# Evaluation of the Convective Weather Avoidance Model for Arrival Traffic

Scot E. Campbell<sup>1</sup>, Michael P. Matthews<sup>2</sup>, and Rich DeLaura<sup>1</sup> MIT Lincoln Laboratory, Lexington, MA, 02420

The effective management of traffic flows during convective weather events in congested air space requires decision support tools that can translate weather information into anticipated air traffic operational impact. In recent years, MIT Lincoln Laboratory has been maturing the Convective Weather Avoidance Model (CWAM) to correlate pilot behavior in the enroute airspace with observable weather parameters from convective weather forecast systems. This paper evaluates the adaptation of the CWAM to terminal airspace with a focus on arrival decision making. The model is trained on data from five days of terminal convective weather impacts. The performance of the model is evaluated on an independent dataset consisting of six days of convective weather over a variety of terminal areas. Model performance in different terminal areas is discussed and the sensitivity of prediction accuracy to weather forecast horizon is presented.

### I. Introduction

A future air traffic system capable of predicting convective weather impacts and proactively issuing TMIs will more effectively use the available airspace, and in turn mitigate the effect of convective weather on the system. The Convective Weather Avoidance Model (CWAM) is a probabilistic model of pilot decision making in the presence of convective weather. CWAM is based on the correlation of spatially filtered weather observations with trajectories of aircraft that penetrated or avoided areas of convective weather in the en route flight regime [1]. The output of the en route CWAM is a three-dimensional {cloud tops, flight altitude, precipitation intensity} Weather Avoidance Field (WAF) that provides the likelihood that a pilot will deviate at a specific position and time given the current and forecasted weather. Outside of the en route phase (e.g. during departure and arrival), aircraft are commonly below the tops of most convection and are subject to different decision mechanisms, both of which are not modeled in the original CWAM. Therefore, in order to model impacts over an entire flight trajectory, CWAM should be adapted to include low-altitude flight phases such as arrival and departure [2].

This paper presents an evaluation of the adaptation of CWAM for arrival operations. Arrival CWAM is trained on approximately 11,000 flights and 1,900 terminal weather encounters over five convective weather days [3]. The training database includes multiple types of weather avoidance decisions that occur during arrival operations to four major metroplex areas (ORD, DFW, CLT, DEN). The decisions types distinguish between strategic and tactical time horizons and encompass both pilot and air traffic management decisions. Additionally, unlike pilots in en route airspace who may have an option to fly at higher altitudes over storms, pilots in arrival airspace are constrained to follow descending trajectories that are typically below the cloud tops. For this reason, the output of the arrival CWAM is a two-dimensional WAF {precipitation intensity, cloud tops}.

The performance of arrival CWAM is evaluated by an independent dataset, where the sensitivity of the model to terminal airspace structure and weather forecast horizon are investigated. The independent dataset contains weather decisions from six convective weather days in a variety of terminal areas (ORD, DFW, DEN, CLT, BOS, JFK/LGA/EWR, DCA/IAD). The most descriptive features of pilot avoidance of convective weather are precipitation intensity and storm height, where a 4 km spatial filter on the 90<sup>th</sup> percentile value of each feature corresponds to the best tradeoff between probability of detection and false alarm rate. The performance of the model

<sup>\*</sup>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.

<sup>&</sup>lt;sup>1</sup> Technical Staff, Weather Sensing, 244 Wood Street, Lexington, MA 02420, AIAA Member.

<sup>&</sup>lt;sup>2</sup> Associate Staff, Weather Sensing, 244 Wood Street, Lexington, MA 02420, AIAA Member.

on the independent dataset matches the training performance well, and is comparable to both en route and departure CWAM, especially when terminal area differences are isolated [4].

## **II.** Arrival CWAM Training

The arrival Convective Weather Avoidance Model (CWAM) is trained on five days of convective weather impacts. Arrival CWAM is based on similar methodology to en route CWAM, which is modeled on weather avoidance decisions that correlate to weather features extracted along flight plans using the CIWS suite of weather products [1]. Arrival CWAM extends the methodology of en route CWAM to include Air Traffic Management (ATM) weather avoidance decisions in addition to pilot decisions.

