Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)

Translating Convective Weather Forecasts into Strategic Traffic Management Decision Aids

Michael Matthews, Mark Veillette, Joseph Venuti, Richard DeLaura, and James Kuchar
Air Traffic Control Systems Group
MIT Lincoln Laboratory
Lexington, MA USA

***Abstract*— This paper presents a method to translate strategic convective weather forecasts into a metric that estimates the impact of convective weather on air traffic flows. The translation method is validated by measuring the flow rates of aircraft using weather impacted airspace in both en route and terminal airspace. Validation results show agreement between the airspace permeability estimates produced by the model and flow rates measured across airspace resources controlling arrival and departure flows. Features from single and ensemble storm-resolving forecasts, combined with two different probabilistic forecasts, were used to generate 0-12 hour estimates of airspace permeability including prediction intervals. The skill of the combined forecast and each contributing forecast was quantified across varying forecast horizons. The algorithms were implemented in a real-time system that was evaluated at several U.S. facilities between 2014 - 2016.**

Keywords–convective weather forecast, air traffic management, impact translation, forecast uncertainty, validation

I. INTRODUCTION

Convective weather accounts for the majority of the delay in the U.S. National Airspace System (NAS) [1]. To mitigate these delays, forecasts of convective weather are used by traffic flow managers to attempt to match traffic demand to capacity constraints of specific air traffic resources such as en route flows or departure fixes via a strategic management plan. Traffic demand for impacted resources is managed through the application of Traffic Management Initiatives (TMI) that either completely remove demand from an impacted airspace resource or that reduce demand by delaying the departure of flights filed through the impacted airspace. Typical strategic TMI programs used by air traffic managers include mandatory playbook reroutes, Ground Delay Programs (GDP), and Airspace Flow Programs (AFP). Since these TMIs require the pre-departure management of demand, the lead time for such decisions may be several hours in advance of the event onset to ensure that the TMI is in place soon enough to capture demand prior to departure. This also allows airline operators to plan for the schedule and fueling consequences of the TMI.

Several weather-only convective forecasts are currently available to the traffic planner in the strategic time domain including the Consolidated Storm Prediction for Aviation (CoSPA) [2], Short Range Ensemble Forecast (SREF) [3], and Collaborative Convective Forecast Product (CCFP) [4]. However, these forecasts do not provide direct guidance about the aviation impact on air traffic resources. The precise location, severity, scale, and timing of operationally-significant storms and the human response to those storms can be notoriously difficult to predict. Therefore, the decision maker is left to make critical TMI decisions based on a subjective assessment of potentially conflicting weather forecast information.

The lack of an explicit translation of weather forecasts into resource constraints is a shortfall in the current weather information available to air traffic managers for strategic traffic flow management. There are several consequences of this shortfall. First, without an explicit translation there is a lack of an operationally-relevant methodology to assess weather forecast resource impact and overall forecast performance. Each participant (e.g., Air Traffic Control System Command Center (ATCSCC), Air Routes Traffic Control Center (ARTCC) Traffic Manager Unit (TMU) and Airline Operations Center (AOC)) comes into the collaborative strategic planning process with their own set of operational objectives, favorite forecast information, risk tolerance, etc. This wide and often divergent range of opinions and goals must somehow be melded into a plan of action. Without shared objective forecasts of weather impacts and estimates of decision risk, there is little common ground on which to base discussions about the best plan of action that addresses the different legitimate concerns of stakeholders. Second, the utility of convective weather forecasts is directly related to the quality of decisions and NAS performance outcomes that the forecasts can support. The definition of explicit, validated weather translations provides an objective and operationally relevant measure of truth against which forecasts can be compared. Without translation-based forecast evaluations, it is difficult to determine how much of an operational shortfall in convective weather mitigation is due to poor weather forecasts and how much is the result of poor interpretation and application of forecast information.

Previous efforts to estimate convective weather impacts have focused either on individual ATC sectors [5] or sector-traversing flows [6]. Such resources are important to tactical operations, as traffic managers seek to avoid sector overloads that can result in sector closures and excessive airborne holding. However, sector-level impacts are a poor match for strategic planning. Strategic planners usually focus on key, large-scale traffic flows that traverse congested en route

airspace or that carry traffic to or from transition airspace for busy metroplexes. Furthermore, the precision of convective weather forecast needed to estimate sector capacities is unachievable in the strategic planning time horizon.

This paper describes a weather translation algorithm that has three main components. First, a given weather situation is mapped into a 0%-100% measure of permeability for defined airspace regions and traffic flow directions. Second, based on analysis of historical flight patterns, permeability is related to maximum achievable and sustainable traffic flow rates and transit times for that airspace. Third, multiple heterogeneous weather forecast products are assimilated to generate a prediction of permeability and flow rate from 0-12 hours in the future. This prediction also includes a measure of uncertainty based on the real-time quality and variability of the weather products contributing to the prediction. The result is a timeline of forecasted airspace impact including uncertainty bounds. The algorithm performing these steps was implemented in a real-time decision support tool that was evaluated in the field during the summers of 2014, 2015, and 2016.

II. Model Overview

The translation of convective weather forecasts into an airspace impact classification began with a translation of weather truth data into an estimate of airspace permeability. The airspace permeability was then validated with an observed real-time operational flow rate that was assumed to represent, to first order, the operational impact of convective weather on the air traffic operations.

The translation model is based upon the Weather Avoidance Fields (WAF) developed as part of the Convective Weather Avoidance Model (CWAM) [8], the definition of airspace resources that are operationally significant and whose capacities are measureable, and the assessment of operational impact of weather on a trajectory. The initial development of the model focused on traffic flow through enroute airspace, with an emphasis on the region where major arrival and departure routes transition between the Cleveland ARTCC (ZOB) and the New York ARTCC (ZNY).

An example of the airspace resource definition for this enroute traffic flow is shown in Fig. 1. The resource definition consisted of three components: airspace crossing; airspace boundary; and airspace traversing trajectories, all of which define a strategic flow through the airspace. The airspace crossing represents an imaginary line for which all aircraft in the strategic flow that traverse this resource will intersect. The airspace boundary represents the region for which the model will evaluate the weather characteristics to estimate the permeability. Finally, the airspace traversing trajectories represent notional routes perpendicular to the aircraft crossing that are possible trajectories through the weather.

