AIAA Modeling and Simulation Technologies Conference
08 - 11 August 2011, Portland, Oregon

AIAA 2011-6536

Convective Weather Avoidance Modeling in Low-Altitude Airspace†

Scot E. Campbell[1] and Rich A. DeLaura[2]
MIT Lincoln Laboratory, Lexington, MA 02420

Thunderstorms are a leading cause of delay in the National Airspace System (NAS), and significant research has been conducted to predict the areas pilots will avoid during a storm. An example of such research is the Convective Weather Avoidance Model (CWAM), which provides the likelihood of pilot deviation due to convective weather in a given area. This paper extends the scope of CWAM to include low-altitude flights, which typically occur below the tops of convective weather and have slightly different operational constraints. In general, the set of low-altitude flights includes short-hop routes and low-altitude escape routes used to reduce the impact of convective weather in the terminal area. This paper will discuss the classification procedure, present the performance of low-altitude CWAM on observed and forecasted weather, analyze areas of poor performance, and suggest potential improvements to the model.

I. Introduction

CONVECTIVE weather is a significant impediment to effective and efficient Air Traffic Management (ATM) decisions, and sometimes results in unnecessary delays to the National Airspace System (NAS). In the NAS, 70% of delays are caused by weather, and of those delays, 60% are specifically accounted for by convective weather [1]. Currently, rerouting decisions made by air traffic managers are aided by weather products such as the Corridor Integrated Weather System (CIWS) and the National Convective Weather Forecast (NCWF) [2, 3]. In a Next Generation ATM system, decision support tools such as the Route Availability Planning Tool (RAPT) will mitigate weather-induced delays by supplementing the situational awareness of an air traffic manager with a forecast of the availability of specific flight routes [4]. RAPT is based on the Convective Weather Avoidance Model (CWAM), which is a probabilistic model of pilot decision making in the presence of convective weather [5].

CWAM is a tool originally developed for the en route flight regime to predict pilot deviation decisions by correlating in-flight deviations of aircraft to the weather features they encounter. The model is based on a database comprised of the deviation decision of each flight and weather statistics along each route, which are obtained from CIWS. Pattern classification experiments on the en route CWAM database show that the most descriptive predictors for deviation are related to echo top height, where the most descriptive is the *difference* in altitude between the aircraft and the echo top height [5]. In the terminal area, deviations are predicted with a different set of features. Several studies of the Dallas and Memphis areas using weather information from the Integrated Terminal Weather System (ITWS) show that deviation decisions are closely related to the radar intensity of the storm and the proximity of the aircraft to the airport [6, 7].

This paper presents the development of a low-altitude version of CWAM which is based on a database composed of weather encounters that occur during level flight between FL100 and FL240. This model is applicable to jet traffic that uses low altitude air routes to 'escape' from terminal areas when weather or volume congestion impacts lead to constraints on high-altitude airspace, or to low-altitude flight by regional jets on 'short hop' routes. Such traffic is common in major metroplex airspaces. In this analysis, flight trajectories are obtained from the Enhanced Traffic Management System (ETMS) database, and weather data are acquired from CIWS for 23 convective weather days across two geographical regions (Chicago and New York). A Gaussian classifier is used to determine a set of deviation predictors and the results are tested on observed and forecasted data. The predictor performance is compared to the existing terminal departure CWAM used in RAPT, and the differences are discussed.

† This work was sponsored by the Federal Aviation Administration under Air Force Contract No. FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

[1] Technical Staff, Weather Sensing Group, 244 Wood Street, Lexington, MA.
[2] Technical Staff, Weather Sensing Group, 244 Wood Street, Lexington, MA.

II. Methods

This section describes the database and classifier training used to develop low-altitude CWAM. The database is comprised of low-altitude flights which encounter weather, where a low-altitude flight is defined as a flight which achieves level flight at or below FL240, and does not climb above FL240 within 20 minutes of departure. Extremely low-altitude flights (< 10k ft) and flights involving light aircraft are excluded from the database. In addition, flights that make a decision to deviate while climbing or descending are not included in the database. The database contains an entry for each flight and includes statistics on the type and severity of weather encountered as well as whether or not the flight deviated. Figure 1 is an example of a flight that deviates around weather on a route between Chicago and Cincinnati. The magenta line represents the flight plan of the aircraft, the blue line is the actual flight path, and the weather is shown as contours of Vertically Integrated Liquid (VIL), which is a measure of precipitation intensity.

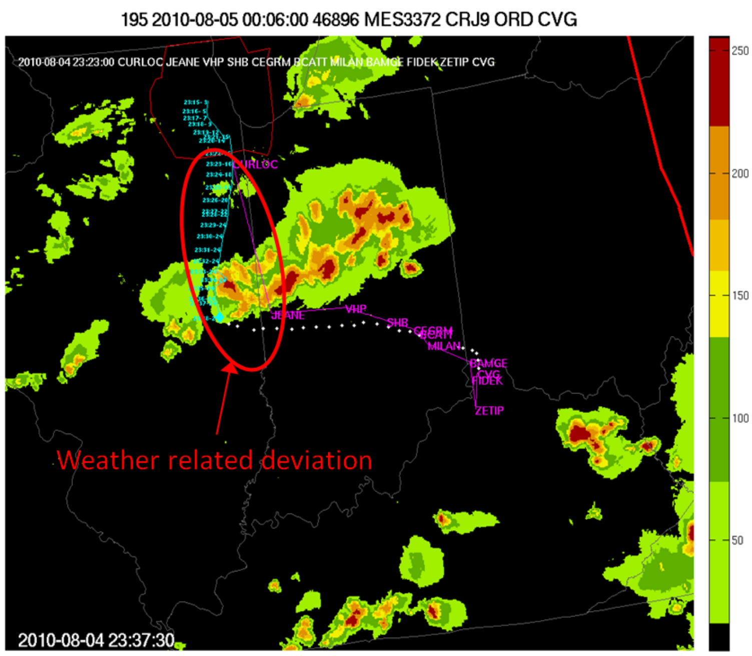


Figure 1. In-flight deviation around convective weather

Flight trajectory data are obtained from the Enhanced Traffic Management System (ETMS) database. The ETMS data provide the three dimensional position of each flight and a list of navigation fixes that describe the flight plan of each flight. Weather data are acquired from the CIWS archive, and weather characteristic fields are created and used as deviation predictors in a classification experiment.

