Identification of Robust Terminal-Area Routes in Convective Weather

Diana Michalek Pfeil, Hamsa Balakrishnan

Massachusetts Institute of Technology {dianam@mit.edu, hamsa@mit.edu}

Convective weather is responsible for large delays and widespread disruptions in the U.S. National Airspace System, especially during summer. Traffic Flow Management algorithms require reliable forecasts of route blockage to schedule and route traffic. This paper demonstrates how raw convective weather forecasts, which provide deterministic predictions of the Vertically Integrated Liquid (the precipitation content in a column of airspace) can be translated into probabilistic forecasts of whether or not a terminal-area route will be blocked. Given a flight route through the terminal-area, we apply techniques from machine learning to determine the likelihood that the route will be open in actual weather. The likelihood is then used to optimize terminal-area operations, by dynamically moving arrival and departure routes to maximize the expected capacity of the terminal area. Experiments using real weather scenarios on stormy days show that our algorithms recommend that a terminal-area route be modified 30% of the time, opening up 13% more available routes that were forecast blocked during these scenarios. The error rate is low, with only 5% of cases corresponding to a modified route being blocked in reality, while the original route is in fact open. In addition, for routes predicted to be open with probability 0.95 or greater by our method, 96% of these routes (on average over time horizon) are indeed open in the weather that materializes.

Key words: Convective weather; Air Traffic Management; Integration of weather forecasts and air traffic management; Route robustness; Airspace capacity

History:

1. Introduction

The increase in demand for air travel over the past few years has been accompanied by an increase in congestion and delays in the United States National Airspace System (NAS), and has made the system more susceptible to weather disruptions. This problem is particularly intense during summer months, when travel demand is high and there are frequent thunderstorms (convective weather activity) over much of the continental U.S. It has been estimated by the Joint Economic Committee, Unites States Senate (2008) that domestic air traffic delays in 2007 cost the U.S. economy $41 billion. It has also been estimated that 76.9% of all delay in the NAS and 25% of all delayed flights in 2007 were weather-related (Joint Economic Committee, Unites States Senate 2008, Bureau of Transportation Statistics 2008). With the demand for air traffic operations expected to grow significantly over the next two decades, it has become increasingly important to develop approaches

that will enable the efficient operation of the airspace system, even in the presence of convective weather (Joint Planning and Development Office 2004). In addition to better incorporation of weather forecasts into traffic flow management, a promising approach identified as part of the NextGen ATM-weather integration plan (Joint Planning and Development Office 2009) involves relaxing the rigid structure of NAS in order to access available airspace and increase achievable capacity. This paper focuses on bridging the gap between available aviation weather forecasts and air traffic flow management algorithms, by using archived weather data to identify robust routes in the terminal area, and by introducing a method to dynamically restructure terminal airspace to use more robustly available routes.

1.1. Motivation and Background

The capacity of airspace resources is strongly influenced by ambient weather, since aircraft need to avoid hazardous atmospheric conditions and may therefore be forced to deviate from their planned trajectories. Air Traffic Flow Management (ATFM) algorithms traditionally handle weather by assuming that the impact of weather on the capacity of a resource at any time is known, and use the deterministic estimates of capacity to route flights between their origins and destinations (Bertsimas and Patterson 2000, Menon et al. 2006, Sun et al. 2006). Algorithms have also been developed to efficiently synthesize routes to avoid weather, given fine-grained and deterministic forecasts as input (Prete and Mitchell 2004, Krozel et al. 2006). However, deterministic capacity estimates based on weather forecasts can be inaccurate under stormy weather conditions. This fact has motivated optimization approaches that assume multiple capacity scenarios for airspace resources, with associated probabilities of occurrence (Bertsimas and Odoni 1997). More recently, robust optimization approaches have been proposed that assume a set of possible capacity values, and try to keep the system safe for any possible realization of the capacity (Bertsimas et al. 2007). There have also been attempts at creating models of capacity, given either deterministic weather forecasts (Krozel et al. 2007), or weather forecasts accompanied by randomly generated regions of uncertainty (Mitchell et al. 2006). In Song et al. (2009), the authors study the correlation between sector throughput and various measures of convective coverage, and conclude that these correlations can be incorporated into an algorithm for sector capacity. Finally, there has been growing interest in operational concepts such as adaptable airspace, which focuses on resolving capacity imbalance by dynamically changing airspace structure (Kopardekar et al. 2009, Klein et al. 2007).

While there has been much prior research on air traffic flow management algorithms that assume accurate convective weather forecasts as input, there has been little work in adapting existing convective forecasts, and in evaluating relevant accuracy and error metrics for use in these applications.

Traditionally, entities that develop forecast products have provided users with statistics based on pixel-by-pixel comparisons of the forecast with actual weather. These statistics, used to evaluate the performance of the forecast product, include rates of false positives, false negatives, bias, and skill scores such as the Critical Success Index. Wolfson et al. (2004), Kay et al. (2006), Weygandt and Benjamin (2004), and Seseske et al. (2006) have conducted historical evaluations of the Lincoln Lab CIWS forecast, Collaborative Convective Forecast Product (CCFP), RUC Convective Probability Forecast, and National Convective Weather Forecast 2 (NCWF-2), respectively. These studies have typically found low accuracy of weather forecasts, but have not indicated how one might reliably use the forecasts for ATFM. There have also been efforts to develop more ATFM-based metrics for evaluating convective weather forecasts. These include metrics based on co-occurance probabilities (Chatterji and Gyarfas 2006), object-oriented approaches to forecast validation (Brown et al. 2004, Mahoney et al. 2004), and studies that evaluate forecast accuracy as a function of the spatial resolution and storm type (Evans et al. 2009). These studies show that while the accuracy increases at the cost of spatial resolution when compared to pixel-based comparisons, the skill scores still tend to be low and show high variability. Overall, the spacial smoothing of these verification techniques, as well as the forecast errors reported in existing validation studies, makes it difficult to use direct predictions in fine-grained traffic-flow algorithms.

There has been some recent research attempting to bridge this gap between convective weather forecasts and ATFM algorithms. Lincoln Lab has developed and validated a model of pilot deviation, which predicts, given convective weather and echo tops data, the probability that a pilot will deviate from around a region of airspace (DeLaura et al. 2008). In addition, the Route Availability Planning Tool (RAPT) uses convective weather forecasts to model jet route blockage deterministically for departures (DeLaura and Allan 2003). Finally, Liu et al. (2008) use historical airport arrival rates to create scenario trees of arrival capacities at any given airport. Although these scenario trees allow them to improve upon previous models of the ground holding problem, capacities are not tuned to reflect day-of weather information, and validation focuses on SFO, which is most often impacted by fog, as opposed to convective weather.

1.2. Contributions of This Paper

While ATFM algorithms assume the availability of reliable probabilistic capacity forecasts, there is still little understanding of the errors present in current forecasts and their effect on downstream algorithms. This paper considers the problem of translating raw convective weather forecasts, which provide deterministic predictions of the Vertically Integrated Liquid (VIL, the precipitation content in a column of airspace) into probabilistic forecasts of route blockage for use in ATFM.

4

Pfeil and Balakrishnan: *Robust Terminal-Area Routing in Convective Weather*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

We adopt a data-driven approach to achieving this objective. We use state-of-the-art aviation convective weather forecasts to identify robust routes, that is, routes that are likely to remain viable in the actual weather that materializes. We consider various features (characteristics) of the forecast weather along arrival and departure routes, and identify features that are highly correlated with route blockage. Using techniques from machine learning, we propose and validate classification algorithms that predict whether a given route is likely to be open or blocked in actual weather, based on the values of different features of the route as determined by the forecast. We evaluate our algorithms using several metrics, such as the accuracy (the fraction of time that the prediction is correct), the false positive rate (the fraction of time that we forecast that the route will be open but it ends up being closed), and the false negative rate (the fraction of the time that we forecast that the route will be closed, but instead it remains viable). We convert this binary prediction into a stochastic forecast by assigning a probability to each prediction, representing the classifier's probability that the route is open. We validate these probabilistic forecasts against actual weather scenarios, and then demonstrate how they can be used to optimally relocate arrival and departure routes within the terminal-area to maximize expected capacity.

2. Convective Weather Forecasts

In this section we introduce the state-of-the-art Corridor Integrated Weather System (CIWS) forecast product developed by MIT Lincoln Laboratory, and discuss limitations in forecast accuracy when using traditional metrics for forecast evaluation.

2.1. CIWS Forecast Product

The 0-2 hour CIWS convective weather forecast product consists of a grid of 1 km $\times$ 1 km pixels covering (in 2D) a large portion of the NAS (Wolfson et al. 2004). Each pixel contains a predicted value of Vertically Integrated Liquid (VIL), indicated by an integer value in the range $[0, 255]$. Figure 1 shows a sample forecast for ATL. These VIL values are divided into seven levels of convective activity, ranging from level 0 (none) to level 6 (very severe). A VIL value above a certain threshold (133, in practice) in the observed data corresponds to weather of severity level 3 or higher, and is commonly considered by pilots to be hazardous (Krozel et al. 2006). Although we use this strict threshold to define hazardous weather in this paper, the threshold of what pilots will fly through is not fixed, and depends on non-weather factors such as airspace demand and proximity to destination airport. Forecasts are issued for horizons spanning between 5 and 120 minutes in 5-minute increments, and are updated every 5 minutes. In other words, at time T_0,

forecasts are available for time $T_0 + 5, T_0 + 10, T_0 + 15, \ldots, T_0 + 120$. The observed VIL values for the same region of airspace at the corresponding times are also available, and can be used for evaluating the quality of the forecast.
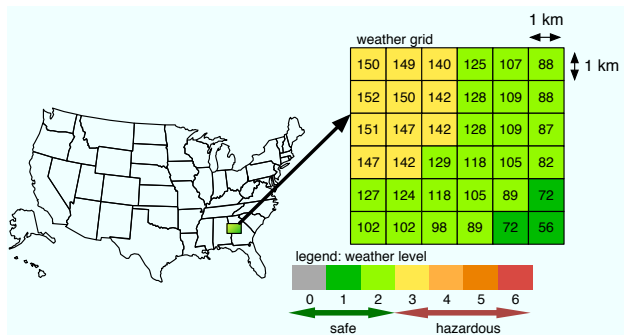

Figure 1   **Sample Lincoln Lab CIWS forecast near ATL.**

2.2. Validation of Pixel-based Forecasts

We compute several standard statistics to evaluate CIWS using data from eight days of high convective activity during the summer of 2008. The results highlight some of the issues associated with integrating convective weather forecasts with ATFM.