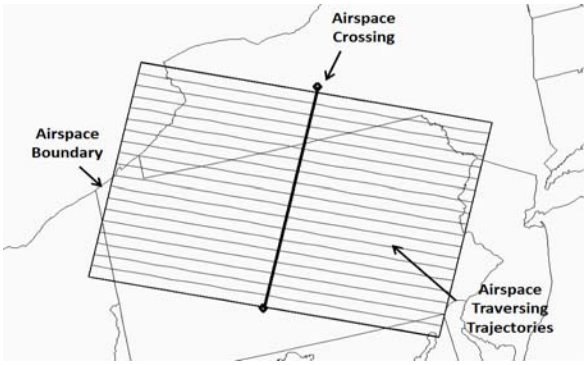

Figure 1. Example enroute airspace resource definitions.

The method of assessing the impact of the weather on a trajectory takes into account the scale and severity of storms that impact the flight trajectory. Storm scale is represented by the length of time that a trajectory spends inside a Convective Weather Avoidance Polygon [9]. Severity is represented by maximum blockage [10] calculated along the trajectory. Each notional route is then assigned an impact of RED (impassable), YELLOW (uncertain), DARK GREEN (passable with acceptable storm-avoiding deviations), or GREEN (passable) based on the two-dimensional heuristic trajectory impact model as shown in Fig. 2.

Finally, the permeability of the airspace, or the availability of passable corridors that traverse the airspace, is estimated by taking a weighted average of the trajectory impacts for the notional routes that traverse the airspace. The four impact categories are weighted as follows: GREEN routes are weighted by 1.0 (100% probability of successfully flying through), DARK GREEN by 0.8 (20% impacted), YELLOW by 0.5 (50% impacted), and RED by 0.0 (route is completely blocked).

Recent efforts have focused on the modification of the model to terminal airspace with an emphasis on the O'Hare International Airport (ORD) operations. An example of the airspace resource definition for this region is shown in Fig. 3. For terminal airspace, the notional routes are no longer perpendicular but are orientated along radials extending out from the airport with a specified minimum and maximum distance. The airspace crossing line is the inner circle which will be used later for model validation.

The method of assessing the impact of the weather on the airspace still takes into account the scale and severity of storms but is based upon the terminal WAF [11] and related Convective Weather Avoidance Polygons. As in the enroute model, each notional route is assigned an impact category as defined in Fig. 2 and the permeability is computed by taking a weighted average of the trajectory impacts.

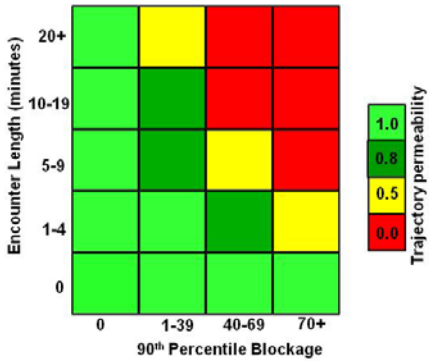

Figure 2. Trajectory impact model.

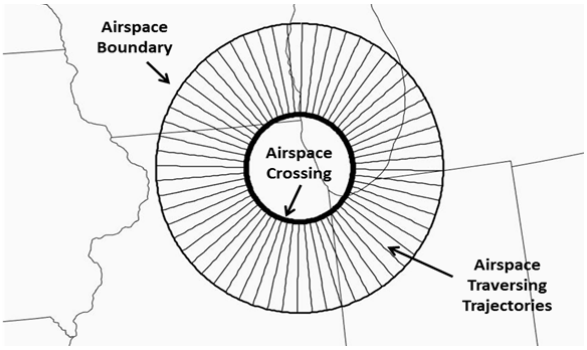

Figure 3. Example terminal airspace resource definitions.

III. Model Validation

In order to validate the model, it is necessary to demonstrate that current-weather permeability estimates correlate to observed operational behavior. High permeability values should support sustained, high traffic throughput, while low permeability values should represent highly constrained airspace. During times of high permeability, it is expected that there may be small, weather-avoiding deviations that will not significantly impact traffic flow rates. During times of low permeability, while it is possible there may still be some traffic passing through the airspace, the majority of traffic should be rerouted or delayed out of the impacted airspace.

An example plot of the enroute traffic flow rate and permeability is shown in Fig. 4 for 13 June 2014 in Cleveland Center (ZOB) / New York Center (ZNY) transition airspace. On this day, a cold front was oriented across the airspace with two lines of convective weather. As shown, the permeability drops starting at approximately 1600 UTC (noon local time). As the spatial extent and intensity of the weather grew, the observed flow rate and permeability estimate decreased synchronously until 1930 UTC. As the weather developed, the air traffic managers were able to reroute the excess demand north into the Boston ARTCC and south into the Washington, DC ARTCC. At 1930 UTC it is observed that the permeability estimate began to increase while the flow rate continued to drop, reaching a minimum of 11 aircraft per hr after 2100 UTC. This noticeable discrepancy between the flow rate and permeability is explained through an analysis of the weather and TMI program logs. As the weather moved eastward, it exited the ZOB/ZNY transition airspace, producing the observed increase in permeability. However, the weather then began to impact the NY metro airports, resulting in Ground Stop programs that held all flights heading into NY, thereby decreasing demand into the transition airspace even though adequate capacity was available in that region.

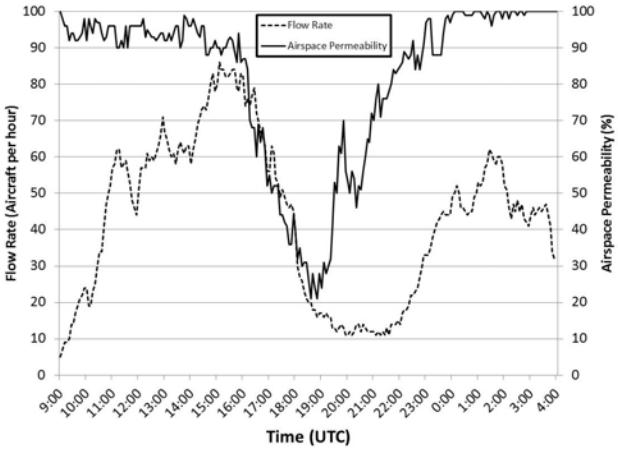

Figure 4. Observed flow rate and airspace permeability estimates.

A statistical validation of the enroute impact model was performed for the ZOB/ZNY transition airspace using data from 139 case days between 2011 and 2015. The data set was filtered to only include hours between 18 UTC and 00 UTC when the airspace experiences the highest demand. Fig. 5 shows resultant permeability estimates binned into increments of 20%. A correlation between the airspace impact model and flow rate is clearly visible. As the convective impact increases (measured by a decreasing permeability), the flow rate decreases accordingly.