The original CWAM development employs an automated process to separate deviations from non-deviations by comparing the distance between the actual and planned trajectories of a flight to a “deviation threshold”. In the en route environment this is an acceptable strategy because flights rarely stray from their planned route. Outside of the terminal area, low-altitude airspace is generally more flexible than en route airspace, allowing ATC to more frequently assign in-flight shortcuts to aircraft. Figure 2 is an example of a shortcut given to an aircraft on a flight from Chicago to Cincinnati. The white dots represent the flown trajectory and the magenta line is the flight plan. The shortcut allows the aircraft to fly a more direct route to its destination. The high frequency of non-weather related deviations in low-altitude airspace makes automated detection difficult and therefore the deviation analysis in this paper is performed manually.

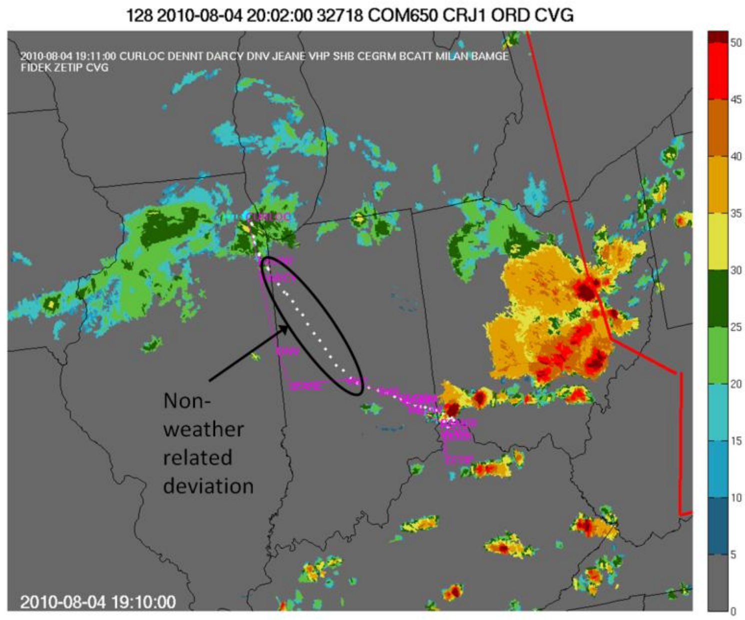


Figure 2. In-flight non-weather related deviation

Once a deviation is identified, the weather characteristics responsible for the deviation are recorded on the planned flight path at the time of the deviation decision. In addition, a non-deviation is recognized as a flight that penetrates weather with VIL ≥ 1 and does not deviate. In this case, the weather features are recorded at the point along the flight path where the flight encounters the highest VIL. The low-altitude database contains flights from 23 days and two regions. The regions consist of Chicago and New York, and the airports are Chicago O'Hare (ORD), Midway (MDW), New York LaGuardia (LGA), John F. Kennedy (JFK), and Newark (EWR). The database is partitioned into training and testing databases. Table 1 lists the number of deviations and non-deviations in each region in the training database. The total number of flights in the training database is 2539, where 1248 of the flights encountered weather and 309 flights deviated because of the weather. It should be noted that "serial deviations", where a flight deviated more than once, are not recorded as multiple deviations. Additionally, weather encounters that occur before or after a deviation are not recorded as multiple encounters.

Table 1. Low-altitude CWAM training database

	ORD, MDW		JFK, LGA, EWR	
Date	Deviations	Non-Deviations	Deviations	Non-Deviations
06/08/2009	43	91	--	--
06/09/2009	--	--	34	133
06/13/2009	--	--	42	122
06/19/2009	33	64	--	--
08/10/2009	--	--	38	51
04/07/2010	23	117	--	--
06/01/2010	--	--	48	100
06/04/2010	17	101	--	--
08/04/2010	31	160	--	--
Total	147	533	162	406

Table 2 lists the days and the corresponding numbers of trajectories in the testing database. The total number of flights in the testing database is 3647, where 1319 of the flights encountered weather and 319 flights deviated because of the weather.

Table 2. Low-altitude CWAM testing database

	ORD, MDW		JFK, LGA, EWR	
Date	Deviations	Non-Deviations	Deviations	Non-Deviations
03/11/2010	7	124	--	--
03/13/2010	--	--	1	117
05/03/2010	31	25	--	--
05/04/2010	--	--	24	112
05/07/2010	18	90	--	--
05/12/2010	11	45	--	--
05/14/2010	--	--	15	65
05/21/2010	34	82	--	--
05/26/2010	25	43	--	--
05/27/2010	--	--	26	41
06/03/2010	--	--	20	48
06/06/2010	--	--	44	100
06/11/2010	43	57	--	--
06/12/2010	20	51	--	--
Total	189	517	130	483

Table 3 presents the weather features which describe possible weather metrics that could influence a pilot's decision to deviate around a storm. The weather features are based on intuition formed from previous work [4-7]. The kernel size is the side-length of the square spatial filter applied at each grid point of the data. For example, VIL8(x,y) is the 90th percentile VIL value in an 8 x 8 km square centered at the grid point (x,y). The variance characteristics are calculated over an 8 km kernel, and in the case of echo tops, the data are pre-processed to exclude values less than 30,000 ft.

Table 3. Set of weather features

VIL1 (90th Percentile Precipitation Intensity, 1km kernel)	VIL8 (90th Percentile Precipitation Intensity, 8km kernel)	VIL16 (90th Percentile Precipitation Intensity, 16km kernel)
ET1 (90th Percentile Echo Top Height, 1km kernel)	ET8 (90th Percentile Echo Top Height, 8km kernel)	ET16 (90th Percentile Echo Top Height, 16km kernel)
VILVAR (90th Percentile VIL Variance, 8km kernel)	ETVAR (90th Percentile Echo Top Height Variance, 8km kernel)	VILCOV (Area Percent Coverage with VIL $\geq$ 3, 16km kernel)
VILpVAR (VIL1 + Maximum VIL Variance, 8km kernel)	ETpVAR (ET1 + Maximum Echo Top Height Variance, 8km kernel)	

The low-altitude training database is input to a Gaussian classifier that uses a diagonal covariance matrix and a linear discriminant function. This is the same technique used in previous work [5]. The classifier finds the combination of predictors that minimize the overall classification error. In addition, it finds the corresponding separating hyperplane that defines the boundary between the deviation and non-deviation spaces. The output from the classification experiment is a set of "best" predictors and combinations of predictors that are used in a series of modeling experiments to confirm their relative performance. The database is then partitioned into histogram bins defined for the best set of predictors, and the observed probability of deviation is found for each bin. The probability of deviation bins are filled out and smoothed using a discretized smoothing spline technique based on the discrete cosine transform [9]. The resulting smoothed tables are tested as candidate WAFs, and the performance of the predictors are compared.