For this study, the training database includes weather avoidance decisions from five convective weather days from four major airports across the country. The weather avoidance decisions are limited to *terminal* impacts, where the terminal area is defined as the area within 150 kilometers of the arrival airport. Additionally, planning decisions (i.e. pre-departure decisions) are excluded from the database. Table 1 lists the trajectory counts for each arrival airport, and gives the number of weather encounters and avoidance decisions. A weather encounter is defined to occur when a flight either makes a weather avoidance decision, or penetrates weather greater than VIP level 1 without a weather avoidance decision.

| Trajectories | Weather Encounters                            | Avoidance Decisions  |
|--------------|---|--|
| 5602         | 1722  | 326  |
| 2358         | 457   | 152  |
| 679          | 161   | 107  |
| 811          | 326   | 269  |
| 9450         | 2666  | 854  |
|              | Trajectories   5602   2358   679   811   9450 | Trajectories Weather Encounters   5602 1722   2358 457   679 161   811 326   9450 2666 |

| Table 1. | Traiect | torv coun   | t for eac | h arrival | airport in | the arrival | CWAM tr | aining database |
|----------|---------|-------------|-----------|-----------|------------|-------------|---------|-----------------|
|          |         | • • • • • • |           |           |            |             |         |                 |

Table 2 presents the frequency of weather decision types in the training database. For a detailed description of the decision types, see Ref. 3. The most common decisions are reroutes and deviations, and there are slightly more ATM decisions than pilot decisions.

|           | <b>TT</b> 7 41 |            | •                                    |          | • • •  | • •     | ATT A | <b>N</b> <i>T</i> <b>( )</b> | • • •       |       |
|-----------|----------------|------------|--------------------------------------|----------|--------|---------|-------|------------------------------|-------------|-------|
| Toble 7   | M/Anthor       | doorgion f | $\mathbf{v}$ <b>n</b> o $\mathbf{t}$ | POOLONON | in tha | OPPHIC  |       | V/I troin                    | ana dotobo  | CO    |
| I ADIE 4. | vvealuer       | UCUSION L  | vije i                               | requency | ни пле | arrivar | A     |                              | ing uatabas | SE.   |
|           |                |            | .,                                   |          |        |         |       |                              |             | ~ • • |

| A       | ATM Decisions      | Pilot Decisions         |                      |  |  |
|---------|--------------------|-------------------------|----------------------|--|--|
| Reroute | Holding & Slowdown | Deviation & Pathfinding | Diversion and Missed |  |  |
| 403     | 138                | 285                     | 28                   |  |  |

The output of arrival CWAM is a Weather Avoidance Field (WAF) that provides the probability that a flight will incur a convective weather-related impact between the top-of-descent and the runway. The WAF is based on the observed probability of convective weather impact partitioned into two-dimensional bins of Vertically Integrated Liquid (VIL) and Echo Tops (i.e. precipitation intensity and cloud tops, respectively). The WAF look-up table is heuristically smoothed and extrapolated to characterize intuitive decision boundary contours. Figure 1 presents the WAF look-up tables for arrival CWAM, where (a) contains a coarse distribution of VIL and Echo Tops, and (b) is based on the resolution of the current departure WAF in the Route Availability Planning Tool (RAPT) [5].



Figure 1. Arrival CWAM weather avoidance field look-up tables.