The observed distributions within each permeability band in Fig. 5 imply a distinction between *sustainable* and *achievable* flow rates. The core box-and-whisker groupings (between 10th and 90th percentiles) span weather-impacted flow rates that air traffic personnel sustained for long periods of time. In contrast, the highest-observed rates (between the 90th percentile whisker and the maximum-observed diamond), were achieved but not sustained beyond a few 5-min periods. From the data that were collected, the achievable high rates were generally observed immediately following the onset of weather events, during times of decreasing permeability, and only for short durations. Higher than expected or desired flow rates may be possible with additional workload and coordination burdens that may not be easily sustained for long durations.

Using the mapping of permeability to flow rate provided in Fig. 5, planners could create traffic management programs that are tailored for the specific scenario of the day. For instance,

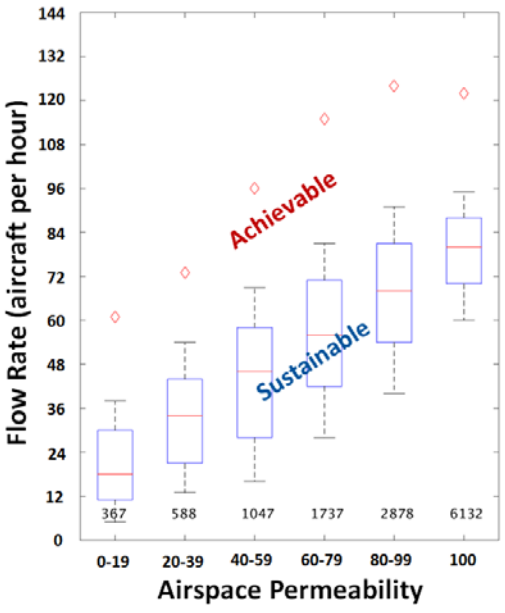

Figure 5. Observed flow rate for the ZOB/ZNY transition airspace.

for short-lived events, air traffic planners may choose to set flow rates near the maximum achievable rate for the permeability estimate forecasted. In this instance, planners may feel confident in the ability to push the workload of the air traffic controllers and sustain high flow rates due to the limited impact duration of the event. On another day, a planner may predict that the convective weather will be long-lived and that it will not be possible to set higher rates due to the difficultly in sustaining a high workload for a long period of time. In this scenario, planners may choose to set rates closer to the median flow rates observed for this airspace.

A statistical validation of the terminal impact model was performed for the O'Hare International Airport from the same 139 case days. The data set was filtered to only include the hours between 12UTC and 00UTC when the airport is under the highest demand. Fig. 6 shows the airport operations count vs. permeability estimates binned into increments of 20%. It is important to note that the y-axis represents the total number of airport operations per hour (both arrivals and departures). A correlation is clearly visible for the terminal impact model with airport operations decreasing as the permeability decreases.

IV. IMPACT FORECASTS

Section II described the translation of a given current-weather situation into a permeability metric. For operational use, it will be necessary to provide forecasts of permeability in the 0-12 hour timeframe. Additionally, the ability to observe the level of uncertainty in that forecast would aid decision-makers in judging the necessity and strength of any TMI responses. Accordingly, a permeability forecasting capability was developed that also conveys uncertainty information. Rather than focus on only a single forecast source, a focused effort was made to incorporate the best attributes from multiple, heterogeneous forecast products.

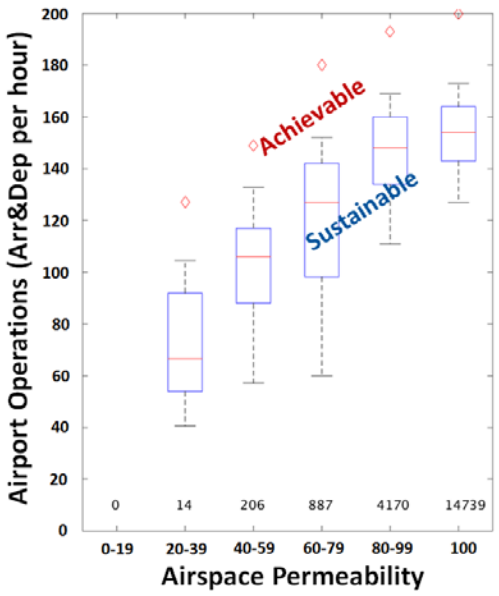

Figure 6. Observed airport operations for O'Hare International Airport.

Convective weather forecasts can generally be categorized as one of two types: storm-resolving or probabilistic. Storm-resolving forecasts provide a single depiction of two- or three-dimensional future weather events over a specified geographic area in a high resolution image. These forecasts are deterministic and attempt to represent the future state of actual meteorological quantities. As a result, storm-resolving forecasts by themselves contain little or no uncertainty information. In contrast, a probabilistic forecast assigns likelihood to a particular weather event (e.g., occurrence of thunderstorms) and displays the probability of this event at any given point and time. The method described below can be applied to any number of forecasts that fall into these categories. In this work, the following four forecast sources were utilized since they form the core set of products typically used in traffic management:

Extrapolation Forecasts: The Corridor Integrated Weather System (CIWS) and CoSPA extrapolation forecast is a storm-resolving forecast that applies motion tracking algorithms to national weather mosaics of precipitation intensity and echo tops [12]. These tracking vectors are used to advect current weather and predict storm location up to 8 hours into the future. The CIWS/CoSPA forecast has 1 km horizontal resolution and a new forecast is issued every 2.5 min. The major advantages of extrapolation forecasts are that they have low latency and typically outperform numerical model forecasts at short forecast lead times (1-2 hours). At later lead times however, forecast performance degrades.

High Resolution Rapid Refresh (HRRR): The HRRR model is a 3 km model that has been under development at the National Oceanic and Atmospheric Administration/Earth System Research Laboratory Global Systems Division since 2008 [13]. The HRRR is capable of providing storm-resolving forecasts of precipitation intensity and echo tops up to 18 hours into the future. This forecast is issued hourly, and typically exhibits 1 to 2 hours of latency. In contrast to the extrapolation forecast, numerical weather models like the HRRR typically perform worse in short lead times due to model spin-up which is often the effect of the model adapting to data assimilation used to initialize the model. The HRRR is currently running operationally at the National Centers for Environmental Prediction (NCEP).

