4.2 EVALUATION OF AN AIRPORT CAPACITY PREDICTION MODEL FOR STRATEGIZING AIR TRAFFIC MANAGEMENT

Shin-Lai (Alex) Tien, The MITRE Corporation, McLean, VA
Sandip Roy, Washington State University, Pullman, WA
Christine Taylor, The MITRE Corporation, McLean, VA
Craig Wanke, The MITRE Corporation, McLean, VA
Rahul Dhal, Washington State University, Pullman, WA

1. INTRODUCTION

Strategic air traffic flow management (TFM) addresses predictions of significant capacity-demand imbalances four or more hours in the future, with the goal of mitigating anticipated delays while maintaining safe operations. In the U.S. National Airspace System (NAS), the current process of formulating a mitigation plan relies heavily on decision makers mentally translating weather forecasts into spatiotemporal impact on air traffic control (ATC) resources, i.e. ATC sectors and airports. In addition, at the strategic look-ahead horizon, the weather phenomena and their impacts are often subject to significant uncertainty, which may influence the design of traffic management plans. As such, impact forecasting in the TFM decision making process often suffers from experience-based subjectivity.

To address this shortfall, quantitative assessment of ATC impact from weather forecast variables is a vital step. Several methods have been proposed that estimate capacity loss of airspace and airports from numerical weather forecast variables, such as Tien14, Dhal13, Dhal14, Xue11, Steiner09, Song09.

The scope of this paper is limited to airport arrival capacity prediction. In the recent literature, a number of airport-capacity models for strategic TFM have been proposed. Buxi and Hansen (2011) generate probabilistic capacity profiles for a full day, by correlating current Terminal Aerodrome Forecasts (TAF) with historical forecasts. Provan et al (2011) and Kicinger et al (2012) find probability distributions for airport capacity given weather forecasts and the airport state (runway configurations, demand, operational standards, procedures, etc.). Also, there are models that leverage data-mining and statistical estimation approaches for predicting runway configuration or capacity (Dhal et al, 2013 Houston and Murphy (2012), Ramanujam and Balakrishnan (2011)). Alternatively, DeLaura et al (2014) developed an Airport Arrival Rate (AAR) prediction model that captures in detail the facility-specific runway selection rules and spacing issues that drive AAR. Although all the existing models study sample airports to support concept development and validation, requiring site-specific calibration or detailed adaptation makes these models less readily applicable to other NAS airports.

Dhal et al. (2014) proposes a model that determines the runway configuration and associated AAR and Airport Departure Rate (ADR) based on 1) weather forecast variables and 2) local preferences observed in the historical data on selecting configurations for given weather conditions. The model attempts to reflect the operators' decision making processes, rather than being entirely data-driven, so the model structure is flexible to changing operational realities (without the need to recalibrate any parameters). Historical data on configuration preferences, thresholds, etc. can be exploited when the inputs of local facilities are not available, to develop a model that can be systematically applied to key NAS airports. When inputs from local facilities are available, these inputs can be integrated in lieu of or in addition to historical preferences.

In this paper, we use empirical data to validate Dhal et al.'s model. We apply the model to FAA's 35 Operational Evolution Partnership (OEP) airports and compare the predicted and observed AAR using both observed weather and the Terminal Aerodrome Forecast (TAF). The differences in prediction performance between weather sources and among airports are examined. Performance under various factors, such as forecast look-ahead time and adverse weather conditions, are also analyzed. The analysis provides insight into outstanding needs in weather forecasting and impact prediction.

2. MODEL DESCRIPTION

Dhal et al.'s model of predicting runway configuration is a mapping from a given combination of weather phenomena – the meteorological condition (MC), wind speed and wind direction – to a set of possible runway configurations. The mapping is deterministic in nature. Each mapped runway configuration is operationally eligible, and there is a set of capacity values – AAR and ADR – associated with such a configuration, derived from either data-mining or local facility inputs. A configuration that is most likely to occur is then predicted from all the eligible ones.