The performance of the model is evaluated by converting the WAF into a binary decision classification using a WAF decision threshold ( $WAF_{threshold}$ ). First, the trajectories in the database are assigned WAF values for each time step k in the observed trajectory ( $WAF_{actual}[k]$ ) as well as the corresponding maximum WAF value in the flight plan ( $WAF_{plan}[k]$ ). Each avoidance decision at time step  $k_D$  is scored based on the difference between  $WAF_{threshold}$  and  $WAF_{plan}[k_D]$ . If  $WAF_{threshold} \leq WAF_{plan}[k_D]$  the decision is labeled a correct avoidance prediction (CAP) and if  $WAF_{threshold} > WAF_{plan}[k_D]$  the decision is labeled a false penetration prediction (FPP). Trajectories that do not make an avoidance decision and are characterized by  $WAF_{threshold} \leq WAF_{actual}[k]$  for any k, are labeled a false avoidance prediction (FAP). Likewise, if  $WAF_{threshold} > WAF_{actual}[k]$  for all k the trajectory is labeled a correct penetration prediction (CPP). Figure 2 presents a notional example of the scoring metrics with  $WAF_{threshold} = 70$ .



Figure 2. Notional example of trajectory scoring metrics. The WAF threshold is 70, the flight plan is the magenta line, and the flown trajectory is the blue dashed line.

Equations 1 and 2 form the metrics that are used to evaluate the effectiveness of the model.

$$Probability of Detection = \frac{CAP}{CAP + FPP}$$
(1)

$$False A larm Rate = \frac{FAP}{CAP + FAP}$$
(2)

Figure 3 shows a receiver operating characteristics (ROC) curve for models with different spatial filter sizes and look-up tables. For this paper, a spatial filter refers to the 90<sup>th</sup> percentile value in a kernel with length of x km. For example, a 4 km spatial filter generates the 90<sup>th</sup> percentile VIL and Echo Tops within a 4 km by 4 km kernel. In Fig. 3, the ideal tradeoff between the probability of detection and false alarm rate occurs at the minimum distance from each curve to the top-left corner of the plot. The dots on the curves correspond to a WAF threshold of 70% probability of deviation. The blue and red curves show the performance of the look-up table of Fig. 1a (coarse resolution) with spatial filters of 1 km (native resolution) and 16 km, respectively. The black and green curves show the performance of the look-up table of Fig. 1b (fine resolution) with spatial filters of 4 km and 16 km, respectively. The magenta curve shows the performance of the departure WAF look-up table currently utilized in RAPT.



Figure 3. Receiver operating characteristics curve for difference spatial filters and look-up tables. The dots correspond to a WAF threshold of 70% probability of deviation. The "Low-Res WAF" corresponds to the look-up table of Fig. 1a, and the "High-Res WAF" corresponds to the look-up table of Fig. 1b.

Figure 4 shows the sensitivity of the ROC curves to spatial filter size. The look-up table for all three filters is the high-resolution table shown in Fig. 1b. As expected, increasing the spatial filter size raises the probability of detection at low WAF thresholds. However, at high WAF thresholds, the 16 km spatial filter results in the highest false alarm rate. This is an artifact of large filters that tend to artificially grow the cores of small storms such that they intersect with the paths of the flights in the database. The 4 km filter results in the best tradeoff between false alarm rate and probability of detection (minimum distance to the top-left corner) for a WAF threshold of 70%.



Figure 4. Receiver operating characteristics curve showing the sensitivity of performance to spatial filter size for the "High-Res" look-up table in Fig. 1b. The dots correspond to a WAF threshold of 70% probability of deviation.

Figure 5 shows the sensitivity of the ROC curves to different WAF look-up tables with a fixed 4 km spatial filter. The RAPT departure look-up table generates a slightly lower probability of detection for a given false alarm rate, however the overall effect of look-up table is largely negligible, especially with a WAF threshold of 70%.



Figure 5. Receiver operating characteristics curve showing the sensitivity of performance to look-up table for a fixed 4 km spatial filter. The dots correspond to a WAF threshold of 70% probability of deviation. The "Arrival WAF" corresponds to the look-up table of Fig. 1a, and the "Arrival HD WAF" corresponds to the look-up table of Fig. 1b.

With these sensitivities in mind, the high-resolution arrival CWAM look-up table with the 4 km filter is chosen as the "best" WAF for the arrival model. The 4 km spatial filter provides the best tradeoff between probability of detection and false alarm rate, and the high-resolution look-up table is an ideal platform for further analysis. This is the WAF that is evaluated in the next section.