Localized Aviation Model Output Statistics Program (LAMP): The LAMP model, developed by the Meteorological Development Laboratory of NOAA's National Weather Service provides probabilistic forecasts of thunderstorms by updating the Global Forecasting System's (GFS) Model Output Statistics (MOS) using observational data (METAR, lightning, radar), output from simple advective models, and geo-climatic data (high resolution topography and relative frequencies) [14]. Probabilistic forecasts of thunderstorms are issued hourly on a 20 km grid. Outputs are available in 1-2 hour intervals. The LAMP forecast is running operationally in the Advanced Weather Interactive Processing System (AWIPS) at Weather Forecast Offices.

Short Range Ensemble Forecast (SREF): The Storm Prediction Center's SREF Calibrated Thunderstorm Probability field is created by post-processing all 21 members of the

SREF along with a three-member time-lagged ensemble and the Weather Research and Forecasting - North American Mesoscale Model (WRF-NAM) [15]. These forecasts are combined to create a single probabilistic forecast of thunderstorms on a 32 km grid. This forecast is issued every 6 hours, and forecasts in hourly time intervals.

A. Forecast Model Development

To combine the different forecast models, we use a supervised machine learning algorithm that extracts features measured from each of these forecasts in an airspace region to predict the posterior distribution of permeability conditioned on these forecasts. This posterior distribution describes the range of possible permeabilities and their respective probabilities. Obtaining a full distribution (rather than a point prediction) allows us to compute uncertainty in our prediction and convey this in the form of a prediction interval.

For this methodology to be effective, a historical database of forecasts along with corresponding validation data must be available. In this study, a two-year history (2013 and 2014) of data from summer months (May – September) was used for model training. The summer of 2015 was kept separate from training and used for validation.

To apply a machine learning methodology, each input forecast for an airspace region must be converted into a vector of features with fixed length. Because the forecasts are of different types (storm-resolving and probabilistic) different features are extracted depending on the type of forecast. For the storm-resolving forecasts, precipitation intensity and echo tops forecasts are combined to create WAF and CWAP forecasts. The WAF and CWAP are processed using the permeability algorithm (Section II) to compute a forecast of permeability for each region. Other features are also computed, such as maximum route WAF, mean storm size within the region, and mean encounter time along each notional trajectory.

In addition to using the latest available forecast, a four-member time-lagged ensemble of HRRR forecasts was used to create the same features as described above. A time-lagged ensemble of forecasts is one method of estimating uncertainty by looking at how the different forecasts evolve at a fixed valid time. Time-lagged ensembles that consistently forecast impacts on an area, in general, indicate a greater level of confidence than ensembles that demonstrate a significant amount of variability.

For probabilistic forecasts, it is not possible to explicitly create WAF or CWAP since these forecasts do not provide precipitation intensity and echo tops information. Instead, various statistical properties of the forecasts were gathered in the area surrounding each airspace region. These features include percentiles of convective probability (10^{th}, 25^{th}, 50^{th} , 75^{th}, 90^{th}) within the region, areal coverage of convective probability exceeding various thresholds (10%, 25%) and average convective probability within quadrants of the region. These features are used to convey the forecasted impact on the region, and are later translated into an estimate of permeability.

Features extracted from each input forecast are represented by vectors $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{N_f}$, where N_f is the number of forecasts (including time-lagged ensemble members), shown as the first stage in Fig. 7. Training data of the form $\left(\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \ldots, \boldsymbol{x}_i^{N_f}, p_i\right)$ are used, representing a set of forecast features all valid at the same time i, and p_i is the observed permeability (based on current-weather CIWS data) at that time. These data were gathered over the summers of 2013 and 2014 around 13 regions in the Northeast quadrant of the U.S. For each day, a new set of forecast features were measured at the top of each hour. For

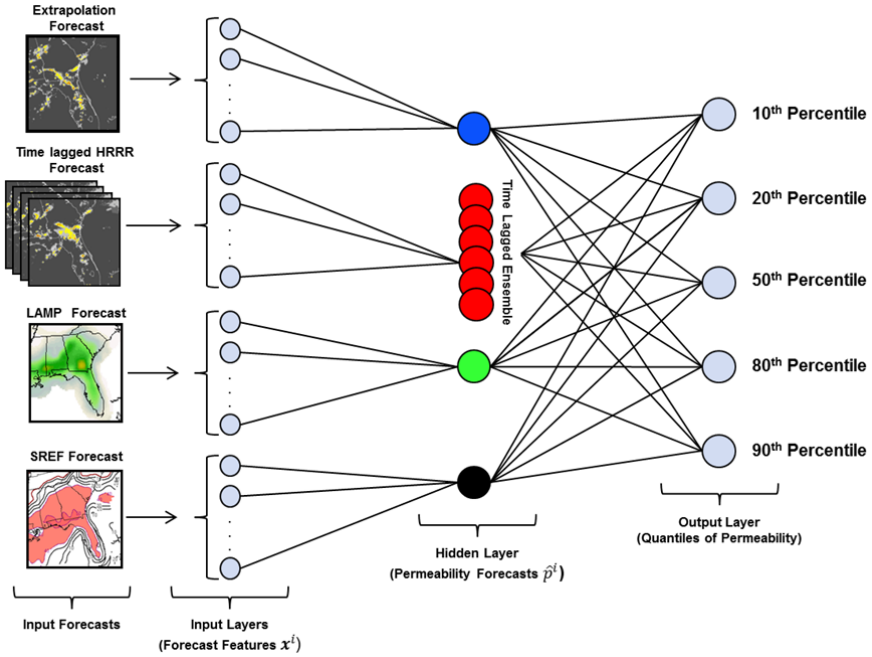

Figure 7. Permeability forecast model.

each of these times, features were gathered at hourly forecast leads out to 12 hours, with the exception of the extrapolation forecast which was only available out to 8 hours. Because forecasts suffer from latency, and because not all forecasts are generated hourly, the most recently-available forecast was chosen to create features for each target time.

The observed permeability, p_i, corresponding to each forecast valid time was measured using precipitation intensity and echo tops from CIWS. The technique outlined below was applied separately for each lead hour using as many of the forecast features as were available.