Figure 1 depicts the analytical steps involved in preparing and running the model. The first step is "Identify Configuration Preferences", which categorizes the historically used configurations by MC, frequency, etc. Although not considered in this paper, further categorization could be based on local hours, arrival/departure banks, or seasons. The output of this step would be a summary table like that shown in Figure 2, which lists the observed configurations, the associated AAR and ADR, and the usage frequency. This summarized preference table can be either obtained from historical observations or from local facility inputs.

The second step is to determine eligible configurations based on the forecast variables of the hour (ceiling, visibility, wind speed, and wind direction). While ceiling and visibility are used to determine the MC category, wind spend and direction are used to determine the eligible configurations that meet the tailwind and crosswind requirements.

The third step is to select one configuration from the eligible ones which is most likely to occur. Several attributes are considered in the selection process, including wind preference, capacity, and the previous hour's predicted configuration. Dhal et al (2014) has more details of the selection process.

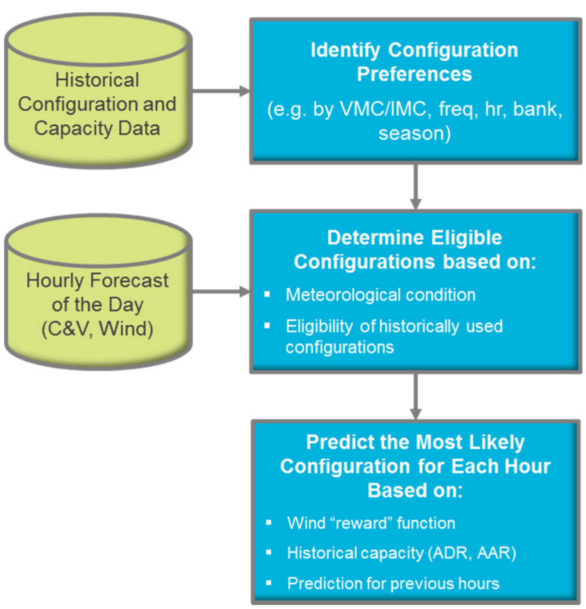

Figure 1. Illustration of Algorithmic Steps in Dhal et al.'s Runway Configuration Prediction Model

Airport	MC	Config	Freq	ADR	AAR
BOS	V	22L 27 | 22L 22R	27%	12	15
BOS	V	4L 4R | 4L 4R 9	24%	14	15
BOS	V	22L 27 | 22R	7%	12	14
BOS	V	27 32 | 33L	5%	9	11
…	…	…	…	…	…
BOS	I	4R | 4R 9	28%	11	8
BOS	I	4R | 4L 4R 9	20%	13	8
BOS	I	22L 27 | 22L 22R	11%	13	14
BOS	I	22L | 22R	9%	10	8
…	…	…	…	…	…

Figure 2. Sample of the Observed Configuration Preferences

3. DATA PREPARATION

The FAA's ASPM database contains quarter-hourly airport operational data (runway configuration, AAR, ADR) and recorded weather variables (meteorological condition, ceiling, visibility, wind speed, wind direction).

Configuration preferences of the OEP35 airports as explained in the first step of Fig 1 were derived from ASPM using data from April to September 2013 in order to capture sufficient variation in historically used configurations. We group airport records by MC and configuration and then calculate the frequency and the average AAR, generating a preference table for each airport, similar to the one shown in Figure 2. Note that this data-mining step is applied systematically to each OEP35 airport and does not take into account the limitations unrevealed in ASPM, i.e. taxiway/runway outage or dependencies among neighboring airports. As a future research effort, local inputs and adaptation can be included to help improve revelation of configuration preferences.

The data for evaluation are collected for July 2014. For each airport, there are 2,976 quarter-hourly observations. Predictions are made using recorded weather from ASPM (described above), as well as archived historical TAFs. The recorded AARs are then compared with the predicted ones.

This paper uses a universal set of thresholds to determine configuration eligibility due to wind, even though the thresholds could be airport-dependent. For VMC, the tolerance is set as 5 knots for tailwind and 20 knots for crosswind; for IMC, 3 knots for tailwind and 15 knots for crosswind.