We first define a few verification statistics that have traditionally been used to evaluate weather forecasts. Consider a pixel for which a forecast and actual weather data exist. A true positive (TP) is when the forecast and actual weather both show hazardous weather at the pixel. A true negative (TN) is when the forecast and actual both show no hazardous weather. A false positive (FP) is when the forecast predicts hazardous weather, but the pixel is clear in actual weather, and the opposite is the case in a false negative (FN). Using these base statistics, we can form four standard measures of forecast skill: the Critical Success Index (CSI) is defined to be $\frac{TP}{TP+FP+FN}$, the probability of detection (POD) is defined to be $\frac{TP}{TP+FN}$, the false alarm rate (FAR) is defined to be $\frac{FP}{TP+FP}$, and accuracy is defined to be $\frac{TP+TN}{TP+TN+FP+FN}$.

Figure 2 shows the performance of the forecast in terms of POD, FAR, CSI, and accuracy. Accuracy posts the best performance with scores above 0.9 for each time horizon, although this high score is misleading as it merely reflects the high proportion of Level 0 pixels in the forecast and the actual data, resulting in a majority of true negatives. The other three statistics, which do not include TN and are therefore more relevant for assessing the quality of the forecast, perform very poorly. The CSI is 0.5 and below across time horizons, meaning that more than half of the time that either the forecast or actual pixel contains weather, the forecast is erroneous. The POD and FAR are similarly low, especially for time horizons of 30 minutes and higher.
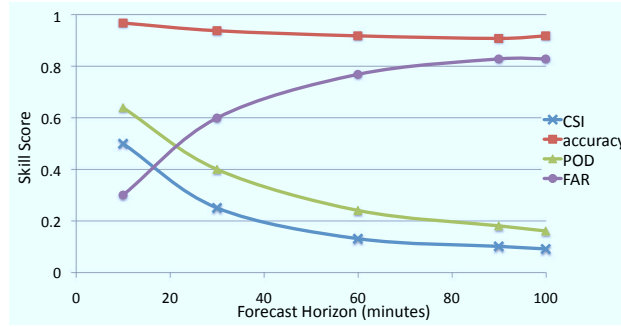
Figure 2   **Four pixel-based skill scores for the CIWS weather forecast, at increasing time horizons.**
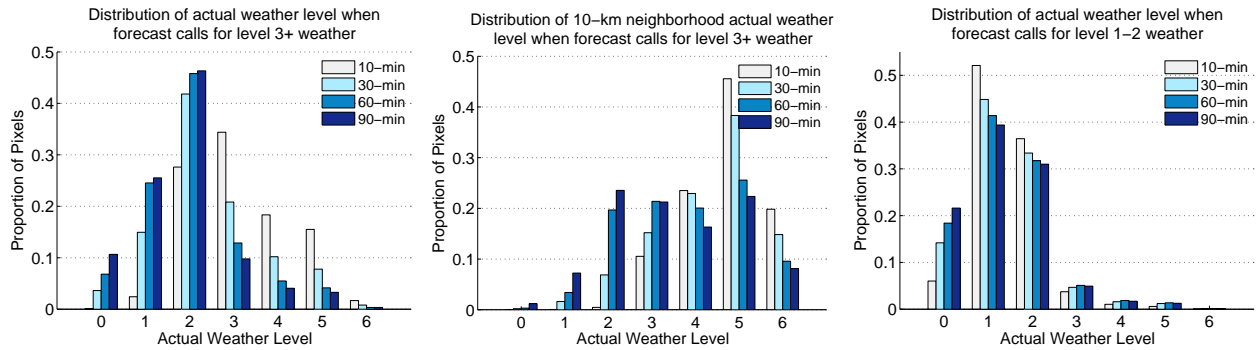

Figure 3   **Histogram of actual weather across a large set of summer 2008 weather scenarios, given a forecast of L3+ weather (left and center), and L1-L2 weather (right) at a pixel. The center histogram validates a pixel against its 10 km neighborhood, to account for small spatial errors in the forecast.**

To further explore the actual weather that arises when weather that is Level 3 or higher is forecast, Figure 3 shows the distribution across time horizons of actual weather when Level 3+ weather is forecast (left). For time horizons of 30 minutes and longer, the actual weather is level 2 and lower at the majority of pixels, indicating an over-forecast of hazardous weather. However, when the definition of true positive is relaxed so that the presence of L3+ weather within a B km neighborhood of the forecast pixel counts as a hit, this distribution shifts right, with a majority of pixels at L3+ (as illustrated in the center histogram for $B = 10$). This reflects a spatial error in the forecast, though the value of B that makes this distribution shift varies, and can be too large to be useful for fine-grained planning of flight plans. The right-most histogram shows the distribution of actual weather when Level 1 or 2 weather is forecast (Level 0 is left out, because it represents a majority of cases, and shifts the distribution to the left), and indicates that the actual weather is Level 2 or lower in the vast majority of cases, indicating a strong TN score.

There are other issues associated with such pixel-based comparisons. First, it is possible that the forecasts get the general weather trends right, but are incorrect in the position of the weather cells. A 10 km × 10 km storm cell forecast, for example, might be displaced by 10 km to the west

when observed, resulting in CSI and POD scores of zero, and a FAR of 1. For planning purposes, however, this forecast may be quite good since moving a planned trajectory could easily be moved 10 km east. Second, a storm consisting of many small sparsely distributed cells of weather activity (a popcorn storm) may have low forecast accuracy in pixel-by-pixel comparisons, and yet have open routes between the cells, resulting in no practical capacity reduction. For these reasons, we adopt a route-based approach to validating and utilizing weather forecasts. We aim to identify opportunities for increased capacity even in the presence of convective activity and inaccurate pixel-based forecasts, by studying persistent routes through weather.

3. Route Robustness Model Setup

Motivated by the pixel-based evaluation of convective weather forecasts, we introduce a route-based model for incorporating weather forecasts into air traffic planning.

3.1. Terminal Airspace Model

We begin by describing the terminal airspace model (illustrated in Figure 4) that is used to identify robust arrival and departure routes, and for optimal fix placement. The input is a terminal-area,
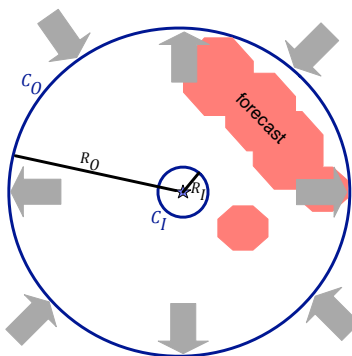

Figure 4 Model of terminal-area flows. Arrival flows enter through C_O and flow into C_I, while departure flows travel in the reverse direction. The red region represents a forecast weather hazard.

defined by two concentric circles: an outer circle C_O of radius R_O, and an inner circle C_I of radius R_I. Aircraft flying under instrument flight rules currently follow their filed flight plans which are represented by standard waypoints connected by airways. Aircraft flows from and to the airport are typically routed through specific way-points on the outer circle known as *fixes*, which are points of entry into or exit out of the terminal-airspace. The circle C_O represents the points at which arriving aircraft first enter the terminal airspace, while C_I represents the point at which aircraft start their landing procedure (for example, the crosswind leg of the approach) into the airport. In contrast,

departures traverse the terminal-area in the reverse direction, entering it close to the airport at C_I and exiting it through the outer boundary C_O. The parameters R_I and R_O are adjustable, and the sensitivity of the model to these parameters is studied in Section 6. In this study, R_I ranges between 10 km and 20 km, and $R_O = 100$ km.

3.2.   Problem Statement

The objective of this paper is to determine routes that are likely to be robust to weather disruptions, by understanding and incorporating the inherent uncertainty associated with weather forecasts. In other words, our problem can be stated as follows:

Given a weather forecast for some time in the future and a set of predetermined potential routes, we would like to best identify those routes that are likely to be open in the actual weather that materializes, and also quantify the uncertainty associated with our prediction mechanism.

In order to solve this problem, we adopt the following definition for an open route.

DEFINITION 1. A route is defined to be **open** or **clear** in the observed weather if there exists a route within a B km neighborhood of the original route that is not impacted by weather. This potentially displaced route is called the **perturbed route**. Note that the perturbed route may be identical to the original one.

This relaxed definition allows for slight deviations in a planned route that reflect the "wiggle room" or the ability of an aircraft to make small adjustments to the planned route. This flexibility is quantified by the parameter B, which will typically be small. The sensitivity of the approach to the choice of B is studied in Section 6.3.2.

The problem stated above is an important one from the air traffic management perspective for several reasons. First, it attempts to capture how the impact of observed weather on routes differs from predicted impact, rather than simply evaluating forecasts using pixel-by-pixel comparisons (Michalek and Balakrishnan 2009a). Second, it accounts for the ability to allow small deviations from planned routes without affecting operations. Third, this approach suggests that in the terminal-area, the theoretical capacity may not be a sufficient metric to measure the impact of weather on air traffic flows (Krozel et al. 2007). This is because while the theoretical capacity might predict that N arrival routes will be open for the next hour, it may not indicate the constraint (which is critical for planning) that these routes all enter the terminal-area from the West. Finally, it is possible for the forecast and realized theoretical capacities to be identical, and to yet require that aircraft use trajectories that deviate significantly from the planned routes.