The first training step is to map features from each input forecast into an estimate of permeability. Prior to model fitting, all forecast features are scaled such that they have mean 0 and unit variance. Forecast features $\boldsymbol{x}^k$ are mapped to a forecast of permeability using a linear combination of the components of $\boldsymbol{x}^k$. The weights in this linear combination are found by fitting a Ridge Regression [16] separately for each forecast type. This involves finding the coefficient vector $\boldsymbol{\beta}^k$ and b_0 that minimize the following objective function:

$$(\boldsymbol{\beta}^k, b_0^k) = \text{argmin}_{(\boldsymbol{\beta},b_0)} \left(\sum_i \left(p_i - (\boldsymbol{\beta} \cdot \boldsymbol{x}_i^k + b_0) \right)^2 + \lambda_k |\boldsymbol{\beta}|^2 \right) \quad (1)$$

where λ_k is a regularization parameter chosen to minimize error on a holdout set chosen using a 10-fold cross validation [17]. As a final post-processing step, the outputs obtained from this step are rescaled so the output values $\{\hat{p}_i^k\}$ over the training set span the interval [0%, 100%]. In the following, $\hat{p}^k = \boldsymbol{\beta}^k \cdot \boldsymbol{x}^k + b_0^k$ represents the estimate of permeability obtained from forecast k.

Estimates of permeability derived from each forecast are then combined (middle of Fig. 7) to estimate the posterior distribution of observed permeability p conditioned on the vector of forecasted permeability $\hat{\boldsymbol{p}} = (\hat{p}^1, \hat{p}^2, \ldots, \hat{p}^{N_f})$, i.e. $F(p|\hat{p}^1, \hat{p}^2, \ldots, \hat{p}^{N_f})$, where F is the conditional Cumulative Distribution Function (CDF) of p. Quantile Regression was chosen as the method for estimating the conditional quantiles of p (these characterize the conditional CDF) [18]. Quantile Regression seeks to find a function $f_\tau(\hat{\boldsymbol{p}})$ that estimates the $100\tau\%$ percentile of the random variable p as a function of $\hat{\boldsymbol{p}}$ by minimizing the following functional over the training set

$$\{\hat{\boldsymbol{p}}_i, p_i)\}\colon f_\tau = \text{argmin}_{f\in H}(\textstyle\sum_i \rho_\tau(p_i - f(\hat{p}_i))). \quad (2)$$

Here, ρ_τ is the "check-mark" function defined as

$$\rho_\tau(t) = \begin{cases} \tau \cdot t, & t > 0 \\ -(1-\tau) \cdot t, & t \le 0 \end{cases} \quad (3)$$

The minimum above is taken over a suitable set of functions H. Typically H is chosen to be all linear functions of the components of $\hat{p}$, however here we introduce non-linear interaction terms between the different forecasts into the model by using an H that covers all second order polynomials of the inputs $\hat{\boldsymbol{p}}$, i.e.

$$H = \left\{ f : f(\boldsymbol{p}) = c_0 + \textstyle\sum_{i=1}^{N_f} c_i p^i + \sum_{i=i,j=i}^{N_f} d_{i,j} p^i p^j \right\}. \quad (4)$$

The optimal values of the coefficients c_i and $d_{i,j}$ for the forecast models considered here was found using the interior point (Frisch-Newton) algorithm [19].

The two steps described here – the translation of weather features into permeability estimates followed by a combination of permeability forecasts into conditional quantiles – can also be viewed as a feed-forward neural network with multiple input layers (one for each of the N_f forecasts) feeding one hidden layer that contains the vector $(\hat{p}^1, \hat{p}^2, \ldots, \hat{p}^{N_f})$. The output layer of the network contains quantiles of the distribution of permeability as shown in Fig. 7.

An example permeability forecast made using this methodology is shown in Fig. 8. The individual model inputs are shown in the top plot and display an overall drop in permeability across the airspace region. The bottom plot shows the resultant combined permeability forecast. The center line represents the median of expected permeability; the upper and lower curves that make up the prediction intervals are given by the 20th and 80th percentiles, respectively. With these choices of quantiles, we expect the prediction intervals to capture at least 60% of all observations. This target for prediction interval accuracy is important when performing validation. The permeability forecast model was assessed using data from the summer of 2015 (recall training data came from the summers of 2013 and 2014). We begin by examining a case study, and then will provide performance statistics for the entire summer.

Fig. 9 shows a sample of verification for a storm impacting New York airspace on 14 July 2015. This plot shows the sequence of 1, 4, and 8 hour permeability forecasts (including uncertainty bounds) made over a period of 60 hours (updating every 15 min). The x-axis represents the valid time of each forecast. The thick black curve in each plot shows the observed permeability measured at each time. Not surprisingly, the 1 hour forecast (top) shows greater skill and thinner prediction intervals than seen in the 4 hour and 8 hour forecasts. Prediction intervals generally become wider when the forecast intensifies (i.e., permeability drops into yellow and red regions). The main drop in permeability at roughly 2100 UTC was detected well across lead times, and the transition was mostly captured by the prediction intervals at 4 hour and 8 hour leads. At approximately 1600 UTC, the 8 hour forecast was late in detecting a drop in permeability, but quickly caught up when storms began to intensify. Both the 4 hour and 8 hour forecasts decayed the storms too quickly, as both had permeability returning to 100% earlier than what was observed.

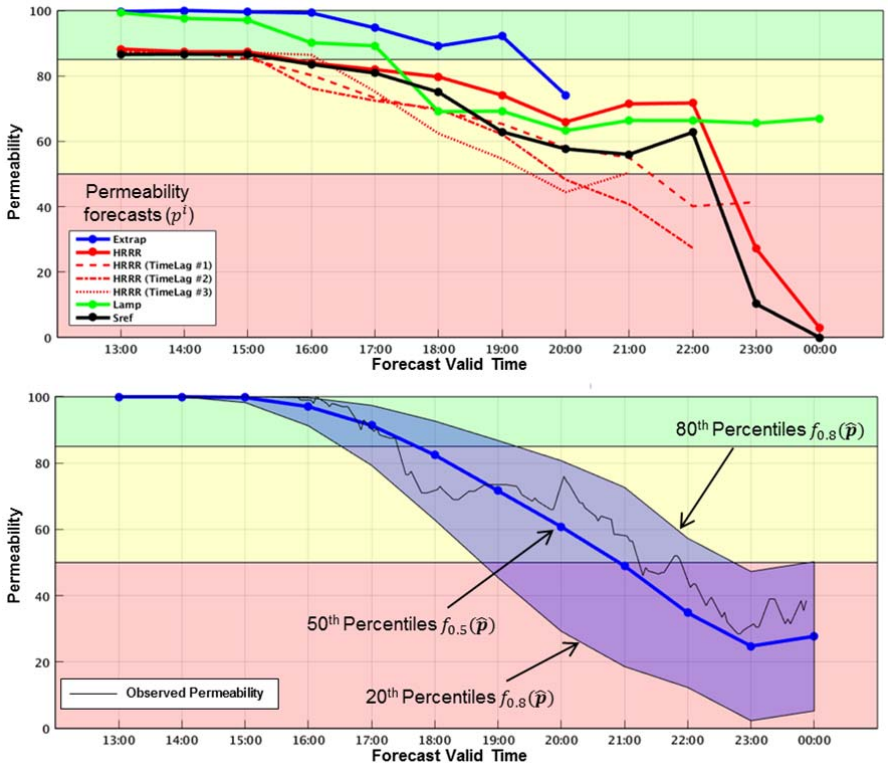

Figure 8. Example permeability verification for 1 hour (top), 4 hour (center), and 8 hour (bottom) forecasts.