The determination of MC could also be airport-dependent. However, this paper assumes a set of universal values to distinguish VMC and IMC.

When using ASPM weather variables for prediction, MC is an input variable, i.e. no determination rule is applied. When using TAF data for prediction, MC is set to IMC when ceiling is below 1000 feet or visibility is below 3 statute miles; otherwise, it is set to VMC. Again, this threshold setting can be revised through adaptation effort for local facilities.

4. RESULTS AND INTERPRETATION

The airport capacity model's performance was evaluated for the OEP-35 airports for July 2014, using the described data. For the performance evaluation, hourly runway configurations and AAR were computed for the OEP-35 airports using the model, for both TAF data and ASPM recorded weather (which serves as a comparison baseline). For the purpose of performance evaluation, the quarter-hourly ASPM AAR data were summed to determine hourly AARs. Statistics of the error (difference) between the model predictions and the ASPM-recorded AARs were determined, for several slices of the data, including for each airport, as a function of look-ahead time. For each data slice, two statistics were computed:

Bias

$$\frac{\sum_{i=1}^{N} Predicted_i - Actual_i}{N}$$

Mean Absolute Error (MAE)

$$\frac{\sum_{i=1}^{N} |Predicted_i - Actual_i|}{N}$$

where $Predicted_i$ and $Actual_i$ are the predicted and recorded capacities, respectively, for each (hour) sample in the data slice and N is the number of samples. The bias indicates whether there is a systematic offset in the model predictions compared to the recorded AARs, while the MAE gives an indication of the extent of error in the rate prediction.

4.1. Model Performance by Airport

The bias and MAE in the model predictions for each airport are shown in Figures 3 and 4, for the case that ASPM recorded weather is used for prediction. In Figures 5 and 6, the recorded-weather-based predictions are compared with TAF-based predictions (where predictions up to a 24 hour look-ahead time are averaged). Figure 7 shows percentage mean absolute error, rather than the absolute value, for the TAF-based predictions.

The statistical analysis shows that only a small number of airports exhibit a significant bias in model predictions, whether ASPM recorded weather or TAFs are used. Only 5 of the 35 airports show a bias of over 10 aircraft per hour (4 are positively biased, 1 is negatively biased) for the ASPM-based predictions, and only 6 show such a bias for the TAF-based predictions. The bias is less than 5 aircraft per aircraft for a significant majority of OEP-35 airports. Thus, for most airports, there is no systematic offset in predictions.

The MAE is also fairly small for most OEP-35 airports, whether ASPM recorded weather or TAFs are used. Only 7 of the 35 airports have an MAE of over 15 aircraft per hour, for both ASPM-based and TAF-based predictions. Almost all of the other airports have an MAE of less than 10 aircraft per hour. Further, the 7 airports with high MAE are precisely the airports with the largest prediction bias. Six of these seven airports also have the largest percentage MAE.

The performance of the model for each OEP-35 airport is very comparable, for TAF-based and ASPM-recorded-weather- based predictions. In particular, the difference in the bias and MAE for the two cases is at most 2 aircraft for almost every airport. Typically, TAF-based prediction incurs a higher error, but in the aggregate this additional error is rather small.

The data analysis indicates that the model performs well for a large majority of the airports, upon use of real forecast data for prediction (with look-ahead of up to 24 hours). Errors in weather forecasts do cause some degradation in the performance, but the impact of forecast error on the whole is small: the error is primarily caused by model inaccuracies, not by weather-forecasting inaccuracies. Furthermore, the few airports with significant prediction error also exhibit significant bias, and in fact the bias is the primary contributor to the MAE. This observation suggests that prediction errors primarily arise due to systematic offsets in AAR prediction at a small number of airports, perhaps due to airport-specific operations that are not represented in the model. For the seven airports with the largest errors, we further examine the operations to identify the main causes of bias and found that:

DFW, MCO, and LAX had ongoing construction in July 2014 limiting runway use, thus leading to an over-prediction of the AAR.