3.3. Dynamic Weather Grid

As aircraft travel along a route, we are interested in whether the route is blocked at a particular location at the time when the aircraft flies through it. In order to model this effect, we construct a *dynamic weather grid* by splicing together weather data for time instants t that increase from the outer to inner circle for arrivals (and from the inner to outer circle for departures). The distance between two concentric circles in the grid (shown in Figure 5(a)) corresponds to the distance flown in 5 minutes by a typical aircraft. These circles are drawn assuming an average aircraft speed of 180 knots in the terminal area; we have also conducted similar analysis for aircraft speeds of 85 knots (corresponding to slower, general aviation aircraft). Suppose the planning horizon of interest is 30 minutes. Then, at each time, we are interested in determining routes for aircraft that are currently 30 minutes from the terminal-area, i.e., we would like to give them recommendations on which fix and terminal-area route to fly through. In order to do this, we need to look at the 30 minute forecast in the outermost annulus of the dynamic weather grid, the 35 minute forecast for the next annulus, the 40 minute forecast for the next one, and so on. Similarly, for validating the performance of this forecast along this route, we will consider the observed weather along the route in each annulus at the time that the aircraft flies through it, i.e., the observations 60 minutes from the current time for the outermost annulus, 65 minutes from the current time for the next annulus, and so on. All weather scenarios considered in the remainder of this paper refer to the dynamic weather grid.
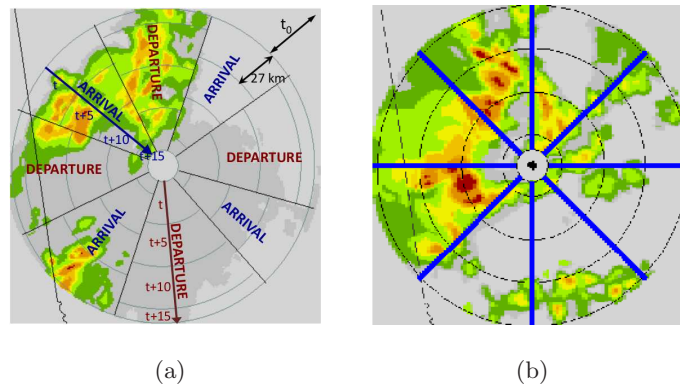

(a)   (b)

Figure 5   **Example dynamic weather scenarios. (a) A sample dynamic weather grid showing arrival and departure sectors, created by splicing together consecutive forecasts. We assume aircraft arrive at C_O at time t. Planning occurs t_0 minutes in advance of aircraft arrival. For departures, the corresponding dynamic grid assumes aircraft arrive at C_I at time t, with the same t_0-minute planning horizon. This grid will therefore be used for planning at the current time, namely, $(t - t_0)$. (b) Illustration of the eight routes sampled for each arrival forecast scenario.**

10

Pfeil and Balakrishnan: *Robust Terminal-Area Routing in Convective Weather*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

4. Generation of Data Sets

This paper presents a technique to identify routes that are likely to be robust to the inaccuracies in the forecast. The approach is based on a large-scale evaluation of the CIWS Convective Weather Forecast, and the difference between the predicted and observed impact on routes. An essential step is the generation of the necessary data sets, which consists of the selection of forecast and observed weather scenarios, the selection of potential arrival and departure routes, and the validation of these routes in observed weather.

4.1. Selection of Weather Scenarios

We focus on the terminal airspace of Hartsfield Atlanta International airport (ATL), which is the busiest airport in the United States and experiences significant delays due to convective weather. A set of weather scenarios during the summers of 2007 and 2008 was identified, and then partitioned by year to form the training and test sets. Each day of CIWS data yields approximately 30 GB of uncompressed binary data. To identify convective weather scenarios for the ATL terminal-area, the forecasts for the airspace surrounding ATL were extracted and visualized to determine the time periods with maximum convective weather activity. This resulted in an average of 4 weather scenarios per day, separated from each other by at least 30 minutes. Separate datasets were created corresponding to planning horizons of 10-, 30-, 60-, 90- and 100-minutes, and for both departures and arrivals. The resulting training sets consisted of weather scenarios from each of the 14 most weather-impacted days during the months of June and July 2007, when ranked according to weather-related delays, while the test sets consisted of scenarios from the 8 most weather-impacted days during the months of June through August 2008. These data sets are referred to as Data2007 and Data2008, respectively, for the remainder of this paper.

4.2. Route Selection and Validation

Potential aircraft trajectories through the forecast grid of each weather scenario were generated by sampling eight straight routes between C_0 and C_I, as depicted in Figure 5(b). These eight trajectories represent a sampling of routes through varying weather forecasts. Arrival trajectories point toward the inner circle, while the departure trajectories are oriented in the opposite direction.

Each route r generated in the manner described above was evaluated using the observed weather data. Recall that r is *open* if there exists a corresponding perturbed route in the observed weather grid within B km of r that does not pass through any weather hazards. This B km neighborhood allows for slight perturbations in the original route (on the order of several kilometers). Open routes were synthesized by solving the following shortest-path problem with turn penalties through the dynamic grid of observed weather (Michalek and Balakrishnan 2009b).

The directed graph $G(\mathcal{N}, \mathcal{A})$ is constructed such that the set of nodes $\mathcal{N}$ contains all pixels within B km of r (in the dynamic observed weather grid) which are free of weather hazards, and such that each set of adjacent nodes forms an arc $a \in \mathcal{A}$. At time t, a unit of flow is sent from a set of source nodes $\mathcal{S} = C_O \cap \mathcal{N}$ (the subset of nodes lying on the outer circle C_O) to a set of sink nodes $\mathcal{T} = C_I \cap \mathcal{N}$. For simplicity, we use a standard transformation and introduce a supersource $\bar{\mathcal{S}}$ and a supersink $\bar{\mathcal{T}}$, and route one unit of flow between the two through the source nodes and sink nodes (Ahuja et al. 1993). For arc (i, j), let parameter $\text{NX}(i, j) \in \mathcal{N}$ be the next node if no turns are taken. In other words, nodes i, j, and $\text{NX}(i, j)$ form a straight line in the observed weather grid. Since we would like to recommend a route which, in addition to being short, requires a minimum amount of maneuvering on the part of pilots, the objective is to find the minimum cost flow f such that out of all minimum cost flows, f has the minimum number of turns.

The problem as posed above, the shortest-path problem with turn penalties, is modeled by the Integer Program (IP) below. Although this problem is known to be solvable in polynomial time (Ahuja et al. 1993), our IP-based approach allows for simple and useful augmentations such as minimizing the deviation of the solution from the original route, sector boundary movements, etc. This problem is solved for each of the selected routes in the data set; the infeasibility of the problem implies that the route is blocked in the observed weather grid, while feasibility implies that the route is considered open. A version of this problem can be also be solved with different sets of sources and sinks and with $B = \infty$ to generate a large set of candidate routes for a given weather forecast scenario (Michalek and Balakrishnan 2009a). Furthermore, although the construction above models the case of arrivals, by modifying the underlying dynamic grid, the same IP can be used to model departures as well.

$$x_{ij} := \text{flow on arc } (i, j) \in \mathcal{A}$$

$$z_{ij} := 1 \text{ if } (i, j) \in \mathcal{A} \text{ is a turn}, 0 \text{ otherwise}.$$

$$\min \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} + \lambda \sum_{(i,j) \in \mathcal{A}} z_{ij}$$

$$\text{s.t.} \sum_{\substack{j \in \mathcal{N}: \\ (i,j) \in \mathcal{A}}} x_{ij} - \sum_{\substack{j \in \mathcal{N}: \\ (j,i) \in \mathcal{A}}} x_{ji} = b_i \quad \forall i \in \mathcal{N} \tag{1}$$

$$z_{ij} \geq x_{ij} - \sum_{\substack{k \in NX(i,j): \\ (j,k) \in A}} x_{jk} \quad \forall (i, j) \in \mathcal{A} \tag{2}$$

$$x, z \in \{0, 1\}^n \tag{3}$$

Constraints (1) are the flow balance constraints, with $b_i := -1$ for a supersource $\bar{\mathcal{S}}$, $b_i := +1$ for a supersink $\bar{\mathcal{T}}$, and $b_i := 0$ for all other nodes i in $\mathcal{N}$. Constraints (2) in conjunction with the penalty

term in the objective function serve to minimize the number of turns in the path without changing the path length, since it is desirable that aircraft trajectories have a limited number of turns for simplicity. All arcs that follow (i, j), *except* (j, k) for $k = \text{NX}(i, j)$, pay a penalty in the objective function. The parameter λ is chosen to be sufficiently small (less than the maximum length of any path) in order to eliminate longer routes with fewer turns. Finally, x and z are binary variables to ensure that flow is not split up.

In the remainder of this paper we are mainly interested in the feasibility of this problem, but use the objective function to identify actual perturbed routes for planning and visualization purposes.

4.3. Dataset Details

Table 1 shows the overall statistics of the route blockage test datasets (using weather scenarios in Data2007) for arrivals and departures at the five planning horizons studied. The datasets correspond to a wiggle room of $B = 8$ km, inner radius $R_I = 10$ km, and outer radius $R_O = 100$ km. Each route is evaluated using the forecasts and observed weather appropriate for the times at which the affected aircraft will traverse the route. For example, for a planning horizon of 30-minutes, forecasts ranging from 30 to 45 minutes will be used for arrivals, as described by the dynamic weather grid (Section 3.3). The parameters B, R_I and R_O are configurable, and are discussed in a sensitivity analysis in Section 6.

	t_0 (min)	# Routes	Forecast Open (%)	Actual Open (%)	Actual Open | Forecast Open (%)	Actual Closed | Forecast Closed (%)
Arrivals	10	408	48	74	98	52
	30	408	47	74	94	57
	60	408	48	74	86	64
	90	408	57	74	86	58
	100	408	62	74	84	59
Departures	10	408	50	76	99	54
	30	408	50	76	94	59
	60	408	48	76	87	66
	90	408	58	76	89	60
	100	408	63	76	88	56

Table 1   Overall statistics for each of the 8 datasets.