III. Results

A. Classifier Training

The classifier is trained on the training dataset presented in Table 1 and with the weather features listed in Table 3. Figure 3 shows the total prediction error of the classifier computed with different feature sets. The error bars show one standard deviation variation from the total prediction error. It is apparent from Fig. 3 that the most important weather feature sets for deviation prediction explicitly include VIL1, VIL8, or VIL16. The echo tops alone do not appear to be a good predictor of deviation. The best predictor of deviation is the set {VIL1, VILpVAR}, but nine other feature sets are within one standard deviation of the minimum total prediction error. Interestingly, the single feature VIL predictors outperform some of the multidimensional feature sets.

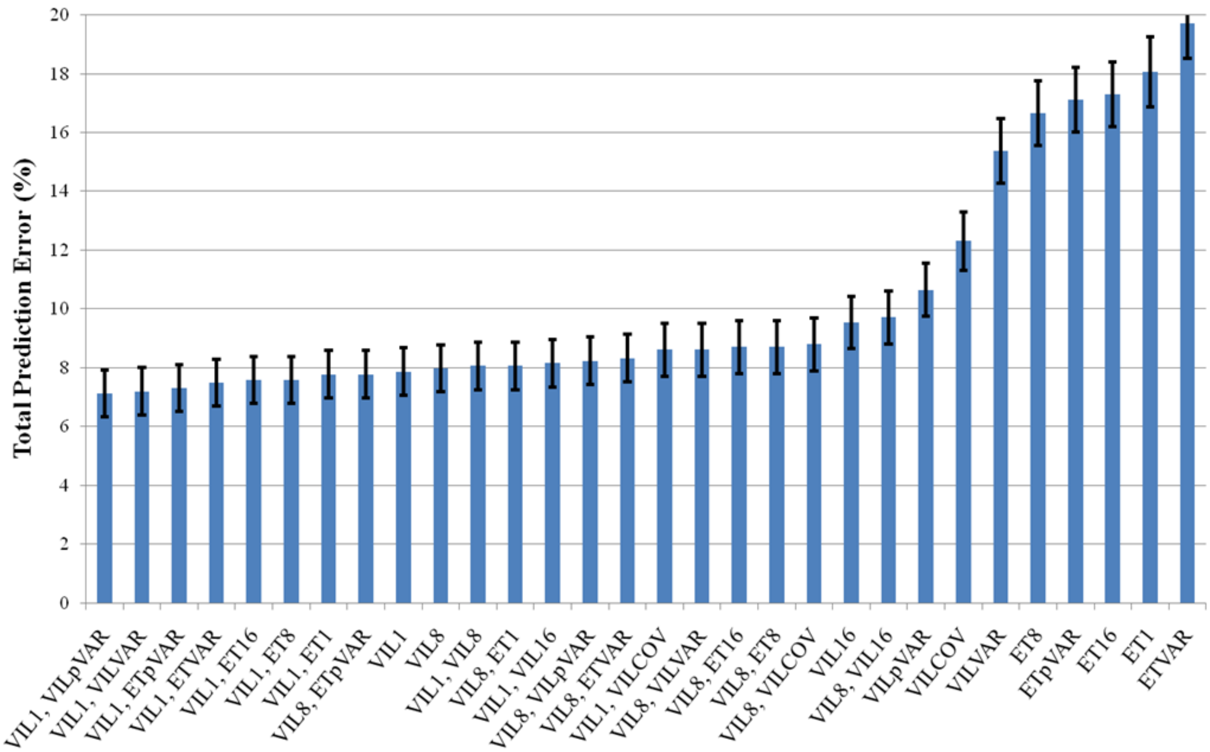
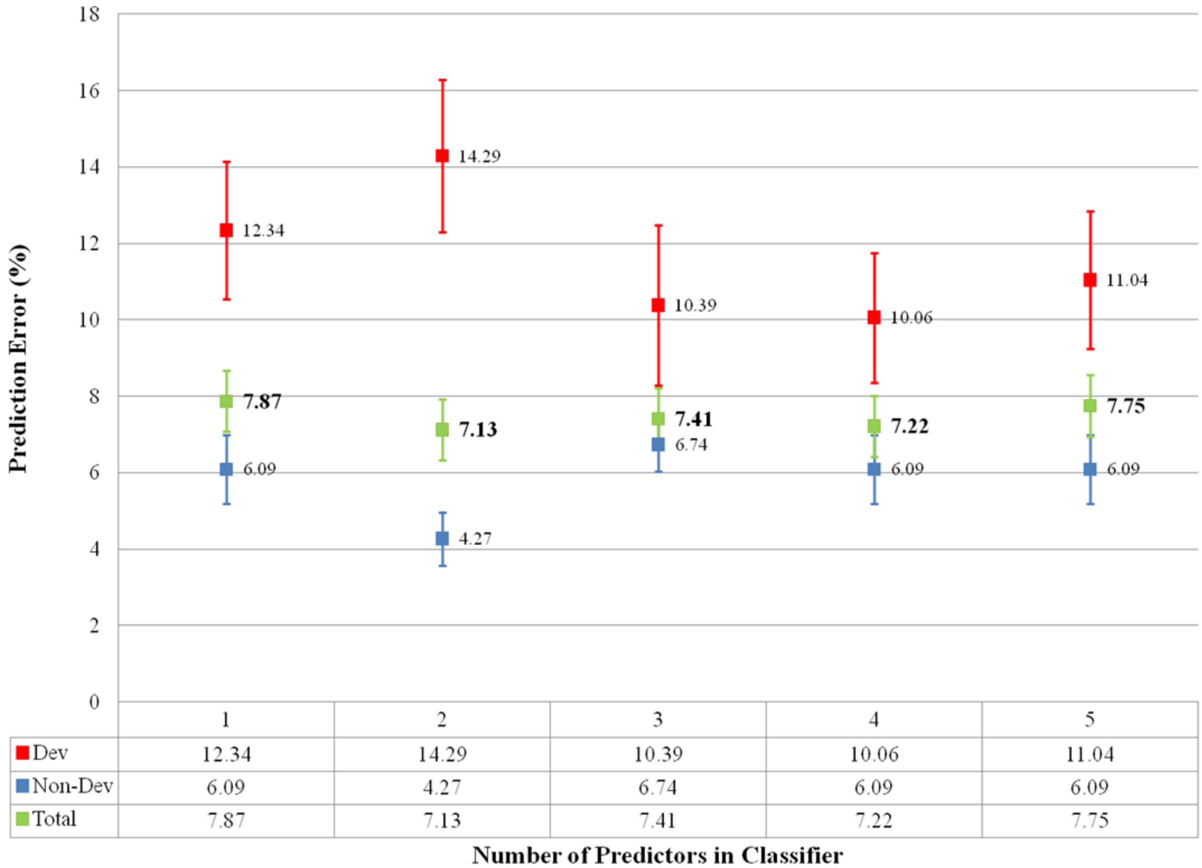