BOS: The most common configuration is being predicted accurately, but there is a discrepancy between the nominal AAR for this configuration and the ASPM-reported AAR. This seems to be an ASPM reporting issue.

LGA, CLT: low capacity configurations are being used at night, which are not predicted correctly; this is not a significant problem for TFM, since congestion is rare during nighttime hours. Excluding the nighttime hours, the model for CLT has Bias=0.83 and MAE=8.5; while LGA has Bias=1.6, MAE=5.0.

CVG: A runway configuration not observed in the 2013 data was frequently used in July 2014, thus causing prediction errors.

We note that the biases in model prediction arise at these airports from identifiable causes and therefore are either not significant for strategic TFM (e.g., nighttime mis-prediction), or can be corrected by allowing input from control-tower personnel or managers. For instance, the configuration-preference table could be updated by operators to reflect changes in allowed configurations due to construction or due to altered runway-selection procedures.

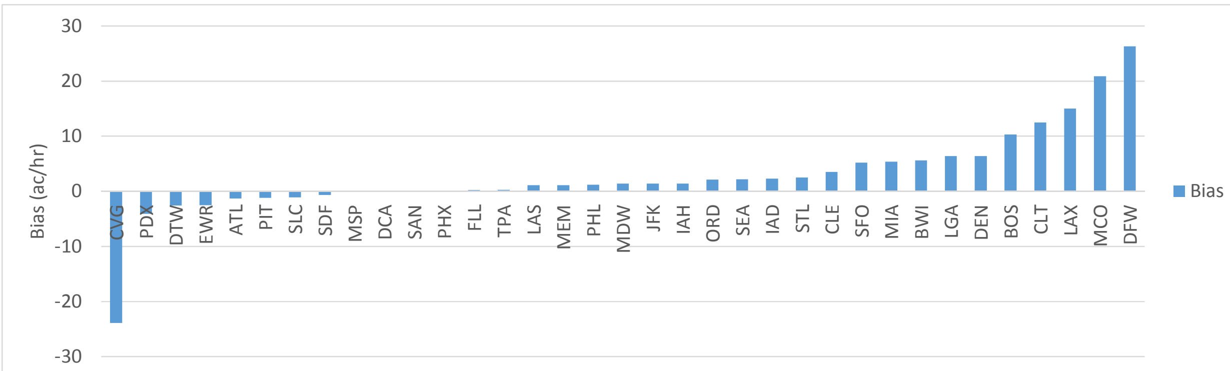

Figure 3: Model bias for each airport, when recorded weather from ASPM is used.

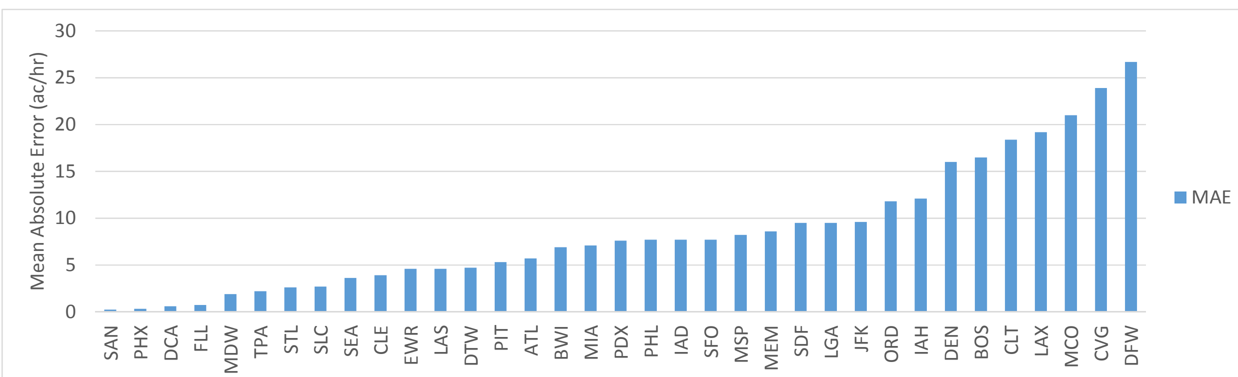

Figure 4. MAE in model predictions for each airport, when recorded weather data from ASPM is used.