Each (test) dataset contains 408 routes, the majority of which are open. The percentages of routes that are forecast open (i.e., routes that do not pass through Level 3+ weather in the forecast) are between 48% and 63% for both arrivals and departures, meaning that approximately half of the routes in the dataset are forecast to be blocked. However, these same routes are open over 74% of the time in the weather that materializes (that is, there is a perturbed route in the neighborhood of the original route which does not pass through Level 3+ weather in the observed weather). The last

two columns indicate how the forecasts and true weather differ for individual routes. Routes that are forecast as open are overwhelmingly open in the observed weather grid, with rates of 84% and above. Arrivals have slightly lower rates than departures, and the rates decrease with increasing planning horizon. Both of these trends are to be expected, because arrivals typically encounter the bottleneck at the end of their route through terminal airspace, where the forecasts are less accurate. Finally, routes that are forecast as closed are closed in the true weather approximately 60% of the time. These low rates reflect the effect of the additional flexibility allowed for finding routes in the actual weather. Figure 6 contains examples of routes synthesized in the forecast grid, along with the same routes validated against the observed weather.
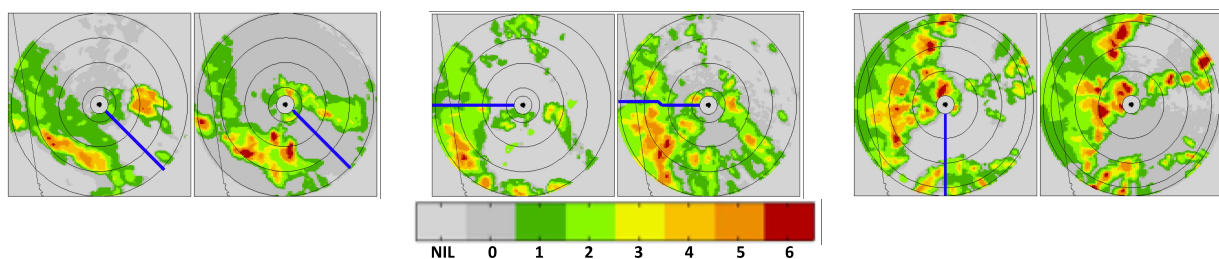

Figure 6   **Each pair of images shows a scenario, with a sample route in a dynamic forecast grid on the left, and the corresponding route in the actual weather on the right. The left-most weather scenario is an arrival route at the 60-minute planning horizon, and depicts a route that is open according to the forecast and ends up being open in actual weather. The scenario in the center shows an arrival route at the 90-minute planning horizon. The precise route is blocked according to the forecast, but a nearby route is available in the true weather grid. The right-most scenario is a departure route at the 30-minute horizon. In this situation, the forecast route is not open in the observed weather grid.**

The raw data suggest that subject to minor adjustments, planning at time horizons up to 100-minutes is quite reasonable, since routes that are forecast to be open end up being overwhelmingly so. Likewise, routes that are forecast to be blocked tend to be open in the observed weather, which indicates potentially underutilized capacity. This is encouraging, and shows that allowing even small adjustments from fixed arrival routes can improve the quality of decision-making based on the forecast. The next sections explore the problem of better predicting blockage using forecast data.

5.   Identifying Features That Influence Route Robustness

Having generated a set of terminal-area routes, the natural next step is to identify characteristics of the convective forecast which may best reflect the likelihood that a trajectory will be open in the observed weather. These features of a route can then be used as predictors of route blockage.

14

Pfeil and Balakrishnan: *Robust Terminal-Area Routing in Convective Weather*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

This section introduces the set of features chosen as potential blockage predictors, and uses mutual information to compare features' predictive power across planning horizons, and across arrival and departures.

5.1. Potential Features of Interest

For each route r, eleven features of interest were identified and each feature was correlated with route blockage:

1. Mean VIL along the route
2. Standard Deviation of VIL along the route
3. Minimum distance to Level 3+ weather along the route
4. Mean distance to Level 3+ weather along the route
5. Maximum VIL in neighborhood of the route
6. Theoretical capacity for the weather scenario
7. Number of segments in the minimum cut
8. Length of the minimum cut segment (bottleneck) that the route passes through
9. Length of tightest bottleneck
10. Maximum pixel density of Level 3+ weather along route
11. Maximum VIL density along route

Features 1-4 are reasonably self-explanatory, and pertain to the VIL along r and the proximity of r to hazardous weather. Feature 5 is the maximum VIL forecast in the neighborhood of radius B along r. Features 6-8 refer to the theoretical capacity of the dynamic forecast grid and the corresponding minimum cut. These features are computed using algorithms from continuous maximum flow theory, which shows that the maximum throughput of a continuous domain (in our case, the annulus from C_O to C_I) corresponds to the minimum cut through a corresponding discrete graph (see Mitchell (1988), Krozel et al. (2007)). Feature 9 contains the length of the minimum bottleneck through which r passes. Feature 10 reflects the intensity of the weather in the neighborhood of r. It is computed by taking a B km neighborhood around r, and finding the strip of pixels perpendicular to r with the largest percentage of Level 3+ forecast pixels. If r is forecast to pass through Level 3+ weather, Features 8-9 will be 0, but Feature 10 may still contain pertinent information about the nature of the hazard. Finally, Feature 11 is similar to Feature 10, except that it considers the largest average VIL in a perpendicular strip.

5.2. Identifying Robust Routes Using Individual Features

A simple model of route robustness can be built based on the correlation of each individual feature with route blockage, giving smooth estimates of the probability of blockage at each feature level. For example, if a route through airspace is forecast open but has an average VIL that is just below the Level 3 threshold, it is more likely to be blocked in the actual weather than a similar route

with an average VIL of zero. This suggests that as the average VIL increases, the route is more likely to be blocked.

Similar correlations can be computed for all 11 features and parameter values (for varying planning horizons and flight directions). These correlations have been explored in prior work, and show that several of the features strongly correlate to route blockage, especially at the shorter planning horizons (Michalek and Balakrishnan 2009b). However, the challenge is to combine the information of these individual features to create a better predictor of route blockage. We develop such a classifier in Section 6.

5.3. Feature Selection

To evaluate features for classification and gain a better understanding of which features best correlate with blockage individually, we compute the Mutual Information (MI) between each feature X_i and the blockage label y (+1 for open, -1 for blocked). Mutual information is an information-theoretic measure of the dependence between two random variables X and Y, and measures how much the uncertainty of X is reduced if Y is observed. This measure considers each feature individually and does not capture situations in which two features combined correlate well with y. In other words, the larger the value of MI for a feature, the greater the correlation of that feature to route availability. For discrete random variables X and Y with joint pmf $p_{XY}(x,y)$, and marginal pmfs $p_X(x)$ and $p_Y(y)$, respectively, their mutual information, $I[X;Y]$, can be expressed as

$$I[X;Y] = \sum_x \sum_y p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

We approximate the distribution functions using their Maximum Likelihood parameter estimates, which are valid when the dataset size is much larger than the domain size of the joint pmf. For features with continuous domains, we do no adopt more complicated approaches to approximating MI for continuous distributions, and instead discretize the data into k equally sized bins.

Figure 7 contains a comparison of MI across features and planning horizons for both departure and arrival datasets. As expected, MI tends to be higher at shorter planning horizons, reflecting the higher forecast accuracy at these time horizons. This also explains why departures have slightly higher MI than arrivals across the board, as arrivals cross the inner circle (a bottleneck region) later in their trip through the (dynamic) terminal area. Figure 7 reveals several insights into the relationship between VIL forecasts and route blockage: for example, it shows that Feature 3, the minimum distance between the route and a weather obstacle (Level 3+ weather) in the forecast, is a worse predictor of blockage than the mean distance of the route to Level 3+ weather (Feature
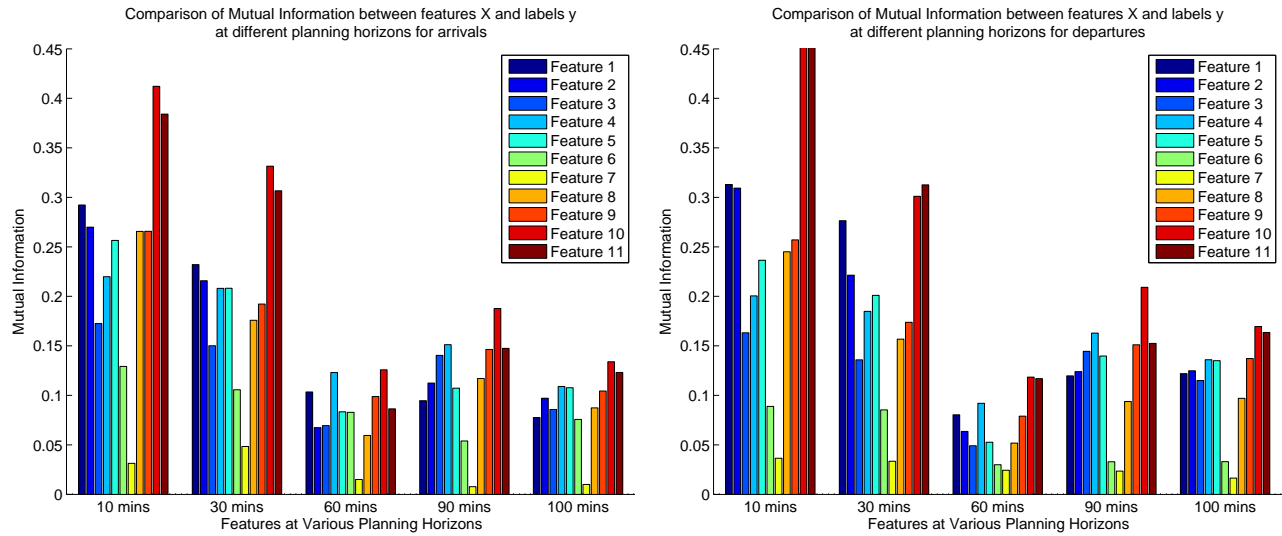
Figure 7   **Comparison of mutual information across features and planning horizons for arrivals (left) and departures (right)**

4). It also suggests that when the standard deviation of the forecast VIL along the route is high, it is more likely to be blocked (i.e., Feature 2 has reasonably high MI). This is likely because a high standard deviation in the VIL means there is much uncertainty in the location of Level 3+ weather, and the obstacles are more likely to change between the forecast and when the route is flown. The maximum density of forecast weather obstacles in the neighborhood of a route (Features 10 and 11) is found to be a strong predictor of route blockage.

Surprisingly, Figure 7 shows that a frequently-cited quantity in planning terminal-area routes, the theoretical capacity (a measure of how many routes into the airport do not pass through forecast weather obstacles, Features 6 and 7) is a poor predictor of route blockage. The reason is that although the theoretical capacity is a prediction of how many routes will be open, it does not give any indication of which ones they would be. For the purposes of routing flights through the terminal-area, it is not sufficient to predict that two arrival routes will be open; we also need to accurately predict whether these will correspond to one each from the North and South, or one from the North and one from the East.

The relative MI of features tends to reflect two principles: the spacial accuracy of a weather forecast decreases with increased time horizon, and the weather forecast in the neighborhood of a route (including the structure of weather cells) has bearing on the viability of a route. Indeed, Features 10 and 11, which reflect both the structure and intensity of forecast weather in the neighborhood of a route, consistently have the highest MI scores. Feature 1 also tends to perform well at shorter planning horizons, when the weather forecast on the route is subject to relatively

low spacial error. On the other hand, Features 4 and 9 outperform Feature 1 at longer planning horizons, since they reflect the neighborhood of the route (not just the weather activity on the actual route).