Figure 3. Classifier performance for different weather features

To better understand the relative differences between predictors with a different number of features, a "best" classifier is selected from each N-feature predictor and the results are compared. Figure 4 lists the "best" predictor(s) for each N-feature classifier and shows a comparison of the deviation and non-deviation prediction errors. The total prediction error, deviation error, and non-deviation error are given as the green, blue, and red squares, respectively. Deviation error is defined as the number of misclassified deviations divided by the number of deviations, and non-deviation error is defined as the number of misclassified non-deviations divided by the number of non-deviations. The relationship between deviation and non-deviation error provides insight into whether the classifier is under or over-predicting deviations. In many cases, two classifiers can have similar total prediction error but vastly different deviation or non-deviation errors. In Fig. 4, the differences in total prediction error for the "best" N-feature classifiers are not statistically significant, but the differences in deviation/non-deviation error spread are significant. The 2-feature classifier exhibits lower non-deviation error and higher deviation error than the other "best" predictors.

Without explicitly assigning weightings to deviation/non-deviation error spread and total prediction error, it is hard to settle on the "best" classifier. Also, the complexity of the classifier increases with the number of predictors in the classifier, which is important because high-dimensional classifiers typically require larger training datasets than low-dimensional classifiers. For this paper, additional weather features in the classifier do not result in a statistically significant increase in performance, therefore the 1-feature classifiers (VIL1, VIL8, and VIL16) are further analyzed to gain a better understanding of their performance.



1 Feature	VIL1
2 Feature	VIL1, VILpVAR
3 Feature	VIL1, VILpVAR, ETVAR
4 Feature	VIL1, VILpVAR, VIL8, ET1
5 Feature	VIL1, VILpVAR, VIL8, ET1, ET16

Figure 4. Comparison of deviation prediction error and non-deviation error for classifiers with different number of predictors.

Figure 5 presents the smoothed WAF tables for the 1-feature classifiers. Weather encounters are divided into 9 equally-spaced bins based on the value of maximum VIL (0-255) as determined by each model's spatial filter. The probability of deviation is calculated by the ratio of deviations to non-deviations inside each bin. In Fig. 5, the vertical lines show the partitions of the 6-level VIP scale, and the color indicates the probability of deviation.

A deviation is predicted when the maximum WAF value along the flight plan of an individual flight is greater than the pre-specified value of the WAF threshold. If a flight is observed to deviate in the database and the maximum WAF value along the flight plan is less than the WAF threshold, the encounter is termed a missed deviation. If a flight does not deviate and the maximum WAF along the flight plan is greater than the WAF threshold, the encounter is labeled a false deviation.

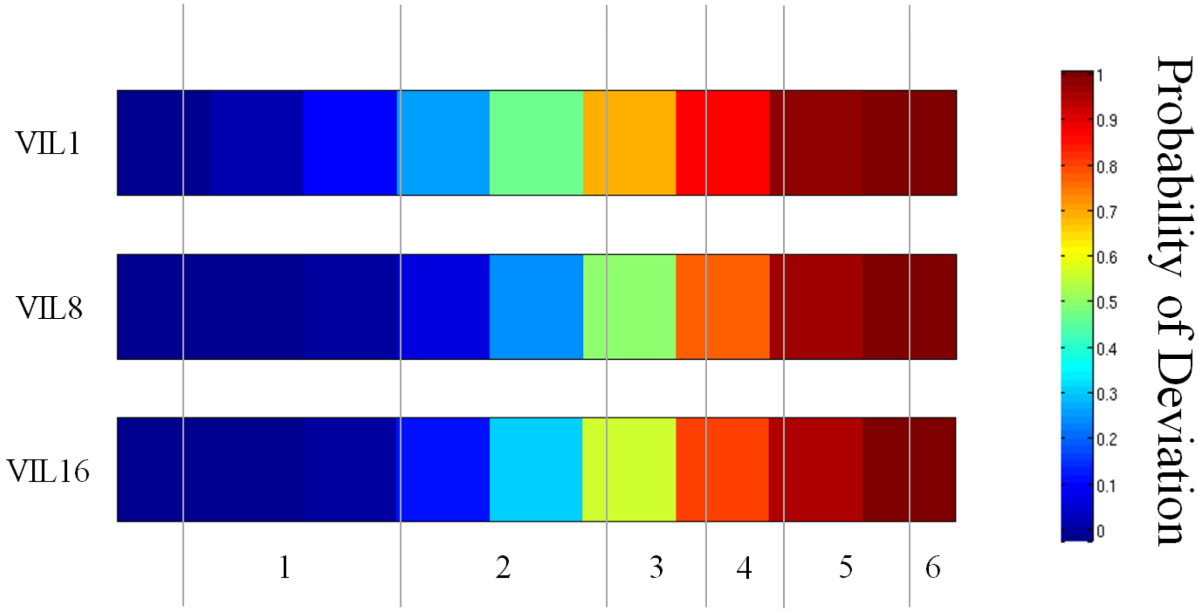


Figure 5. WAF lookup table for three 1-predictor models with different spatial filter size.

Figure 6 shows the relative performance of the 1-feature models and the existing terminal area departure WAF currently used in RAPT in terms of the probability of detection, false alarm rate, and critical success index. Probability of detection, false alarm rate, and critical success index are calculated as a function of WAF deviation threshold. The probability of detection (*PoD*) is shown in Eq. 1, where *hits* are correct predictions of deviation and *misses* are missed deviations.

$$PoD = \frac{hits}{hits + misses} \tag{1}$$

The false alarm rate (*FAR*) is given in Eq. 2, where *false* is the number of false deviations.

$$FAR = \frac{false}{hits + false} \tag{2}$$

The critical success index (*CSI*) is a measure of the overall skill of the predictor and is given in Eq. 3.

$$CSI = \frac{hits}{hits + misses + false} \tag{3}$$

The red, blue, and black lines represent the performance of the model applied to the Chicago, New York, and combined training databases, respectively. The green line shows the performance of the current departure WAF in RAPT when tested on the combined training database.