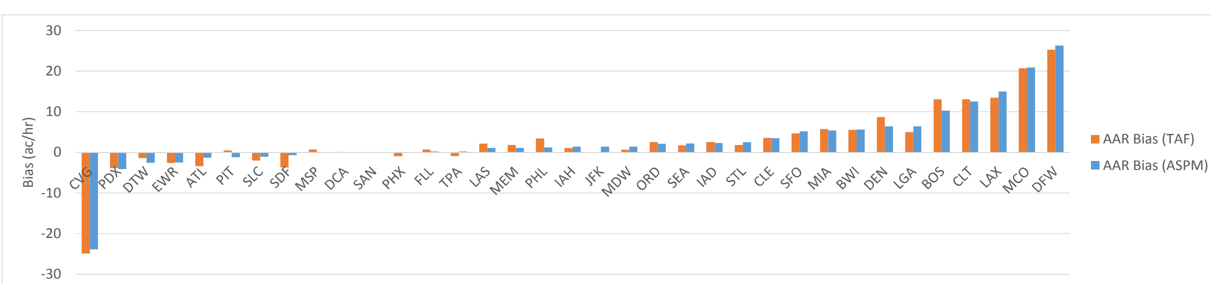

Figure 5. Bias in model predictions by airport, comparison of TAF and ASPM based predictions.

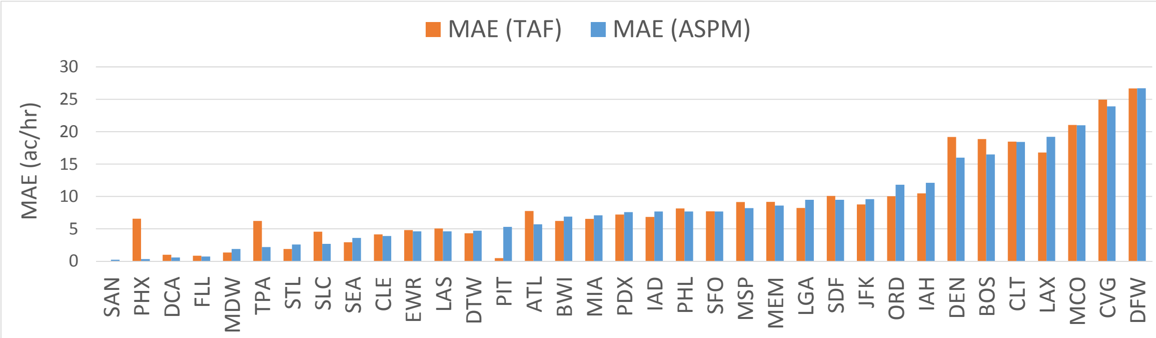

Figure 6. MAE in model predictions by airport, comparison of TAF and ASPM-based predictions.

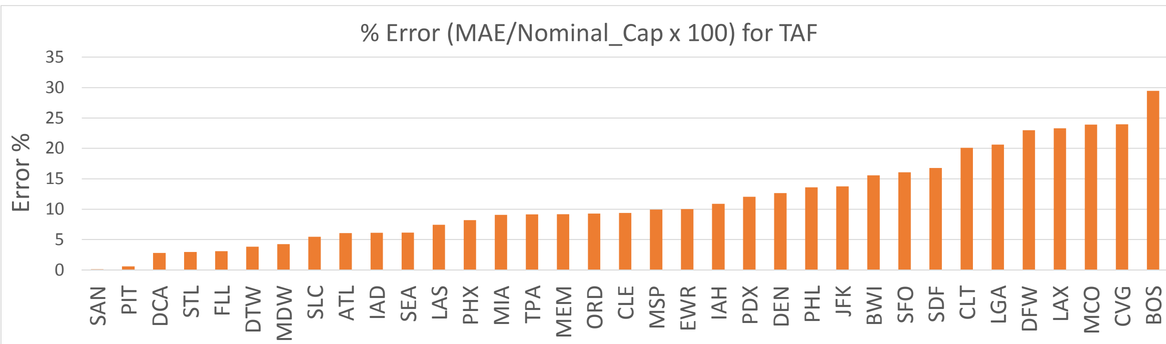

Figure 7. Percentage MAE by airport, TAF-based predictions.