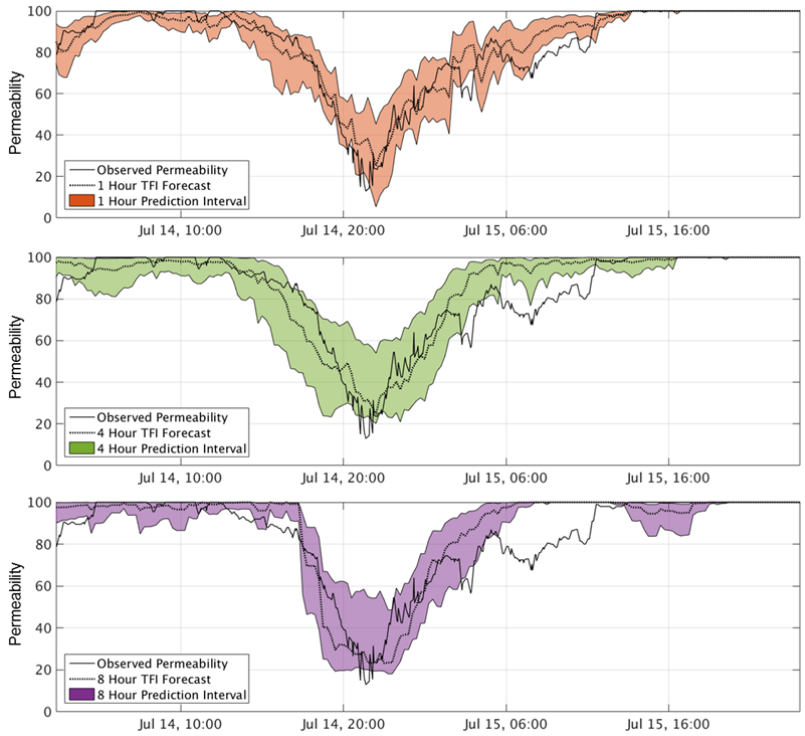

Figure 9. Example permeability verification for 1 hour (top), 4 hour (center), and 8 hour (bottom) forecasts.

B. *Forecast Model Assessment*

As one means to quantitatively assess model skill, the correlation coefficient *r* was used:

$$r = {cov\left(p\ ,\ f_{0.5}(\hat{\boldsymbol{p}})\right)}\Big/{\sigma_p \sigma_{f_{0.5}(\hat{\boldsymbol{p}})}} \quad (5)$$

The correlation coefficient is a measure of how well the observed permeability is related to the median curve in the combined forecast.

The square of the correlation coefficient, r^2, for forecasts made in three disjoint airspace regions (Chicago, New York, and Washington, DC) is shown in Fig. 10 as a function of forecast lead time. Because the dataset is dominated by cases with no weather impact, cases with observed permeability of 100% were removed prior to computing correlation, leaving approximately 6,000 forecasts per hourly forecast lead out to 12 hours. In terms of correlation, the combined forecast offers a clear improvement over the permeability estimates derived from individual forecasts alone. This improvement is most striking in the 3-4 hour lead time.

It interesting to examine the interrelationships between forecasts in the permeability model shown in Fig. 10. For short lead times, permeability measured from the extrapolation forecast dominates the other input forecasts, but quickly loses skill past 3 hours. As the skill of the extrapolation forecast drops, the HRRR and LAMP forecasts become the most skillful at predicting permeability. Permeability estimated from the SREF probabilistic forecast generally showed the least skill of the four forecasts considered here. We believe that this is mostly due to infrequent updating of the SREF (every 6 hours) compared to the other forecasts considered here (which are updated at least hourly). One conclusion is that frequently-updating forecasts are better suited for estimating permeability than are forecasts that update infrequently. Note that the translation of weather forecasts into a specific metric (permeability) enables us to quantitatively compare model performance in terms of airspace impact. Such relationships have been postulated qualitatively in the past; this work represents the first known analysis providing a quantitative comparison of the relative performance of different forecast sources.

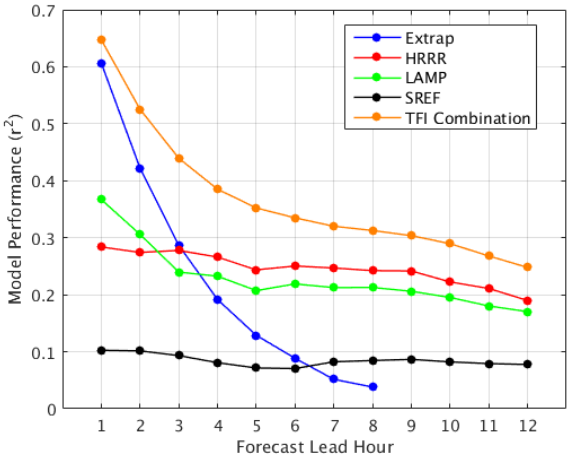

Figure 10. Skill of individual and combined permeability forecasts

V. Operational Field Evaluations

The weather impact algorithms described above were implemented in a real-time decision support tool called Traffic Flow Impact (TFI) in 2014 - 2016. The primary display format for TFI (Fig. 11) includes the 0-8 hour CoSPA weather forecast shown graphically on the top, with a timeline of the impact category for the chosen airspace regions shown across the bottom in a manner similar to that used by the operational Route Availability Planning Tool (RAPT). Each row in the timeline represents a different airspace resource (Flow Constrained Area, FCA).

Each column in the timeline along the bottom of Fig. 11 represents the permeability impact category going out from 0 to 12 hours using a three-color categorization scheme. Permeability values above 85 were colored green in the timeline cell (signifying minor impact), values between 50 – 85 were colored yellow (moderate impact), and values less than 50 were colored red (significant impact). In addition, the user could select a given FCA row and view a drill-down impact timeline, shown on the top left in Fig. 11.

