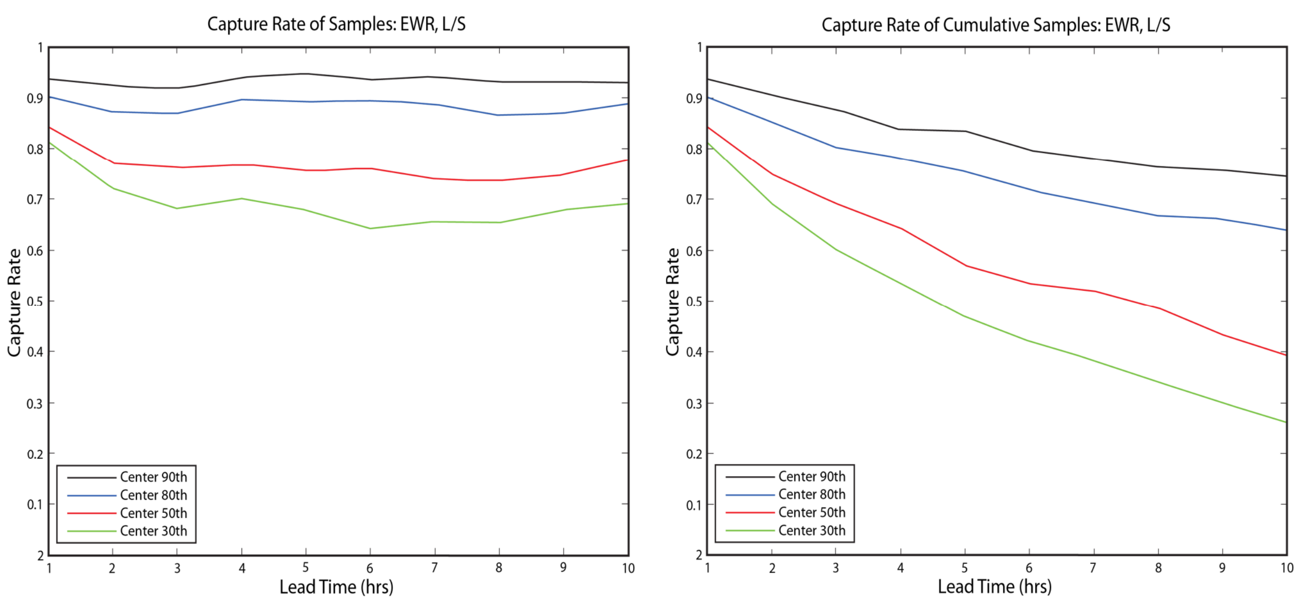

Figure 5. Hourly (left) and cumulative (right) capture rates for EWR

percentiles are high at the early lead times, the capture rates gradually decrease and eventually drop below the target values at 10 hours.

5. CONCLUSIONS, IMPROVEMENTS, & FURTURE WORK

The weather forecast products used in this effort provide WTMG with satisfactory information, but enhancements are necessary. Although the TAF is the preferred product in aviation weather, it is not detailed enough to adapt to fast-changing weather events. It does well as a general, strategic forecast product, but a higher resolution product is better for tactical decision making. On the other hand, the LAMP is a sufficient supplement to the TAF, particularly in the tactical time frame. With a high temporal resolution and a frequent update cycle, the LAMP provides the most up-to-date information when it is required. However, the lack of spatial resolution in the LAMP hinders its performance in the WTMG. Since the resolution of the LAMP is restricted to the terminal area, any weather occurring beyond the area is unable to be accounted for in the forecast.

Through this research, a number of important model improvements have been found to be necessary, especially for future TFM work:

(1) Improved probabilistic weather forecasts.
(2) Higher resolution thunderstorm forecasts both spatially and temporally.
(3) Advanced model physics and algorithms that can accurately parameterize mesoscale phenomena in and around the terminal area.
(4) Higher resolution forecasts in the terminal area.

Future work includes further refining the WTMG to effectively utilize current forecast products and the integration of the experimental gridded LAMP products (Ghirardelli 2011) as well as LAMP Convection (Charba et al. 2011). Both of these products may help provide the necessary weather information in the surrounding terminal area that can improve the airport arrival predictions made by WTMG.

6. ACKNOWLEDGEMENTS

The research described in this paper was funded by NASA Ames Research Center and FAA Systems Operations Planning and Procedures.

7. REFERENCES

Caldwell, D., 2008: Terminal Aerodrome Forecasts. [Available online at http://www.nws.noaa.gov/directives/010/pd01008013g.pdf].

Chan, W., S. Grabbe, L. Cook, C. Provan, and J. Evans, 2011: Integration of Weather into Air Traffic Management (ATM) Initiatives: Recommendation of Enhancement. FAA System Operations, Planning and Procedures, AJR-52.

Charba, J. P., F. G. Samplatsky, and P. E. Shafer, 2011: Experimental LAMP 2-H Convection Guidance Forecasts on a 20-km Grid. Preprints, *91th AMS Annual Meeting,* Seattle, WA, Amer. Meteor. Soc., J19.3

Ghirardelli, J. E., 2011: Gridded LAMP Pre-implementation Briefing. [Available online at http://www.nws.noaa.gov/mdl/lamp/presentations/GLMP_pre-impl_brief_NCEP_final.pptx].

Ghirardelli, J. E., and B. Glahn, 2010: The Meteorological Development Laboratory's Aviation Weather Prediction System. *Wea. Forecasting*, **25**, 1027–1051.

Provan, C. A., L. Cook, and J. Cunningham, 2011: A Probabilistic Airport Capacity Model for Improved Ground Delay Program Planning. *ATM Capacity Improvements*, 30th Digital Avionics Systems Conference, Seattle, WA.