The most apparent observations from Fig. 6 are the qualitative differences in the performance curves between the Chicago and New York datasets and the statistically significant improvement in maximum CSI compared to the current RAPT departure WAF. It is interesting that the WAF threshold for maximum CSI is much lower for the RAPT departure WAF compared to the low-altitude WAF models developed in this paper. This implies that the RAPT departure WAF is under-predicting deviations in the low-altitude flight regime. In other words, flights in the low-altitude regime deviate around less severe weather than initially expected in the RAPT development. Additionally, the RAPT departure WAF does not perform as well in the tradeoff between probability of detection and false alarm rate. A good way to qualify the best tradeoff between probability of detection and false alarm rate is to see which data points are closest to the top left corner of the figure (PoD = 1.0, FAR = 0.0). All three predictors (VIL1, VIL8, VIL16) show a strong “kink” in the PoD vs. FAR curves which indicates there is a clear choice for the best WAF threshold.

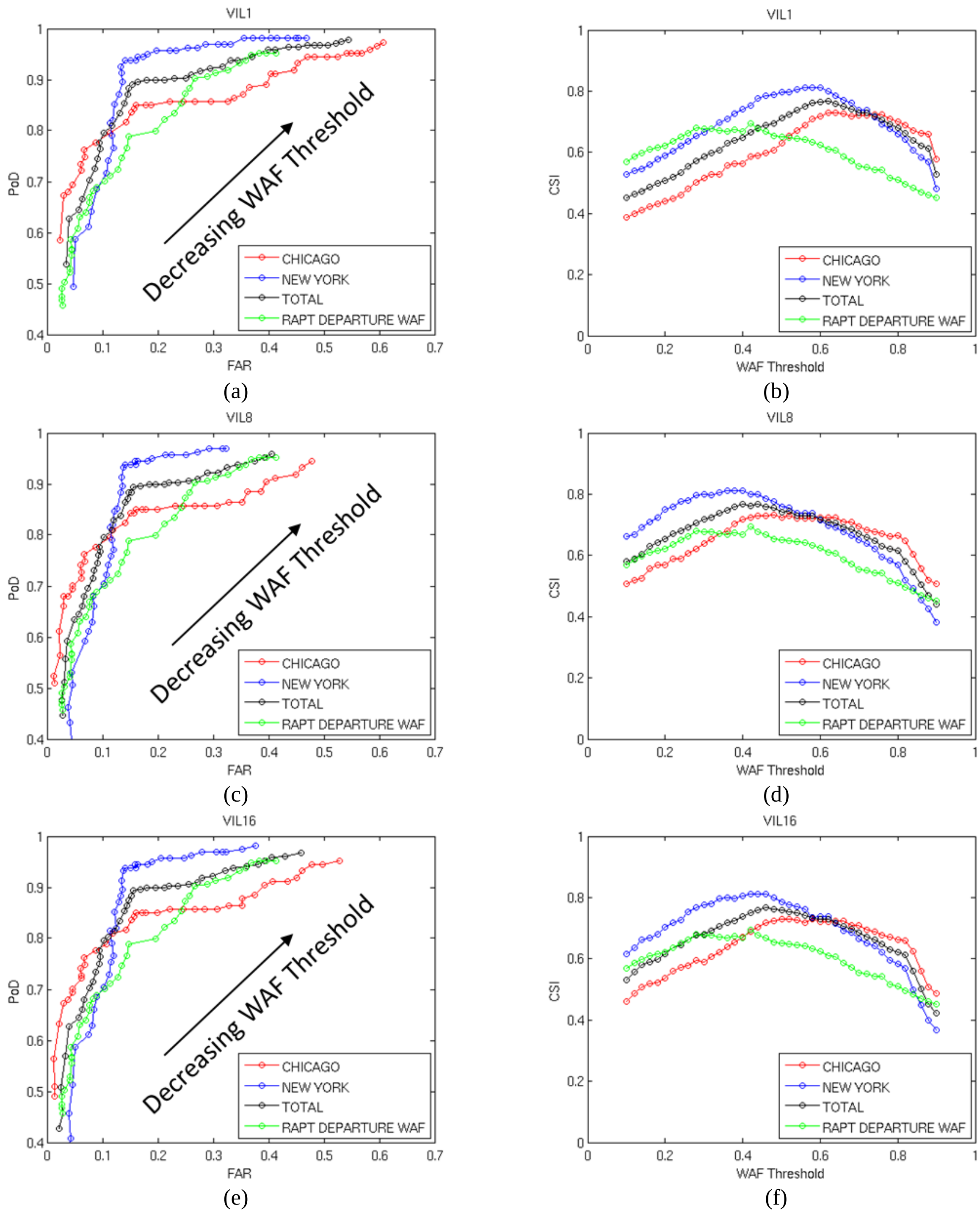
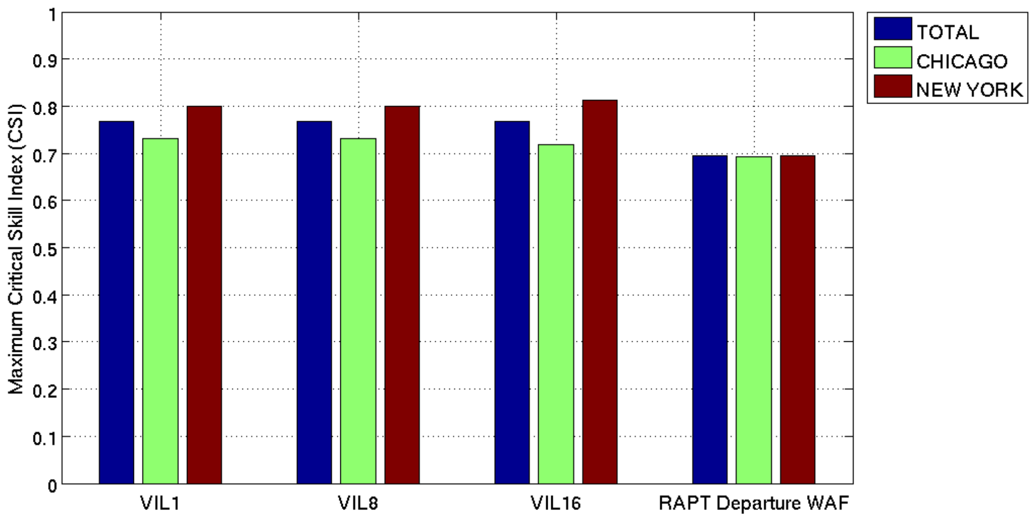
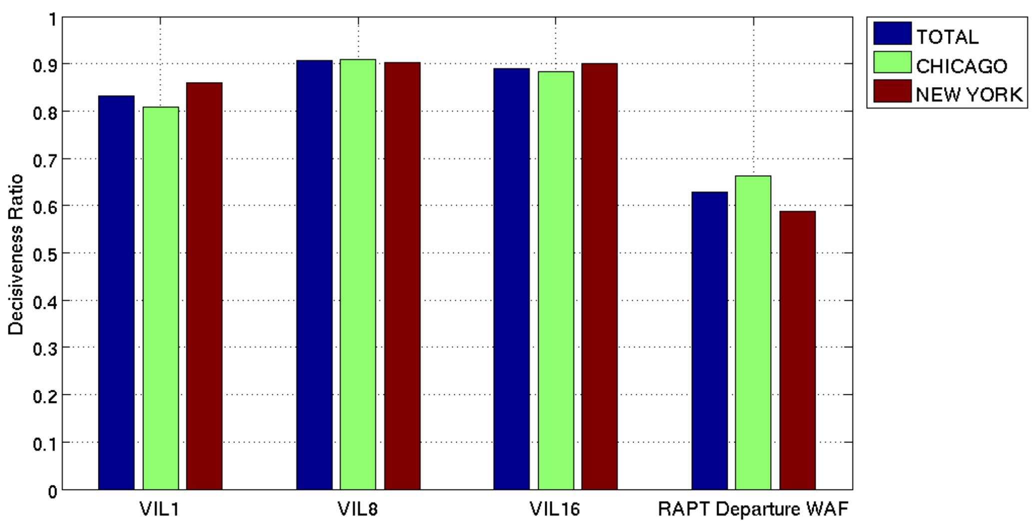