## **III. Arrival CWAM Evaluation**

The performance of the arrival adaptation of CWAM is evaluated on an independent dataset of 12,064 trajectories in which 5,097 weather decisions are made. Like the training dataset, the testing dataset is structured to include multiple decision classes that are initiated by both pilots and air traffic management. Weather avoidance decisions are determined in the same fashion as in the training dataset, where decision classification is performed manually for each trajectory. Table 3 lists the trajectory counts for each arrival airport, the number of weather encounters, and the number of weather avoidance decisions.

| Table 5. Trajectory count for each arrival an port in the arrival C wAw testing database. |              |                    |                     |  |  |  |  |  |  |
|---|--------------|--------------------|---------------------|--|--|--|--|--|--|
| Airport   | Trajectories | Weather Encounters | Avoidance Decisions |  |  |  |  |  |  |
| ORD   | 3218         | 709                | 285                 |  |  |  |  |  |  |
| DFW   | 942          | 64                 | 7                   |  |  |  |  |  |  |
| CLT   | 778          | 279                | 174                 |  |  |  |  |  |  |
| DEN   | 1052         | 311                | 152                 |  |  |  |  |  |  |
| BOS   | 522          | 185                | 97                  |  |  |  |  |  |  |
| DCA/IAD   | 1722         | 317                | 141                 |  |  |  |  |  |  |
| JFK/LGA/EWR   | 3830         | 476                | 212                 |  |  |  |  |  |  |
| All   | 12064        | 2341               | 1068                |  |  |  |  |  |  |

Table 3. Trajectory count for each arrival airport in the arrival CWAM testing database

Table 4 presents the frequency of weather avoidance decision types in the testing database. The most common decisions are reroutes and deviations, and there are slightly more ATC decisions compared to pilot decisions.

| Table 4. | Weather | avoidance o | lecision t | ype fi | requenc | y in t | he arrival | CWAM | l training | g datab | ase. |
|----------|---------|-------------|------------|--------|---------|--------|------------|------|------------|---------|------|
|----------|---------|-------------|------------|--------|---------|--------|------------|------|------------|---------|------|

| A       | TC Decisions       | Pilot                   | Decisions          |
|---------|--------------------|-------------------------|--------------------|
| Reroute | Holding & Slowdown | Deviation & Pathfinding | Diversion & Missed |
| 534     | 122                | 352                     | 60                 |

The objective of this evaluation is to verify the correctness of the model such that it can be utilized in the prediction of weather impacts in the terminal area. However, there are limitations in scoring the performance of air traffic management decisions because the trajectory data used to validate the model is insufficient to identify false avoidance predictions. For example, consider the case in which weather impacts the terminal area, with WAF<sub>plan</sub>>WAF<sub>threshold</sub>, but no decision is made by ATM to close the route and the flights filed on the route are not close enough to the weather to deviate. In this instance, the ATM decision to keep the route open cannot be judged because there is no traffic data (e.g., pilots refusing to use the airspace) to suggest that the airspace is unacceptable to the pilot. In other words, pilot decisions are made on a very tactical timescale, where the decisions can be easily correlated with the current weather. ATM decision point cannot be clearly identified). For this reason the focus of the performance evaluation will be on pilot decision modeling

The first step in the evaluation is to assess the statistical goodness of the model. In the context of this evaluation, statistical goodness is interpreted by observing the frequency of weather decisions in a specified range of weather avoidance field values. Figure 6 presents a histogram of the testing dataset, where the bars represent the frequency of pilot deviation decisions and pilot penetration decisions for bins of predicted weather avoidance probability. As expected, the vast majority of penetrations occur at a low maximum probability of avoidance, whereas deviation decisions dominate high values of maximum probability of avoidance. The frequency of penetrations and deviations is approximately equal when the maximum probability of avoidance is 0.7 in both the training and testing datasets. An additional characteristic of the model is that the penetrations and deviations predominately occur at either high or low values of predicted probability of avoidance. In other words, the model is able to *decisively* predict deviations

and penetrations. This is a very beneficial characteristic in decision support, where in many cases the usability of the tool is directly related to its decisiveness.