The objective of the summer operations observations was to gain valuable insight into the potential usefulness of the TFI permeability forecast and guide the development of the algorithms and validation process. Input from experienced operational traffic flow managers familiar with the complexity of the NAS and who perform the daily decision-making that impacts NAS performance is critical to developing future advanced methods and products. An initial, informal evaluation was conducted in the summer of 2014. At that time, the concept of TFI was shown to several traffic managers in the ATCSCC to gather initial feedback on the display and its functionality while the underlying algorithms were being further developed and matured.

Formal field demonstrations of TFI were conducted during the previous two convective seasons. Once from 9 July 2015 to 31 October 2015 and then again from 1 June to 31 October 2016. These demonstrations involved FAA, National Weather Service, and commercial airlines. During the field demonstrations, the 0-8 hour CoSPA weather forecasts were viewable by operational users via web-based displays such as that shown in Fig. 11. Given the importance of convective

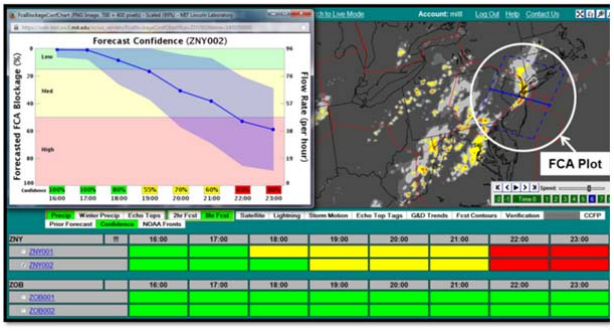

Figure 11 Traffic Flow Impact (TFI) display.

forecasts to air traffic management in the New York metroplex, we conducted focused field operational evaluations covering eight days when storms were forecast to develop across the Eastern U.S. Observers simultaneously visited four FAA ARTCCs (Boston, Washington DC, New York, and Cleveland) and the ATCSCC. Five airline operations centers (Delta, Southwest, American, United, and JetBlue) were also visited during field observations. Observers embedded at each facility documented the use of TFI in the strategic planning process for planning Airspace Flow Programs, Ground Delay Programs, and enhanced reroute planning. Since TFI was a new application available to traffic planners, personnel performed training during the months preceding the release and conducted additional in-situ training during the field observations as needed. Each time a user referred to, modified, or commented on TFI or other traffic management tools, the observer would record the time, event, and actions taken using a standardized data-entry form.

More than 600 person-hours of observations were gathered across the five FAA and five airline facilities during the eight observation days. ATCSCC planners and specialists reviewed TFI on each of the mornings during the strategic planning period (approximately 1000-1600 UTC) to aid in developing collaborative traffic management decisions. A total of 124 events were recorded in which personnel referred to or used TFI in traffic management discussions or decision-making. Of the 124 observed events, TFI was used to support general situation awareness 89 times, to determine parameters for AFPs 19 times, and in other activities (such as reroute planning) 16 times. Users were observed viewing the drill-down plots of the TFI application in order to track the forecasted convective impact on the timeline. In particular, users specifically concentrated on the starting and ending times of forecasted convective events in order to determine the start time and duration of any TMIs that were going to be issued.

Subjective feedback on TFI was also solicited from 47 users at the end of each summer (17 in 2015 and 30 in 2016) using a survey form with a scale from 1 (strongly disagree) to 7 (strongly agree) for each question. Fig. 12 shows the response distributions from two of the questions. As expected from a new concept implemented in a preliminary engineering configuration without formal integration into decision-making processes or training, the user responses were varied. Fig. 12 (a) illustrates that the users generally agreed that TFI was helpful in understanding weather impacts, and (b) was helpful in facilitating discussions with other stakeholders. This feedback supports the general hypothesis that TFI can provide users with information that aids them in collaboratively evaluating weather impacts, but also points to the need to further explore how TFI-like information should be more explicitly integrated into traffic management workflows across stakeholders. It is also apparent that additional work is warranted to assess the best settings for thresholds of impact categories, and to ascertain how uncertainty information can best be displayed and applied in operational decision-making. In some cases, indications of large uncertainty may lead traffic managers to implement less aggressive but more flexible decisions so as to leave open opportunities to adjust as more information is gathered over time.

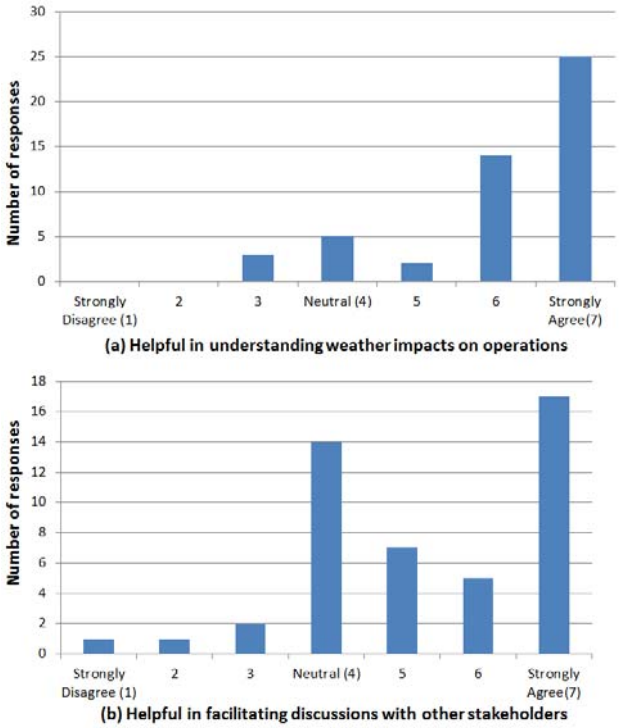

Figure 12. Selected subjective responses.

VI. CONCLUSIONS

This paper has presented the results of research to predict the impact of convective weather on operations using a flow-based permeability measure. The results have shown good agreement between the permeability estimate and measured flow rates for a major air traffic resource controlling flow east/west bound for the NY metro airports and for a terminal area around Chicago O'Hare International Airport. The permeability estimate forms the basis for a weather impact assessment that air traffic planners could use to anticipate the constraints on the air traffic control system and set rates as well as start times of traffic management initiates.

By translating convective weather forecast information into the parameters used in selecting TMIs (e.g., time of onset, level of impact [permeability and flow rates], and duration), it is hypothesized that more effective and timely TMIs can be formulated and assessed in operations. Additionally, we believe that communicating forecast uncertainty as expressed using those same decision variables provides an objective, quantitative basis to better understand and communicate the risks and benefits of various levels of TMI strategies. However, more research and evaluation is needed to verify these hypotheses and ensure that decision support information meets user needs.