Figure 6. Predictor performance (probability of detection vs. false alarm rate and critical skill index vs WAF threshold) for predictors of different spatial filter size. Top (a,b): 1km kernel. Middle (c,d): 8km kernel. Bottom (e,f): 16km kernel.

The maximum CSI for the predictors is compared in Fig. 7a. The maximum CSI scores for the VIL1, VIL8, and VIL16 predictors are statistically identical on the total dataset, but are more than one standard deviation better on the New York dataset compared to the Chicago dataset. The RAPT departure WAF is not statistically different on the New York and Chicago datasets, but shows a more than one standard deviation decrease in performance on the total dataset when compared to the VIL1, VIL8, and VIL16 predictors.



(a)



(b)

Figure 7. Comparison of maximum CSI (a) and decisiveness ratio (b) for the classifiers analyzed in this paper.

Figure 7b shows the decisiveness ratio of the predictors, where the decisiveness ratio is the fraction of flights which encounter WAF values greater than 70% and less than 30%. The decisiveness ratio is calculated with Eq. 4, where $N_{enc>70\%}$ is the number of encounters which penetrate a WAF contour greater than 70%, $N_{enc<30\%}$ is the number of encounters which penetrate a WAF contour less than 30%, and N_{tot} is the total number of encounters.

$$Decisiveness\ Ratio = \frac{N_{enc>70\%} + N_{enc<30\%}}{N_{tot}} \tag{4}$$

Generally speaking, the decisiveness ratio gives a sense of the fraction of flights that can be identified as either a deviation or non-deviation with a high level of confidence. The most decisive predictor is VIL8, followed by VIL16, VIL1, and lastly the RAPT departure WAF. Decisiveness is an important metric because it is a measure of the certainty of the predictor. For example, a highly certain predictor enhances situational awareness in ATM by providing “yes/no” advice on route blockage instead of “maybe”.

B. Low-Altitude CWAM Performance Evaluation

Uncertainty in pilot decision modeling during convective weather involves the convolution of two sources of error: uncertainty in the pilot decision and uncertainty in the weather forecast. This section will analyze the performance of the classifiers by applying the set of classifiers to actual (observed) and forecasted weather data from the testing dataset. The effect of weather forecast uncertainty is inferred by comparing the model with observed data to the model with forecasted data. Lastly, the testing database is partitioned by geographical region to determine the sensitivity of the model to different airspaces.

The performance of the classifiers is evaluated on the testing database presented in Table 2. The database includes observed and forecasted data from 14 days in 2010 which encompass a wide range of severe weather. The total number of flights in the testing database is 3647, where 1319 of the flights encountered weather and 319 flights deviated because of the weather. Figure 8 shows the performance of the predictors in terms of probability of detection and false alarm rate for different forecast periods.

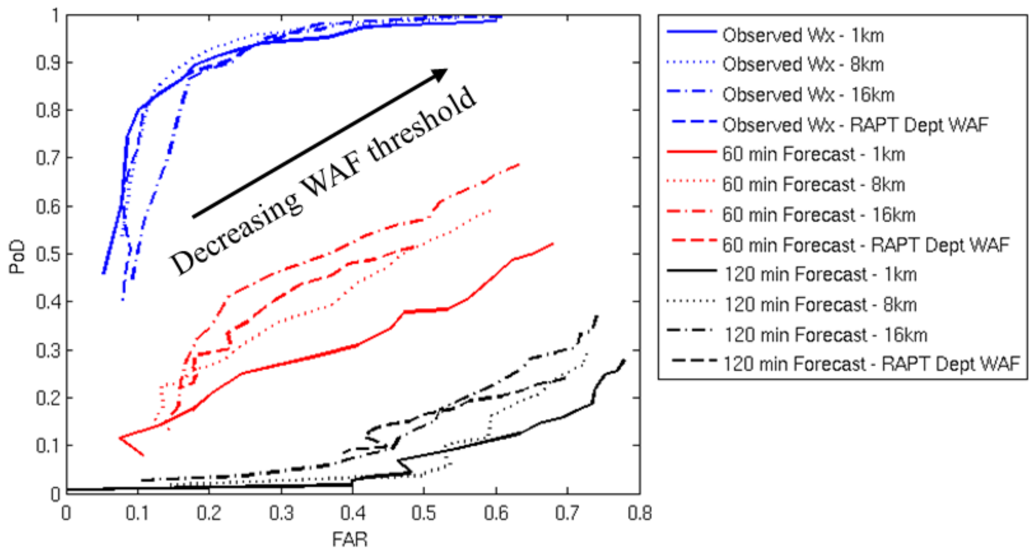


Figure 8. Comparison of classifier performance for predictors of varying spatial filter size. The classifiers are evaluated on the testing database for different forecast periods.