The above analysis provides a better understanding of how well the features of a convective weather forecast correlate with route blockage. In the next section, the selected features will be used to predict robust routes, through the use of methods from machine learning.

6.   Classification

Using the route datasets described in Sections 4 and 5, we adapt techniques from machine learning to better predict route blockage. Specifically, a classifier is trained to predict, given the features of a route in forecast weather, whether the route will be open or blocked in the actual weather that materializes. This prediction is also associated with a probability, which is determined by the performance metrics of the classifier.

6.1.   Training Objectives

When evaluating a classifier, the class predictions are compared with the actual classes of a test set, according to the standard *two-class confusion matrix*:

	Predicted Open	Predicted Blocked
Actual Open	TP (True Positive)	FN (False Negative)
Actual Blocked	FP (False Positive)	TN (True Negative)

We note the difference between these definitions and those in Section 2.2: here, a positive is defined to be an open route (i.e. a route not containing hazardous weather). This difference is a natural consequence of the move from a pixel-based to a route-based evaluation of forecast accuracy.

Although it is typically desirable to maximize the accuracy (total number of correct predictions) of a classifier on a test set, the context of aviation weather warrants a modified objective. Due to safety concerns, it is more important to correctly predict a route that ends up blocked than one that ends up open. This emphasis on correctly predicting members of the blocked class (minimizing false positives) is complicated by the fact that the dataset is imbalanced, having fewer blocked examples than open, making it inherently harder to perform well on the blocked (minority) class. As we will see in Section 6.3, there is an inherent tradeoff between the false positive rate and the accuracy, which translates into a tradeoff between increased safety and underutilized capacity.

In addition to the FP and FN rate, we compute the following (standard) performance metrics to the evaluate our classifier: $a^- = \frac{TN}{TN+FP}$, $a^+ = \frac{TP}{TP+FN}$, g-mean $= \sqrt{a^- \times a^+}$, and accuracy $= \frac{TP+TN}{n}$, where n is the total number of routes in the data set. In particular, we are interested in a^-, also known as recall, which is a measure of how well the classifier performs on members of the minority class. We seek to maximize this value through classification.

6.2. Ensemble Classifiers

The Machine Learning literature has shown that ensemble classifiers tend to perform well on imbalanced datasets, outperforming non-ensemble methods (Hulse et al. 2007, Chen et al. 2004). We trained two classifiers using the R language for statistical computing (Liu et al. 2006): an ensemble of Support Vector Machines (EnsSVM), and a weighted random forest (WRF). Two additional classifiers, namely, a (regular) SVM with an RBF kernel and a decision tree with a weighted loss function, were trained on the route blockage data set, but EnsSVM and WRF outperformed them in terms of maximizing recall. In this paper, we focus on the use of the EnsSVM classifier.

We created training and test data sets from the Data2007 route data set, by partitioning it randomly such that the training set and test set had 70% and 30% of instances, respectively. To ensure an unbiased data set, weather scenarios from the same date were not split up. The training set was then further processed when setting up the ensemble: m blocked instances of the training set were set aside, and N bootstrap samples of size m were created from the open instances. The blocked set was then combined with each of the bootstrap samples to create N training bootstrap training sets. Each of the N bootstrap sets therefore had a balance between open and blocked instances. In our experiments, $N = 11$, and m varied somewhat by planning horizon and other dataset parameters, but averaged to 55 (20% of the dataset size).

These N bootstrap samples were then used to train the two types of classifiers. EnsSVM was trained with an RBF kernel, using 5-fold cross validation to tune the parameters. The resulting ensemble classifier uses the majority vote of the ensemble to classify new routes.

6.3. Results for Ensemble SVM

Table 2 shows the results for the ensemble SVM classifier at all five planning horizons of interest, for both arrivals and departures. Results are compared against the deterministic raw weather forecast along each route, denoted by RWF; a route is open according to the RWF if each pixel along the route is open. All metrics shown are averaged over 5 runs of the classifier on the test Data2008 data set, and are hence independent of the training set. At the shorter planning horizons of 10-, 30-, and 60-minutes, the table shows that the ensemble classifier outperforms the RWF in terms of accuracy, but not recall. This situation is reversed at the longer planning horizons of 90- and 100-minutes, where there is improvement in the recall rate of the classifier at a cost to accuracy. Indeed, arrivals at 90-minutes post a 3% improvement in recall rate over the RWF, with similar cost to FP rate. At 100-minutes, the improvement in recall rate is 17%, with a slightly larger cost to accuracy. The results for departures show similar trends.

		10 min		30 min		60 min		90 min		100 min	
		EnSVM	RWF	EnSVM	RWF	EnSVM	RWF	EnSVM	RWF	EnSVM	RWF
Arrivals	Acc	88.57	82.14	81.87	77.68	75.80	76.34	76.07	77.68	58.04	72.77
	a^-	83.89	91.67	73.33	77.78	62.78	66.67	75.00	72.22	87.22	69.44
	a^+	89.47	80.32	83.51	77.66	78.30	78.19	76.28	78.72	52.45	73.4
	g-mean	0.87	0.86	0.78	0.78	0.70	0.72	0.76	0.75	0.66	0.71
	% TP	75.09	67.41	70.09	65.18	65.71	65.62	64.02	66.07	44.02	61.61
	% FP	2.59	1.34	4.28	3.57	5.98	5.36	4.02	4.46	2.05	4.91
	% TN	13.48	14.73	11.79	12.5	10.09	10.71	12.05	11.61	14.02	11.16
	% FN	8.84	16.52	13.84	18.75	18.22	18.3	19.91	17.86	39.91	22.32
Departures	Acc	90.71	79.91	81.07	77.68	77.05	74.55	69.2	73.66	68.66	74.11
	a^-	78.97	84.62	78.46	82.05	62.05	64.10	61.54	64.10	71.8	66.67
	a^+	93.19	78.92	81.62	76.76	80.22	76.76	70.81	75.68	68.00	75.68
	g-mean	0.86	0.82	0.80	0.79	0.70	0.70	0.66	0.70	0.69	0.71
	% TP	76.97	65.18	67.41	63.39	66.25	63.39	58.48	62.50	56.16	62.50
	% FP	3.66	2.68	3.75	3.12	6.61	6.25	6.70	6.25	4.91	5.80
	% TN	13.75	14.73	13.66	14.29	10.80	11.16	10.72	11.16	12.50	11.61
	% FN	5.62	17.41	15.18	19.20	16.34	19.20	24.11	20.09	26.43	20.09

Table 2 **Validation results for the Ensemble SVM classifier, compared to the raw weather forecast (RWF)**

These results follow intuition, since the recall rates of the RWF are found to be high at shorter horizons where weather forecasts are known to be more accurate, leaving little room for improvement. However, the accuracy rates at these short planning horizons can be improved, due to the wiggle room introduced when validating routes, allowing for slight shifts in the original routes, often uncovering routes around weather cells present in the area.

In summary, EnsSVM combines the features of a given weather scenario to predict route blockage, resulting in higher recall rates than the RWF at longer planning horizons. This improvement in the false positive rate comes at a tradeoff with accuracy, explained by the conservative objective function placed on the learning algorithm, and the imbalance between open and blocked routes.

This trade-off between false positive rates and accuracy can be controlled in the Weighted Random Forest classifier through the weight in the training loss function, which places an explicit penalty on misclassified blocked routes. A detailed discussion of the WRF is omitted here, but can be found in prior work (Michalek and Balakrishnan 2009b). In practice, an airspace operator could select an operating point, choosing between higher accuracy on one hand, which would mean a more aggressive strategy and hence higher throughput, and a lower false positive rate on the other, which would be a more conservative strategy with a potential under-utilization of capacity.

6.3.1.   Sensitivity of Classifier to Inner Radius, R_I: In our terminal-area model, arriving aircraft begin their final merge or the crosswind leg upon crossing the inner radius R_I. As such, R_I is meant to be a parameter which can be set depending on the specific characteristics of a given airport. Ideally, the results from Section 6.3 should not be overly sensitive to R_I. In this section we show the classifier results as R_I varies from 10 km to 30 km, keeping the wiggle room B fixed

to 8 km. This represents the range of realistic values of R_I across airports. Datasets are created for each value of R_I (note that feature values will change with R_I), for all planning horizons, and for both arrivals and departures. The EnsSVM classification algorithm is run for each dataset, and the same statistics are calculated as in Section 6.3.
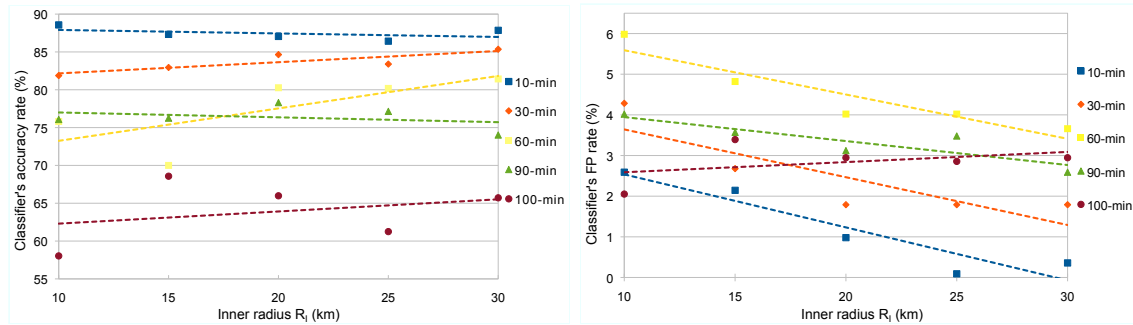

Figure 8   Sensitivity of EnsSVM to R_I, as measured in terms of accuracy (left) and false positive rates (right).

Figure 8 shows the effect that R_I (varied from 10 to 30 km) has on false positive and accuracy rates, for the case of arrivals. The accuracy rates stay fairly consistent with increasing R_I, although there is greater variability within a fixed planning horizon as R_I increases. The false positive rates tend to decrease with increasing R_I, effectively giving better performance when routes are shorter. We hypothesize that this is because shorter routes inherently have fewer potential hazards to avoid, and because a larger diameter around the airport relaxes the bottleneck for flow.