Figure 6. Frequency of deviations and penetrations in the testing dataset.

Figure 7 compares the ROC performance of the model using the testing and training datasets, where a measure of ideal performance is the minimum distance between the curve and the top-left corner of the plot. The overall shape of the curves is similar, however the testing curve is rotated slightly counter-clockwise. The difference is most likely explained by variation in weather type and/or terminal area between the datasets.



Figure 7. A comparison of the receiver operating characteristics curves for the training and testing datasets.

The training dataset is composed of weather encounters in terminal areas that are structured in a "corner post" fashion. The testing dataset includes both "corner post" and highly dense "complex" terminal areas (DCA/IAD,





Figure 8. A comparison of the receiver operating characteristics curves for the training and testing datasets, where the testing dataset is partitioned by terminal area type. The dotted line represents model performance in "corner post" terminal areas and the dash-dot line represents model performance in "complex" terminal areas.

As expected, when the testing dataset is restricted to weather encounters in corner post terminal areas, the ROC curve more closely matches the training dataset. This implies that the performance of arrival CWAM is sensitive to the geographical area in which it is applied. At high probabilities of detection (greater than 0.6), the model performs better on the testing dataset than the training dataset, which is most apparent when the model is tested on data from the more dense terminal areas (DCA, IAD, JFK, LGA, EWR). In other words, the model evaluated on dense traffic data is less likely to have more false weather penetration predictions than the model evaluated on corner post traffic data.

Often it is useful to have a single metric to describe the performance of a predictive model. One such metric is the critical success index (CSI), which is a measure of correct avoidance prediction accuracy discounted by all model errors. The CSI is given by Eq. (3).

Critical Success Index = 
$$\frac{CAP}{CAP + FPP + FAP}$$
 (3)

Figure 9 shows the variation of CSI with WAF threshold for the corner post and complex partitions of the testing dataset. The blue line shows the CSI of the model using the corner post partition and the red line uses the complex partition. The arrows indicate the WAF threshold corresponding to the maximum CSI. The model performs slightly better in the complex TRACON compared to the corner post TRACON based on the maximum CSI value. Additionally, the WAF threshold corresponding to the maximum CSI is higher for operations in a complex TRACON.



Figure 9. Variation of critical success index with WAF threshold for the corner post and complex partitions of the testing dataset. The arrows indicate the WAF threshold for maximum CSI.

In a sense, the maximum CSI value corresponds to the best tradeoff between the ability to predict avoidance and non-avoidance decisions. The fact that the model generates a higher CSI in a complex TRACON indicates that, on average, the model has more skill in predicting pilot behavior in complex terminal areas. However, differences in model performance between disparate TRACON types does not necessarily mean that distinct CWAM models should be developed for each specific area. One potential solution is to utilize an adjustable WAF threshold to maximize the model performance given the terminal area. In essence, an adjustable WAF threshold enables the model to operate at the ideal tradeoff over a range of terminal areas. The approach of adjusting WAF threshold values for different airspace characteristics has been used successfully in the operational RAPT prototypes in New York and Chicago.

For the remainder of this section the performance of the model is described from the perspective of the decision maker, where Eqs. (4-5) calculate the metrics used to evaluate the model.

Probability of Correct Avoidance Prediction = 
$$\frac{CAP}{CAP + FAP}$$
 (4)

Probability of Correct Penetration Prediction 
$$= \frac{CPP}{CPP + FPP}$$
 (5)

The variable CAP represents correct avoidance prediction, FAP is a false avoidance prediction, CPP is a correct penetration prediction, and FPP is a false penetration prediction. In practical terms, the probability of correct avoidance prediction (PCAP) is the probability that a deviation prediction is true, given that a deviation is predicted. Essentially, it measures the accuracy of a deviation prediction. The probability of correct penetration prediction (PCPP) is the probability that a penetration prediction. The probability of correct penetration prediction (PCPP) is the probability that a penetration prediction is true, given that a penetration prediction (PCPP) is the probability that a penetration prediction is true, given that a penetration is predicted. The ideal predictor corresponds to a curve that has a point on the top-right corner, where the PCAP = PCPP = 1.