The VIL1, VIL8, and RAPT departure classifiers perform similarly on the observed weather data. The VIL16 classifier exhibits a higher FAR for a given PoD, which is likely a result of the predictor over-filtering (spatially expanding) the weather data. The classifier performance decreases significantly with increasing forecast horizon. In both the 60 and 120 minute forecasts, the VIL16 classifier shows the best performance, which is because the larger spatial filter is better able to capture trends in uncertain weather forecast data. Figure 9 shows the variation of CSI with deviation threshold for the predictors shown in Fig. 8.

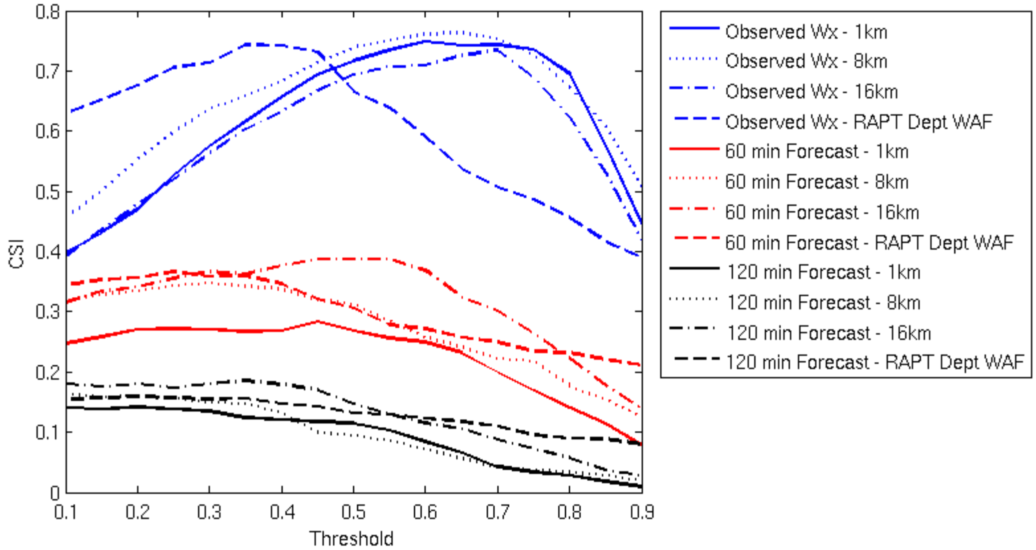


Figure 9. Comparison of Critical Success Index for predictors of varying spatial filter size applied to the testing database.

Figure 10 presents histograms of missed predictions and false alarms for the 1 km and 16 km predictors applied to different forecast horizons. The 1 km and 16 km filters are chosen to illustrate the sensitivity of missed predictions and false alarms to spatial filter size. The histograms are generated assuming a WAF deviation threshold of 70%, and the bins are labeled to show the maximum WAF value for that bin. For example, the label of 40 WAF implies $\{30 \leq WAF_{max} < 40\}$. When the model is applied to observed weather, missed deviations predominately occur with maximum WAF values between 30% and 60%, which is expected because this is where the model is most indecisive.

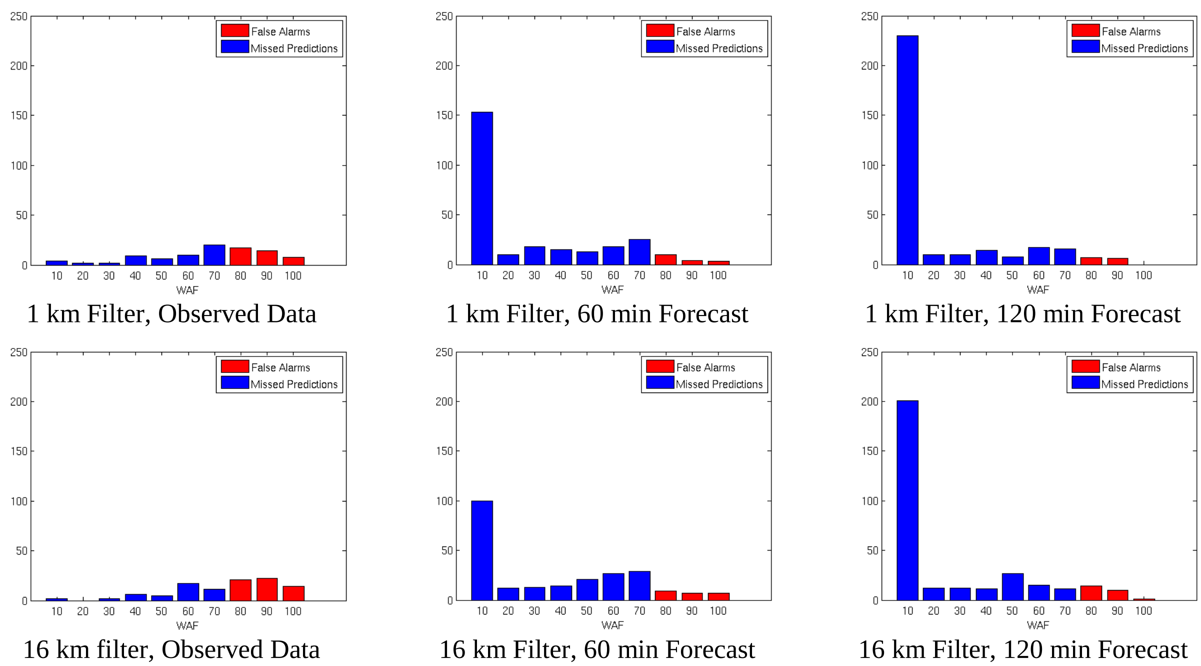
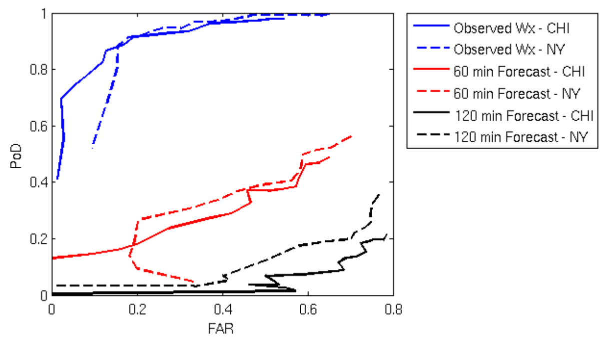


Figure 10. Histograms of missed predictions and false alarms for predictors with 1 km and 16 km spatial resolutions, assuming a WAF deviation threshold of 70%.