6.3.2.   Sensitivity of Classifier to Wiggle Room, B: Another parameter that was fixed in the classification results in Section 6.3 is B, the amount of wiggle room used to determine route blockage in actual weather. Although B was previously set to 8 km, it is an adjustable parameter meant to represent the maneuverability that an aircraft is allowed without having to change its original routing plan. In this section we study the sensitivity of the classification results to the parameter B.

Figure 9 contains a box plot of the classification accuracy and false positive rates as B varies between 0 km (no wiggle room allowed) and 20 km, across all planning horizons, for both arrivals and departures, and for inner radius R_I set to 10 or 20 km. The average accuracy rates tend to increase with B, while average false positive rates tend to decrease with B. These trends can be explained as follows: increased wiggle room for any given weather scenario makes it more likely that the route will be open (and the classifier learns that increased maneuverability makes routes more likely to persist), which decreases the number of potential false positives. At the same time, the imbalance between open and blocked routes results in increased accuracy when more routes are
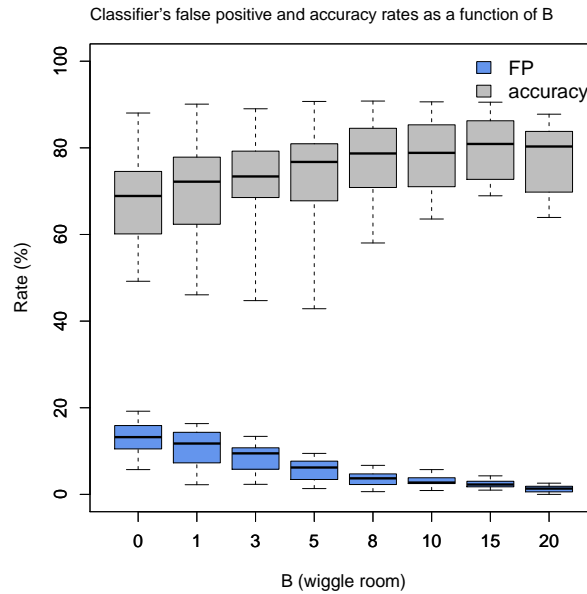
Figure 9   **Box plot of the classifier's sensitivity to the wiggle room parameter, B. The end points represent the minima and maxima of the data.**

classified as open. The results show smaller variance as B increases, since the wiggle room removes some of the randomness in the spatial prediction. It is worth noting that the improved classification performance with larger wiggle room must be traded off against a diminished capability for fine-grained planning (for example, with Required Navigation Performance or Area Navigation (RNAV) routes (Joint Planning and Development Office 2004)).

Until now, we have evaluated the performance of the EnsSVM classifier in providing a deterministic prediction of whether a route is blocked or not, but not the likelihood of a route being open, and comparing it to the deterministic, raw convective weather forecast. In the following sections, we consider the problem of deriving probabilistic predictions of route blockage from the deterministic, raw convective weather forecasts.

6.4. Probabilistic Predictions for Route Blockage

A probabilistic prediction of route blockage enables the selection of an optimal route, or most robust route, through the terminal-area. This section describes the creation and validation of probabilistic predictions that are then used in Section 7 to optimize the placement of terminal-area fixes.

The EnsSVM classifier has a natural mapping into a probabilistic prediction: the classifier consists of N (dependent) ensemble members, each of which are trained to give a separate SVM prediction of route blockage based on the features of a potential route. This binary prediction can be turned into a probabilistic prediction by fitting a logistic distribution using maximum likelihood to the decision values of the classifier. We use the mean of these N probabilities to give the

ensemble prediction. This fusion strategy (equivalent to the sum of experts) was chosen because it is a simple and effective fusion strategy (see Alkoot and Kittler (1999)) that outputs a probability rather than a binary value, and because its predictions were found to be similar to the majority vote (the fusion strategy typically used in ensemble methods).
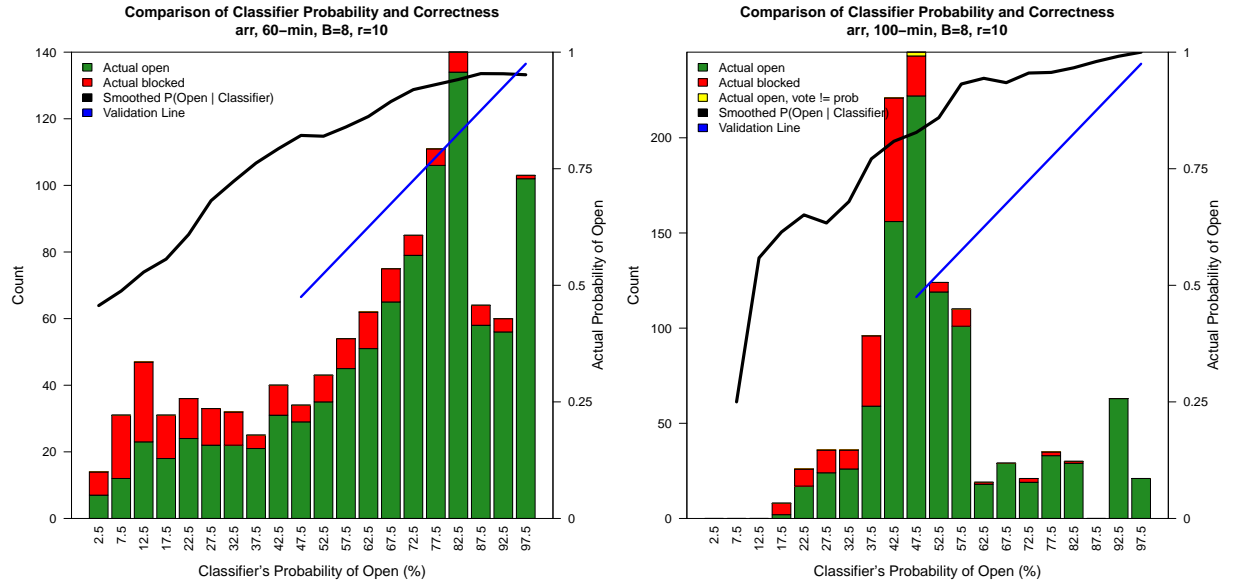

Figure 10 **Validation of classifier's probability of blockage on the test data set, for arrivals at 60-(left) and 100-minutes (right). The black trend line tends to stay above the $x = y$ validation line, indicating a conservative forecast of blockage probability.**

Figure 10 validates the mean probability fusion strategy for assigning probabilities to route blockage predictions, and shows a histogram of the (binned) probability assigned to the prediction (a probability of at least 50% means that the route is classified as open), with routes that end up blocked in red, and those that end up open in green. The fraction of open routes in each bin represents the empirical conditional probability that a route will be open given its classification probability. The black trend line is a smoothing of these empirical probabilities, performed by taking the average of each five-bin window, weighted in proportion of the number of data points in each bin. This smoothed line represents the actual probability that a route will be open given the classification probability. In each case, the smoothed probability line falls slightly above the validation line at $x = y$, indicating that the classifier's probabilistic estimate tends to be a conservative estimate of true probability. This trend is shown for specific parameter settings here, but is representative of other planning horizons and flight directions, and can be explained by the classification training function. Training placed more weight on avoiding false positives than on accuracy (which

is dominated by being correct on the open routes), thereby lowering the confidence on open routes. For this reason, we use the raw classifier probabilities to perform optimization in the next section, but note that any increasing fit of the probability data points would give the same optimization results, and would differ only in probability values.

Also noteworthy are the contrasting distributions seen in Figure 10's histograms. At the shorter planning horizon of 60-minutes (left), the histogram data is spread across the bins, with no peak around 50% (the bin with highest uncertainty), indicating consensus in the prediction. In contrast, the longer planning horizon of 100-minutes (right) has its histogram bars concentrated closer to the 50% probability bin, indicating a lower certainty amongst the ensemble members. This trend is seen across a range of parameter settings, and could be explained by the inverse relationship between forecast accuracy and planning horizon.

Having thus validated the stochastic forecasts of robust routes, we use them to optimize terminal-area fix locations.

7.   Optimization of Arrival and Departure Fixes

In the terminal-area, a flight follows a Standard Instrument Departure (SID) when departing an airport, and a Standard Terminal Arrival Route (STAR) into its destination airport. The corresponding arrival fixes, departure fixes, and terminal airspace sectorization, are fixed, even when the presence of hazardous weather renders them unusable. There is clear potential to recover lost capacity by dynamically altering the terminal airspace structure in the presence of adverse weather. This concept of relaxing the rigid structure of airspace and the resultant potential increase in airspace capacity has been identified by the NextGen ATM-Weather integration plan (Joint Planning and Development Office 2009). In this section, we adopt the concept and present an algorithm for optimally placing arrival and departure fixes, using the stochastic route robustness model developed in previous sections.

We focus on ATL for our analysis. At ATL, as at many other airports, the terminal airspace is divided into a set of arrival and departure sectors, each containing a gate of one or more fixes. Figure 11 contains a diagram of terminal airspace for ATL, which follows a four corner-post structure: there are 4 arrival sectors alternating with 4 departure sectors. The fixes (green triangles) are placed approximately 100 km away from the airport, and typically merges and landing patterns occur within a 20 km radius of the airport.
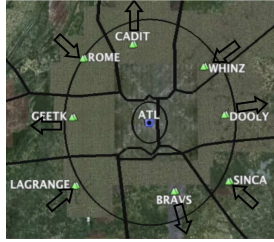
Figure 11   **ATL terminal area and fixes. The outer circle has a radius of 100 km, the inner one has a 20 km radius.**
© Google, Map Data US Navy. Image © 2009 DigitalGlobe. Image USDA Farm Service Agency.

7.1.   Algorithm for Relocation of Terminal Fixes

There are several ways in which terminal airspace can be restructured to optimize capacity in the presence of hazardous weather, with varying degrees of complexity. One method with relatively low complexity fixes the sector boundaries, and moves fix locations within each sector so as to minimize the probability that the arrival route through the fix is blocked by hazardous weather. Keeping the sector boundaries fixed limits the additional complexity and air traffic controller workload that arise from redefining sectors. In this paper, we focus on this simple version of the problem, in order to study the potential benefits of even a small amount of operational flexibility.

Since sector boundaries are fixed, each arrival and departure sector is optimized independently (Figure 12), according to the procedure below.