The effect of weather forecast horizon on model accuracy is especially important for application to ATM decision support tools. Figure 10 shows the tradeoff between the PCAP and PCPP of the model on observed weather, the 60 minute CIWS forecast, and the 120 minute CIWS forecast. It is obvious that forecast uncertainty significantly affects the overall performance of the model. The probability of correct penetration prediction remains above 0.8 for both forecast horizons, but the probability of correct avoidance prediction drops from approximately 0.7 to 0.38 in the 60 minute forecast to 0.25 in the 120 minute forecast. The steep decrease in performance is, in part, a result of small scale forecast errors that affect the performance of the model but are not necessarily operationally significant. Additional spatial filtering such as the route blockage algorithm in RAPT would increase the operational performance of the model.

American Institute of Aeronautics and Astronautics



forecast (black).

#### **IV.** Conclusions

This report presents an evaluation of the arrival convective weather avoidance model (CWAM). Arrival CWAM is based on approximately 11,000 flights and 1,900 terminal weather encounters over five convective weather days. The training database includes weather avoidance decisions of multiple types that occurred during arrival to four major metroplex areas (ORD, DFW, CLT, DEN). The most descriptive features of the model are VIL (precipitation intensity) and Echo Tops (cloud height), and a 4 km spatial filter on the 90<sup>th</sup> percentile value of each feature provides the model the best tradeoff between probability of detection and false alarm rate.

The performance of arrival CWAM is evaluated by an independent dataset of 12,064 trajectories and 5,097 weather decisions, where the sensitivity of the model to both terminal airspace structure and forecasted weather are investigated. The independent dataset contains weather decisions from six convective weather days in a variety of terminal areas (ORD, DFW, DEN, CLT, BOS, JFK/LGA/EWR, DCA/IAD). The performance of the model on the testing dataset is similar to the training performance, especially when terminal area differences are isolated. The model does not perform as well on forecasted weather data, which is primarly a result of small scale forecsting errors in the 60 and 120 minute CIWS forecasts. However, many of these errors are not operationally meaningful and can be removed by post-processing the WAF such as done in RAPT. Further work is needed to investigate the true impact of weather forcast horizon on arrival CWAM performance. This includes a study on the effect of spatial filtering on the performance of the model with forecasted weather data as input. For example, it is possible that a larger spatial filter in the model will decrease the sensitivity to weather forecast uncertainty and result in better performance.

#### References

<sup>1</sup>DeLaura, R., Robinson, M., Pawlak, M. L., Evans, J. E., "Modeling Convective Weather Avoidance in Enroute Airspace," 13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), New Orleans, LA, Amer. Meteor. Soc., 2008. <sup>2</sup>Rhoda, D. and Pawlak, M., "An Assessment of Thunderstorm Penetrations and Deviation by Commercial Aircraft in the Terminal Area," Project Report NASA/A-2, June 1999, MIT Lincoln Laboratory. <sup>3</sup>Matthews, M. and DeLaura, R., "Modeling Convective Weather Avoidance of Arrivals in the Terminal Airspace," Second Aviation, Range and Aerospace Meteorology Special Symposium on Weather-Air Traffic Management Integration, Seattle, WA, Amer. Meteor. Soc., 2011.

<sup>4</sup>Rubnich, M., and DeLaura, R., "Initial Validation of a Convective Weather Avoidance Model (CWAM) in Departure Airspace," 30<sup>th</sup> Digital Avionics Systems Conference (DASC), Seattle, WA, Oct. 16-20, 2011.

<sup>5</sup>DeLaura, R., 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," AMS 13th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), New Orleans, LA, 2008.