4.2. Model Performance vs. Look-ahead Time

When forecasts such as the TAF are used for AAR prediction, it is natural to study whether the forecast look-ahead influences model performance and therefore we explore the MAE as a function of look-ahead time (aggregated across all airports). Specifically, the prediction MAEs for look-ahead times in 3 hour bins are shown in Figure 8. The performance shows remarkably little dependence on the LAT, with nearly uniform performance for look-ahead times of less than 15 hours. A slight degradation in performance is seen after 15 hours, with a more significant degradation after 24 hours. Based this analysis, the TAF appears to provide sufficiently accurate forecast data of wind and ceilings over a 24 hour look-ahead, to permit effective AAR prediction. However, it is worth cautioning that this analysis is aggregated over wind/MC conditions: the forecasts and hence the model may show a more pronounced dependence on look-ahead time during poor weather conditions.

Comparing the daytime and overall performance shown in Figure 8, the model also exhibits better performance during day-time hours, with a 40% improvement in performance during 6AM-11PM as compared to the nighttime hours. The performance degradation during the nighttime reflects the use of sub-optimal configurations (lower-capacity configurations), due to reduced demand pressure. Because of low demand, congestion is not a concern during nighttime hours, so the poorer performance during nighttime hours is not a significant concern for strategic TFM.

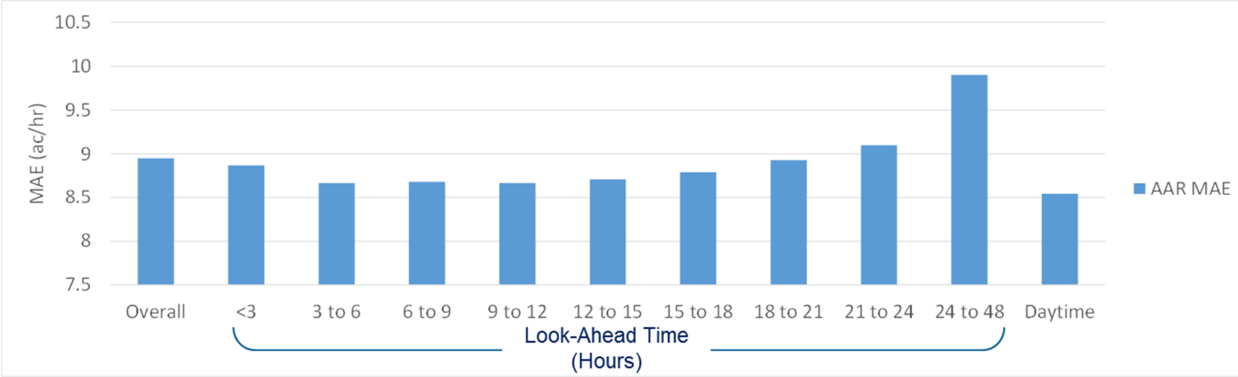

Figure 8. Dependence of MAE on look-ahead time (using TAF).

5. CONCLUDING REMARKS

Arrival capacity prediction of Dhal et al.'s model is "operationally-structured" and is promising for quick application to the OEP35 airports. Observed weather and TAF are applied for performance validation. It is found that model performance using TAF is comparable to performance using recorded weather.

To capture enough variation in runway configurations, data from summer months in 2013 are used to derive configuration preferences. For airports that demonstrate a consistent trend in configuration selection, the prediction model performs well. On the other hand, airports undergoing recent changes (i.e. runway/taxiway closure, etc.) have predicted AARs that have a larger error. More recent data could be included in the generation process of configuration preferences in order to improve model performance.