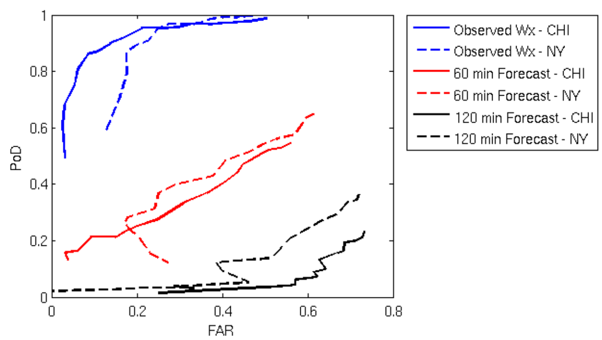
The histograms for 60 and 120 minute forecast horizons show a disproportionate amount of missed predictions in which the maximum WAF is between 0 and 10. This implies that the forecasts did not predict weather that was observed.

It is apparent that the weather uncertainty dominates the pilot decision uncertainty when the classifier is applied to forecasted weather data. It is also interesting that the 16 km filter performs better on the forecasted data relative to the other spatial filters. An explanation for this is that the larger filter is more able to capture the spatial uncertainty in the weather forecast than the smaller filters. It is also an indication that a pixel-based forecast is not an ideal forecast to implement in a pilot decision model. A more robust technique may be to use a route-based forecast, as is explained in some newly published work [10].

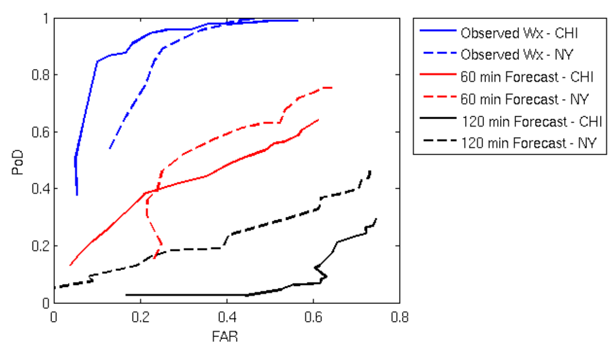
The sensitivity to geographical region is investigated by dividing the testing database into two regions (New York and Chicago), and applying the model to each region. Figure 11 shows the performance of the classifiers in New York and Chicago for varying spatial filter size and forecast period.



a) 1 km Filter



b) 8 km Filter



c) 16 km Filter

Figure 11. Comparison of predictor performance based on geographic area

The model consistently performs better on the Chicago dataset when applied to observed data and better on New York when applied to forecasted data. This implies that the model more easily predicts pilot decision making in the Chicago region, but the weather around Chicago is more difficult to forecast. This observation may be a result of limited data in each region, therefore the datasets should be expanded to get more reliable insight into the differences between the regions.

IV. Conclusions

This study introduces a low altitude version of CWAM, which is an extension of the existing CWAM developed for the en route flight regime. The paper presents a probabilistic model of pilot deviation decision making for flights between FL100 and FL240 in the presence of convective weather. The model is trained on a database of nearly 1000 encounters with convective weather, of which 309 resulted in deviations. The dominating predictor of deviation in the low-altitude flight regime is precipitation intensity, specifically the VIL level. Moreover, there is little added benefit to including additional predictors to the VIL-based classifier model.

The model is tested on an independent database of low-altitude flights. The testing confirms the observations formed during the classifier training, where precipitation intensity is the dominant predictor of deviation in low-altitude airspace. The effect of weather forecast uncertainty is inferred from the difference in classifier performance on observed and forecasted data. As expected, the classifier does not perform as well on the forecasted data, which is a result of the spatial uncertainty in the weather forecasts. Weather forecast uncertainty dominates pilot decision uncertainty when the model is applied to forecast horizons of 60 and 120 minutes. Future research should explore more robust weather forecasting techniques, as well as expand the database and explore a wider range of geographical regions.

References

[1] Clifford, S.F., et al., "*Weather Forecasting Accuracy for FAA Traffic Flow Management*", The National Academies Press, Washington DC, 2003, pp. 2.

[2] Klingle-Wilson, D., Evans, J., "Description of the Corridor Integrated Weather System (CIWS) Weather Products", MIT Lincoln Laboratory Project Report ATC-317, August, 2005.

[3] Mueller, C.K., et al., "National Convective Weather Forecast Product", 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX, January, 1999.

[4] DeLaura, R.A., Robinson, M, Todd, R.F., MacKenzie, K., "Evaluation of Weather Impact Models in Departure Management Decision Support: Operational Performance of the Route Availability Planning Tool (RAPT) Prototype", 13th Conference on Aviation, Range, and Aerospace Meteorology, New Orleans, LA, 2008.

[5] DeLaura, R.A., Robinson, M, Pawlak, M.L., Evans, J.E., "Modeling Convective Weather Avoidance in En Route Airspace", 13th Conference on Aviation, Range, and Aerospace Meteorology, New Orleans, LA, 2008.

[6] Rhoda, D.A., Pawlak, M.L., "The Thunderstorm Penetration/Deviation Decision in the Terminal Area", 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX, 1999.

[7] Rhoda, D.A., et al., "Commercial Aircraft Encounters with Thunderstorms in the Memphis Terminal Airspace", 9th Conference on Aviation, Range, and Aerospace Meteorology, Orlando, FL, 2000.

[8] Matthews, M.P., DeLaura, R.A., "Assessment and Interpretation of En Route Weather Avoidance Fields from the Convective Weather Avoidance Model", 10th AIAA Aviation Technology, Integration, and Operations Conference, Fort Worth, TX, 2010.

[9] Garcia, D., "Smoothing of Gridded Data in One and Higher Dimensions with Missing Values", Computational Statistics and Data Analysis, vol. 54, 2010, pp. 1167-1178.

[10] Pheil, D. M., "Optimization of Airport Terminal Area Air-Traffic Operations under Uncertain Weather Conditions," Ph.D. Thesis, MIT, 2011.