1. The set of potential fixes and corresponding routes is generated by sampling straight-line routes at incrementally increasing angles from the airport. Each potential route begins at R_I and ends at R_O; the intersection of the route with C_O defines the associated fix.

2. For each potential fix, the feature values for the associated route are computed, and the route is run through the classifier, giving a probability of being open (Section 6.4).

3. The output is the fix with maximum probability of being open (as long as the probability exceeds 50%). Ties are broken by picking the route that is closest to the standard fix.

Since sectors are optimized independently, and since there is an implicit B km of wiggle room in the actual route (and fix) flown, the algorithm as described does not necessarily maintain separation between adjacent routes. To ensure strict separation of at least $2B$ km, it is necessary to simultaneously solve across all sectors. Furthermore, the perturbed routes may violate sector boundaries, especially for larger values of B. However, these violations are rare in our experiments (there were no fix separation violations, and one boundary crossing violation of under 1 km, among the 224 cases studied in the next section), so for simplicity we do not consider them further.
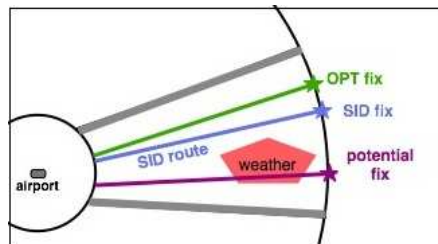
Figure 12   **Diagram of the sector optimization procedure for the case of a departure sector. Potential fixes and**
associated routes are sampled, and the optimal fix is the one with highest probability of being open.

7.2.   Analysis of Fix Relocation Algorithm

The optimization algorithm was run on the Test2008 data consisting of 28 weather scenarios. The parameters were set to $B = 8$ km, $R_I = 20$ km and $R_O = 100$ km, and the classifiers used were those from Section 6.3, trained on Train2007 data.

Table 3 shows the overall performance of the algorithm for varying planning horizons, and for arrival and departure sectors. Each row corresponds to one planning horizon and direction combination, and represents 112 data points (28 weather scenarios, each with 4 sectors per flight direction). The computed metrics reflect the effectiveness and trade-offs of the dynamic fix movement algorithm.

	Horizon t_0 (min)	Movements (%)	Fix classified blocked (%)	Fix blocked given classified blocked (%)	Potential avoidable blockage (%)	Avoided blockage (%)
Arrivals	10	21	18	50	75	60
	30	20	18	55	55	35
	60	19	22	32	48	36
	90	32	24	30	44	37
	100	34	36	23	48	40
Departures	10	22	21	43	65	61
	30	29	26	34	59	48
	60	34	32	25	72	64
	90	36	32	25	58	56
	100	48	40	24	51	42

Table 3   **Overall results for fix optimization.**

The first metric reported, movements, refers to the percentage of fixes moved, and gives a measure of how often the algorithm recommends an alternate fix. This number increases with planning horizon, and tends to be larger for departures than for arrivals.

The second metric, classified fix blockages, refers to the percentage of (original) fixes for which the classifier predicted blockage. This number increases with increasing planning horizon. Of course, a predicted fix blockage does not necessarily mean the fix will actually be blocked, and this situation is captured in the next metric, the percentage of actual blocked fixes given that the fix is predicted to be blocked. Here we find that the longer planning horizons are accompanied by lower values, indicating that the classifier's prediction is less accurate at longer planning horizons.

26

Pfeil and Balakrishnan: *Robust Terminal-Area Routing in Convective Weather*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

Potential avoidable blockage shows the percentage of predicted-blocked fixes for which the algorithm recommends an optimal fix (which is predicted to be open). We find that at shorter planning horizons, the potential to avoid blockages is predicted to be greatest, staying well above 50%, meaning that the algorithm gives an alternate routing possibility more than half the time. Finally, avoided blockages refers to the percentage of predicted-blocked fixes for which the optimal fix recommended by the algorithm is open in actual weather. This number tends to decrease with planning horizon, and tends to be higher for departures than arrivals. The gap between the last two columns gives a measure of accuracy on predicted-blocked routes, though it does not distinguish between fixes assigned a 90% probability of being open and those with 60% probability. Clearly the accuracy should depend on these probabilities, and we explore this correlation later.

	Horizon t_0 (min)	Number of Movements	Orig. & Opt. Open (OO)	Orig. Blocked, Opt. Open (BO)	Orig. Open, Opt. Blocked (OB)	Orig. & Opt. Blocked (BB)
Arrivals	10	24	15	4	0	5
	30	22	16	2	0	4
	60	21	16	0	2	3
	90	36	31	1	0	4
	100	38	33	0	0	5
Departures	10	25	18	6	1	0
	30	32	25	3	4	0
	60	38	31	3	2	2
	90	40	33	3	4	0
	100	54	47	2	3	2

Table 4   Analysis of fix movements.

Table 4 provides a closer look at the fixes that are moved to some optimal fix by the algorithm. When a fix is moved, there are four possible outcomes: the original fix and the optimal fix are both open in the observed weather (OO), both are blocked (BB), the original fix is blocked while the optimal is open (BO), or the original fix is open while the optimal is blocked (OB). Ideally, we would want that the cases where the algorithm makes a mistake in moving a fix, BB and OB, be few in number, while BO (especially) and OO be many.

The table indicates several trends. First, OO accounts for more than 63% of fix movements across all categories, while OO and BO together account for more than 76% of fix movements, indicating that the optimal fix is usually likely be at least as good as the original. A movement of a fix that turns out to be open may seem undesirable, but as the confidence in the optimal fix and associated route is greater than the original fix (since the moved fix is associated with a higher probability of being open), making it the more conservative choice with higher expected capacity. There are very few data points in the other three categories, indicating possibly large sampling error, so we only perform modest analysis of these cases. Nevertheless, at the 10-minute planning

horizon for both arrivals and departures, it is a good decision to move the fix, which is consistent with the short-term accuracy of the pixel-based forecast. Thus, tactical decisions to move fixes can be relied on, although more care and validation must be pursued at longer and more strategic planning horizons.

	Horizon t_0 (min)	% open | $p \in (0.95, 1.00]$	% open | $p \in (0.75, 0.95]$	% open | $p \in (0.50, 0.75]$
Arrivals	10	97.92	-	-
	30	94.38	-	-
	60	-	91.46	-
	90	96.30	88.90	-
	100	-	-	95.35
Departures	10	96.00	-	-
	30	97.53	-	-
	60	98.11	87.50	-
	90	91.05	88.46	-
	100	-	92.31	88.24

Table 5   **Validation results for fix optimization, as a function of the predicted probability of being open, p.**

Table 5 shows algorithm performance as a function of classifier prediction probability p, giving a sense of the accuracy of probability values given with each fix movement recommendation. Each column lists, for a set of input parameters and for a range of values of p, the empirical percentage of optimal fixes that were open in actual weather. For instance, for arrivals at a 90-minute planning horizon, among those optimal fixes (corresponding to actual weather scenarios) that were predicted to be open by the classifier with probability between 0.95 and 1.00, 96.30% of them were open in the actual weather that materialized. Blank entries correspond to cells with fewer than 13 data points, which were removed to eliminate cells with sample variance over 0.05. This procedure also eliminated cells in which all data points corresponded to open routes, resulting in a sample variance of zero. As was noted in Section 6.4, the longer the forecast planning horizon, the lower the likelihood of high-probability predictions from the classifier. This explains the uneven distribution of data points among the three probability levels studied. The table shows that the percentage of open routes tends to stay within the predicted percentage when there are enough data points. This means that decision-makers can be more confident in fixes that have a high probability of being open. The table also shows that the validation is less accurate with increased planning horizon and with decreased probability interval, as expected based on the classification results.

Thus, the predicted probabilities correlate well with actual rates of route availability, and can be used to inform fix movement decisions in marginal weather conditions.

8. Dynamic Fix Movement

In the fix relocation approach that has been described so far, the recommended fix and route are designed for aircraft that begin their flight through the terminal at a specific time t, and takes their entire trajectory (including the time at which they reach different points along their route) into account. However, as a weather scenario evolves, it is not necessary that the same fix will be appropriate for flights that arrive at the boundary later, say, at $t + 20$. In practice, it would be necessary to decide when to initiate a dynamic fix movement, and to understand the frequency and scale of any necessary adjustments, as weather evolves. This section introduces a dynamic variant of the fix relocation algorithm (DYN from now on), and describes how our approach would work in practice as weather scenarios evolve.

Given a weather scenario with starting time t, planning horizon t_0, and fixed parameter settings as before, we run the algorithm presented in Section 7.1 for times $t, t + 5, \ldots, t + 55$, using the planning horizon t_0 in each iteration. We break any ties among potential optimal fixes during period t_i by selecting the fix closest to the optimum from period t_{i-1}. The first time period uses the standard fix as a tie breaker, as in the original algorithm.
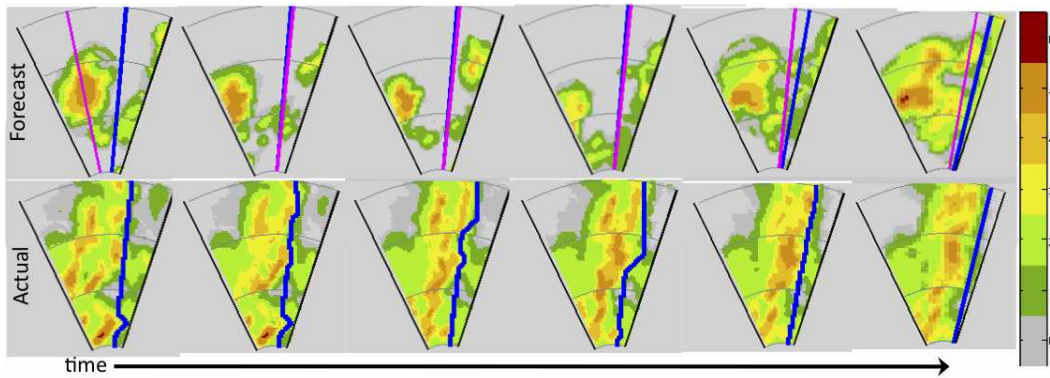

Figure 13   **Visualization of dynamic fix algorithm, showing how the fix and associated route evolve for flights entering the terminal area during six consecutive time periods, each 5 minutes long.**

Figure 7.2 shows the fix locations and associated routes recommended by DYN for one departure sector, as a weather scenario evolves over 6 consecutive periods (i.e., 30 minutes). The top row contains the (dynamic) weather forecast, with the route and fix from the previous time-period in magenta, and the new recommended optimal route and fix in blue. The bottom row contains the actual weather corresponding to each time period, as well as the perturbed route closest to the optimal in blue (indicating feasibility: note that this route is not necessarily a straight line, but avoids weather hazards). In other words, the leftmost images correspond to the fix location and

routes (shown by a blue line) assigned to flights that arrive in the first 5 minutes of the scenario, the images immediately to the right of that correspond to the fix location and routes for flights that arrive in the next 5 minutes, and so on.