Ongoing research is focused on performance improvement. The inputs of local facilities could be included in the modeling process to address the account for events that are not observable in the historical data. Also, the model could be assessed with other forecast products such Short-Range Ensemble Forecast (SREF) to capture the impact of weather uncertainty on airport capacity prediction.

6. REFERENCES

[Buxi11] Buxi, G., and Hansen, M., Generating probabilistic capacity profiles from weather forecast: a design-of-experiment approach, The 9th U.S.A./Europe ATM R&D Seminar, Berlin, Germany, 2011.

[Kicinger12] Kicinger, R., Krozel, J. A., and Steiner, M., Airport capacity prediction incorporating ensemble weather forecasts, AIAA Guidance, Navigation, Control Conference, Los Angeles, CA, June, 2012.

[Dhal13] Dhal, R., Roy, S., Taylor, C., and Wanke, C., Forecasting Weather-Impacted Airport Capacities for Flow Contingency Management: Advanced Methods and Integration, 2013 Aviation Technology, Integration, and Operations Conference, American Institute of Aeronautics and Astronautics, 2013.

[Houston12] Houston, S., and Murphy, D., Predicting Runway Configurations at Airports, Proceedings of the Annual Meeting of Transportation Research Board, Washington D.C., 2012.

[Ramanujam11] Ramanujam, V.; Balakrishnan, H., Estimation of Maximum-Likelihood Discrete-Choice Models of the Runway Configuration Selection Process, Proceedings of the American Control Conference (AAC 2011), San Francisco, CA.

[DeLaura14] DeLaura, R. A., Ferris, R. F., Robasky, F. M., Troxel, S. W., Underhill, N. K., Initial Assessment of Wind Forecasts for Airport Acceptance Rate (AAR) and Ground Delay Program (GDP) Planning, Project Report ATC-414, MIT Lincoln Laboratory, Lexington, MA, Jan 2014.

[Tien14] Tien, S., Taylor, C., Wanke, C., Comparing and Clustering Ensemble Forecast Members to Support Strategic Planning in Air Traffic Flow Management, AMS 94th Annual Meeting, Atlanta, GA, February 2014.

[Dhal14] Dhal, R., Roy, S., Taylor, C., Tien, S., Wanke, C., An Operations-Structured Model for Strategic Prediction of Airport Arrival Rate and

Departure Rate Futures, AIAA Aviation 2014, Atlanta, GA, June 2014.

[Song09] Song, L., Greenbaum, D., and Wanke, C., The Impact of Severe Weather on Sector Capacity, The Eighth USA/Europe Air Traffic Management Research and Development Seminar, Napa Valley, CA, 2009.

[Xue11] Xue, Mengran, S. M. Zobell, Sandip Roy, Christine Taylor, Yan Wan, and Craig Wanke. "Using stochastic, dynamic weather-impact models in strategic traffic flow management." In *Proceedings of the 91st American Meteorological Society Annual Meeting. Seattle, WA*. 2011.

[Steiner09] Steiner, Matthias, and J. Krozel. "Translation of ensemble-based weather forecasts into probabilistic air traffic capacity impact." In *Digital Avionics Systems Conference, 2009. DASC'09. IEEE/AIAA 28th*, pp. 2-D. IEEE, 2009.

[Provan11] Provan, Christopher A., Lara Cook, and Jon Cunningham. "A probabilistic airport capacity model for improved ground delay program planning." *Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th*. IEEE, 2011.

ACKNOWLEDGEMENTS

The authors thank Jonathan Hammer, Mike Klinker, Dr. Lixia Song, Dan Greenbaum Joe Hollenberg of The MITRE Corporation for their valuable comments and support.

The authors would like to thank Tudor Masek, Rachel Ethier, John Huhn, Mary Hokit, Scott Mills, Matt Fronzak of The MITRE Corporation for their contributions in TFM concept and software development, and Dave Chaloux of VenCore for software integration.

The author Sandip Roy acknowledges the support of the MITRE/UNT subcontract as well as NSF Grant CNS-1035369.