Future model development should continue to validate the permeability estimates for a wide range of strategic flows through airspace resources with different configurations, traffic demands, and weather scenarios. The validation methodology should also look at ways to measure and account for times

where high flow rates are maintained at the expense of higher controller workload. This will be important to answer the question of whether the weather impact was managed through reduced flow rates or higher workload. In the future, given objective forecasts from strategic weather forecast products such as CoSPA or SREF and a translation model, as described in this paper, it would be possible to develop disciplined TMI decision making methodologies to manage an appropriate flow rate while not overtaxing the air traffic control personnel.

ACKNOWLEDGMENTS

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited. This material is based upon work supported by the Federal Aviation Administration under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Federal Aviation Administration. We are especially appreciative of the support from William Chan (NASA Ames Research Center) for sponsoring initial development, Rob Hunt, Yong Li, and Mike McKinney (FAA) for funding work to develop en route and terminal airspace algorithm improvements, and Tom Webster (FAA) for the funding and support to conduct field evaluations. We also wish to thank Anthony Tisdall, Lead of System Operations at the ATC Command Center, and the other dedicated staff at the FAA and airline facilities for their time and insights into the strategic decision-making process. Finally, we wish to thank Haig Iskenderian, Kim Calden, Brad Crowe, Peter Erickson, Fulvio Fabrizi, Richard Ferris, Darin Meyer, and Ngaire Underhill of Lincoln Laboratory for their contributions to this work.

REFERENCES

[1] Klingle-Wilson, D., Evans, J. E., "Description of the Corridor Integrated Weather System (CIWS) weather products", Project Report ATC-317, MIT Lincoln Laboratory, Lexington, MA, 2005.

[2] Wolfson, M. M., Dupree, W. J., Rasmussen, R., Steiner, M., Benjamin, S., Weygandt, S., **"**Consolidated Storm Prediction for Aviation (CoSPA)," 13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), New Orleans, LA, *Amer. Meteor. Soc.*, 2008.

[3] Du, J., S. L. Mullen, and F. Sanders, "The NOAA/NWS/NCEP Short Range Ensemble Forecast (SREF) system: Evaluation of an initial condition vs multiple model physics ensemble approach," *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction, Seattle, WA, Amer. Meteor. Soc.* Vol. 21.

[4] Huberdeau, Mark, and Jennifer Gentry, "Use of the Collaborative Convective Forecast Product in the air traffic control strategic planning process," *Journal of Air Traffic Control* (2004).

[5] Cho, J. Y. N., Welch, J. D., Underhill, N. K., "Analytical workload model for estimating en route sector capacity in convective weather," 9th USA/Europe Air Traffic Management R&D Seminar, Berlin, Germany, June 2011.

[6] Song, Lixia, Craig Wanke, and Daniel Greenbaum, "Predicting sector capacity for tfm decision support." *6th AIAA Aviation, Integration, and Operations Conference*. 2006.

[7] Matthews, M. P. and DeLaura, R. "Decision risk in the use of convective weather forecasts for trajectory-based operations", 14th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Atlanta, Georgia, GA, 2014.

[8] DeLaura, R., and Evans, J., "An exploratory study of modeling enroute pilot convective storm flight deviation behavior" 12th Conference on Aviation, Range, and Aerospace Meteorology, Atlanta, GA, 2006.

[9] Rubnich, M. and DeLaura, R. "An algorithm to identify robust convective weather avoidance polygons in en route airspace", 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX, 2010.

[10] Robinson, M, DeLaura, R., Underhill, N., "The Route Availability Planning Tool (RAPT): Evaluation of departure management decision support in New York during the 2008 convective weather season," Eighth USA / Europe Air Traffic Management Research and Development Seminar (ATM2009), Napa, CA, 2009.

[11] Matthews, M. and R. DeLaura, "Modeling convective weather avoidance of arrivals in the terminal airspace", 2nd Aviation, Range, and Aerospace Meteorology (ARAM) Special Symposium on Weather-Air Traffic Management, Seattle, WA, *Amer. Meteor. Soc.*, 2011.

[12] Klingle-Wilson, D. and J. E. Evans, "Description of the Corridor Integrated Weather System (CIWS) weather products"**,** Project Report ATC-317, MIT Lincoln Laboratory, Lexington, MA, 2005.

[13] Weygandt, Stephen S., et al., "Forecast improvement from radar data assimilation within the Rapid Update Cycle, Rapid Refresh, and High Resolution Rapid Refresh forecasts initialized with RUC/RR grids", 14th Conference on Aviation, Range and Aerospace Meteorology, Atlanta, GA, 18-21 January 2010.

[14] Ghirardelli, Judy E., "An overview of the redeveloped Localized Aviation MOS Program (LAMP) for short-range forecasting", 21st Conference on Weather Analysis and Forecasting / 17th Conference on Numerical Weather Prediction, Washington, DC, Amer. Meteor. Soc., 13B5, 2005.

[15] Bright, David R., "Post processed short range ensemble forecasts of severe convective storms", 18th Conference on Probability and Statistics in the Atmospheric Sciences, Atlanta, GA, 29 January - 2 February 2006.

[16] Hoerl, Arthur E., and Robert W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", Technometrics 12.1 (1970): 55-67.

[17] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning, Vol. 1. Springer, Berlin: Springer series in statistics, 2001.

[18] Koenker, R. *Quantile Regression*. Cambridge University Press, 2005.

[19] Portnoy, S. and R. Koenker, "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators", *Statistical Science* (1997): 279-300.

AUTHOR BIOGRAPHIES

Michael Matthews is Associate Technical Staff in the Air Traffic Control Systems Group at MIT Lincoln Laboratory. He has a BS in Meteorology from the University of Massachusetts at Lowell.

Mark Veillette is Technical Staff in the Air Traffic Control Systems Group at MIT Lincoln Laboratory. He has a PhD in Mathematics from Boston University.

Joe Venuti is an aviation weather specialist in the Air Traffic Control Systems Group at MIT Lincoln Laboratory. He has an MS in Meteorology from Penn State University and a BS in Meteorology from the University of Massachusetts at Lowell.

Rich DeLaura is Technical Staff in the Air Traffic Control Systems Group at MIT Lincoln Laboratory. He earned a BA in Chemistry and Physics from Harvard University.

James Kuchar leads the Air Traffic Control Systems Group at MIT Lincoln Laboratory. He holds SB, SM, and PhD degrees from MIT in Aeronautics and Astronautics.