This dynamic variant to the fix movement algorithm was evaluated against 28 weather scenarios with 8 sectors each, with parameters set to $t_0 = 60$ minutes, $B = 8$ km, $R_I = 20$ km, and $R_O = 100$ km. Across 12 time periods, the average sector was blocked for 1.4 periods and contained 2.4 fix movements. The average movement size across all fix movements was 9.5 degrees (equivalent 16.6 km at the outer circle) with standard deviation of 10 degrees. Moreover, the majority of fixes were highly stable. In particular, 74% of fixes during a given time period were predicted to be open with probability of 0.95 or greater. Among these fixes, 99.6% of them remained open in the subsequent time period, averaging a movement size of only 1.2 degrees (2.4 km).

We conclude that over a typical 1-hour period, fixes tend to remain open and stable, especially once a high-probability fix is identified. This dynamic approach is also applicable to TRACON operations for non-metroplex airports, which have the type of simple arrival and departure routes studied. Moving fixes during severe weather scenarios has the potential to increase terminal capacity and to improve controller workload as far as dealing with pilot deviations, since target fixes would be more likely to be flyable. However, further studies on when to initiate dynamic fix movement, and the effect on controller complexity are needed before this approach can be adopted in practice.

9.   Conclusion

This paper presented an approach for identifying routes that are likely to be robust to the inaccuracies of convective weather forecasts. In contrast to prior efforts which assumed the presence of accurate capacity forecasts as input, the proposed approach evaluated different features of weather forecasts, and selected ones that had high correlation with route blockage in observed weather. These features were then utilized in classification algorithms based on machine learning techniques to predict, given a set of potential routes and a weather forecast, which routes were likely to be blocked in the observed weather. The performance of the proposed classifiers was evaluated and compared to the raw forecast predictions, using several metrics including the false positive rate and the overall accuracy. It was shown that classifiers can be optimized to minimize the FP rate (important for aviation applications), and the trade-offs between overall accuracy and the FP rate were illustrated. The route robustness model was also applied to the problem of dynamically restructuring terminal area fixes without changing current sector boundaries. The potential benefits of fix relocation were discussed, using the ATL terminal area as a case study.

30

Pfeil and Balakrishnan: *Robust Terminal-Area Routing in Convective Weather*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

Acknowledgments

This research was supported by NASA under the NGATS-ATM Airspace Program (NNA06CN24A) and by NSF (ECCS-0745237). The authors are grateful to Marilyn Wolfson, Rich De Laura, Mike Matthews, Kim Calden, and Mike Robinson at MIT Lincoln Lab for help with CIWS data and fruitful discussions, and to John Robinson (NASA) for his interest in this work and insightful comments.

References

Ahuja, Ravindra K., Thomas L. Magnanti, James Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall.

Alkoot, F. M., J. Kittler. 1999. Experimental evaluation of expert fusion strategies. *Pattern Recogn. Lett.* **20**(11-13) 1361–1369.

Bertsimas, D., D. B. Brown, C. Caramanis. 2007. Theory and applications of robust optimization. Preprint.

Bertsimas, D., A. Odoni. 1997. A critical survey of optimization models for tactical and strategic aspects of air traffic flow management. Tech. rep., NASA Ames Research Center.

Bertsimas, D., S. Stock Patterson. 2000. The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Science* **34** 239–255.

Brown, Barbara G., Randy B. Bullock, Christopher A. Davis, Jason Halley Gotway, Michael B. Chapman, Agnes Takacs, Eric Gilleland, Kevin Manning, Jennifer L. Mahoney. 2004. New verification approaches for convective weather forecasts. *11th Conference on Aviation, Range, and Aerospace*.

Bureau of Transportation Statistics. 2008. Understanding the reporting of causes of flight delays and cancellations. `http://www.bts.gov/help/aviation/html/understanding.html`.

Chatterji, Gano B., Brett Gyarfas. 2006. Convective weather forecast quality metrics for air traffic management decision-making. *6th AIAA Aviation Technology, Integration and Operations Conference*.

Chen, Chao, Andy Liaw, Leo Breiman. 2004. Using random forest to learn imbalanced data. Tech. rep., Department of Statistics, UC Berkeley. `http://www.stat.berkeley.edu/users/chenchao/666.pdf`.

DeLaura, R., S. Allan. 2003. Route selection decision support in convective weather: A case study. *USA/Europe Air Traffic Management R&D Seminar, Budapest*.

DeLaura, R. A., M. Robinson, M. L. Pawlak, J. E. Evans. 2008. Modeling convective weather avoidance in enroute airspace. *13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM)*. Amer. Meteor. Soc.

Evans, J.E., M.E. Weber, M.M. Wolfson, D.A. Clark, O.J. Newell. 2009. Roadmap for weather integration into traffic flow management modernization. Tech. rep., MIT Lincoln Laboratory.

Hulse, Jason Van, Taghi M. Khoshgoftaar, Amri Napolitano. 2007. Experimental perspectives on learning from imbalanced data. *ICML '07: Proceedings of the 24th international conference on Machine learning*. ACM, New York, NY, USA, 935–942.

Joint Economic Committee, Unites States Senate. 2008. Your flight has been delayed again. Tech. rep.

Joint Planning and Development Office. 2004. Next Generation Air Transportation System Integrated Plan.
http://www.jpdo.aero/integrated_plan.html.

Joint Planning and Development Office. 2009. *Joint Planning and Development Office ATM-Weather Integration Plan*, 0th ed.

Kay, Michael P., Jennifer L. Mahoney, Joan E. Hart. 2006. An analysis of collaborative convective forecast product performance for the 2005 convective season. *12th Conference on Aviation Range and Aerospace Meteorology*. American Meteorological Society.

Klein, Alexander, Parimal Kopardekar, Mark D. Rodgers, Hong Kaing. 2007. "airspace playbook": Dynamic airspace reallocation coordinated with the national severe weather playbook. *7th AIAA Aviation Technology, Integration and Operations Conference*.

Kopardekar, Parimal, Karl Bilimoria, Banavar Sridhar. 2009. Initial concepts for dynamic airspace configuration. *7th AIAA Aviation Technology, Integration and Operations Conference*.

Krozel, J., C. Lee, J. S. B. Mitchell. 2006. Turn-constrained route planning for avoiding hazardous weather. *Air Traffic Control Quarterly* **14**(2) 159–182.

Krozel, J., J. S. B. Mitchell, V. Polishchuk, J. Prete. 2007. Capacity estimation for airspaces with convective weather constraints. *AIAA Guidance, Navigation and Control Conference and Exhibit, Hilton Head, South Carolina*.

Liu, Pei-Chen Barry, Mark Hansen, Avijit Mukherjee. 2008. Scenario-based air traffic flow management: From theory to practice. *Transportation Research Part B: Methodological* **42**(7-8) 685–702.

Liu, Yang, Aijun An, Xiangji Huang. 2006. Boosting prediction accuracy on imbalanced datasets with SVM ensembles. *Lecture Notes in Computer Science*, vol. 3918. Springer Berlin / Heidelberg.

Mahoney, Jennifer Luppens, Stacey Seseske, Joan E. Hart, Mike Kay, B.G. Brown. 2004. Defining observation fields for verification of spacial forecasts of convection: Part 2. *11th Conference on Aviation, Range, and Aerospace*.

Menon, P. K., G. D. Sweriduk, T. Lam, G. M. Diaz, K. Bilimoria. 2006. Computer-aided Eulerian air traffic flow modeling and predictive control. *AIAA Journal of Guidance, Control and Dynamics* **29** 12–19.

Michalek, D., H. Balakrishnan. 2009a. Building a stochastic terminal airspace capacity forecast from convective weather forecasts. *Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration, Phoenix, AZ*.

Michalek, D., H. Balakrishnan. 2009b. Identification of robust routes using Convective Weather Forecasts. *USA/Europe Air Traffic Management R&D Seminar, Napa, CA*.

Mitchell, J. S. 1988. On maximum flows in polyhedral domains. *SCG '88: Proceedings of the fourth annual symposium on Computational geometry*. 341–351.

Mitchell, J. S. B., V. Polishchuk, J. Krozel. 2006. Airspace throughput analysis considering stochastic weather. *AIAA Guidance, Navigation, and Control Conference, Keystone, Colorado.*

Prete, J., J. S. B. Mitchell. 2004. Safe routing of multiple aircraft flows in the presence of time-varying weather data. *AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, Rhode Island.*

Seseske, S. A., M. P. Kay, S. Madine, J. E. Hart, J. L. Mahoney, B. Brown. 2006. *2006: Quality Assessment Report: National Convective Weather Forecast 2 (NCWF-2).* Submitted to FAA Aviation Weather Technology Transfer (AWTT) Technical Review Panel.

Song, Lixia, Daniel Greenbaum, Craig Wanke. 2009. The impact of severe weather on sector capacity. *Eighth USA/Europe Air Traffic Management Research and Development Seminar.*

Sun, D., S. D. Yang, I. Strub, A. M. Bayen, B. Sridhar, K. Sheth. 2006. Eulerian trilogy. *Proceedings of the AIAA conference on Guidance, Navigation and Control.*

Weygandt, Stephen S., Stanley G. Benjamin. 2004. Ruc model-based convective probability forecasts. *11th Conference on Aviation, Range, and Aerospace.*

Wolfson, M., B. Forman, K. Calden, W. Dupree, R. Johnson, R. Boldi, C. Wilson, P. Bieringer, E. Mann, J. Morgan. 2004. Tactical 0-2 hour convective weather forecasts for FAA. *11th Conference on Aviation, Range and Aerospace Meteorology, Hyannis, MA.*