MEASURING THE UNCERTAINTY OF WEATHER FORECASTS SPECIFIC TO AIR TRAFFIC MANAGEMENT OPERATIONS*

Michael P Matthews†
Marilyn M. Wolfson
Richard A. DeLaura
James E. Evans
Colleen K. Reiche
Massachusetts Institute of Technology, Lincoln Laboratory
244 Wood Street, Lexington, MA 02420

Hamsa Balakrishnan and Diana Michalek
Massachusetts Institute of Technology
Cambridge, MA 02139

1. INTRODUCTION

The efficient management of air traffic during convective weather in regions of congested air space is a very difficult task. Significant efforts have been underway for many years to provide improved forecasts of convective weather to traffic flow managers to help them increase air space usage efficiency during these times. However, due to the increased workload created by convective weather impacting the air traffic routes and other factors, increased forecast performance may not translate directly into more efficient operations. In recent years, automated decision support tools have been under development to assist traffic managers in more effectively utilizing the improved forecast products. These tools are designed to integrate deterministic or probabilistic forecasts of convective weather into impacts on predetermined air traffic routes in the national air space (NAS).

The usefulness of the automated guidance is dependent upon both the accuracy of models to interpret the air traffic control impact of the weather and the accuracy of the weather forecasts themselves. MIT Lincoln Laboratory, in collaboration with NASA Ames, developed a quantitative Convective Weather Avoidance Model (CWAM) to translate convective weather into impacts on air traffic operations (DeLaura and Evans, 2006 and DeLaura et al., 2008). The results of the CWAM allowed the creation of a Weather Avoidance Field (WAF) which provides an estimate of the probability of pilot deviation around convective weather at each altitude in en route airspace. Significant efforts are underway to understand and validate the CWAM model by analyzing the pilot's behavior as it relates to various measures of the weather. However, to date, limited effort has been invested in providing quantitative, reliable estimates of forecast error that can be readily translated into impact uncertainty. The best modeling of pilot behavior in the presence of convective weather is of limited use for forecasting future route blockage or sector capacity when errors associated with the forecasted weather are unknown.

In this paper, we develop a novel way to measure the accuracy of weather forecasts based upon the impact on air traffic flows. This method uses new techniques developed as part of the CWAM that consider the complicated interaction between pilots, air traffic controllers and weather. This technique, known as the blockage model (Martin et al., 2006), differentiates between minor deviations performed by pilots around convective weather and their larger deviations due to fully blocked air routes that require air traffic control interaction. This blockage model is being used by the automated Route Availability Planning Tool (RAPT) to predict route blockage for NYC departures. RAPT integrates the Corridor Integrated Weather Systems (CIWS) deterministic 0-2 hour forecasts of precipitation and echo tops into route specific forecasts of impact on air traffic in the congested east coast corridor. Applying the blockage model to the entire CIWS weather domain as a metric for scoring the performance of the forecast algorithms is shown to be an excellent approach for measuring the adequacy of the forecast in predicting the impact of the convective weather on air traffic operations.

*This work was sponsored by the Federal Aviation Administration (FAA) under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.
†*Corresponding author address*: Michael P Matthews, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: mpm@ll.mit.edu

2. BACKGROUND

The Corridor Integrated Weather System is a prototype aviation weather system developed to aid air traffic controllers in the situational awareness of weather impacting the U.S. National Airspace System (NAS; Evans and Ducot 2006). The CIWS provides a collection of weather products for display that can be easily interpreted by air traffic control managers. Of significant use to managers are the forecast products that depict a deterministic picture of the location of precipitation and echo tops out to 120 minutes in the future (Wolfson et al., 2004 and Dupree et al., 2005). These products can be used to plan route and sectors impacts and proactively adjust traffic flows through reroutes or closures. The CIWS coverage was increased to cover the CONUS in June, 2008 and the system has now been wholeheartedly accepted by the operational air traffic control user community.

The Route Availability Planning Tool is an automated air traffic management decision support tool that was developed for air traffic control users to assist in determining the impact of weather on departing flights from the major NYC airports (DeLaura et al., 2008). The RAPT uses the 0-120 minute forecasts of precipitation and echo tops produced by the CIWS and a model of pilot behavior in the presence of convective weather developed for RAPT. The Convective Weather Avoidance Model developed in conjunction with NASA Ames, estimates the probability of aircraft deviation around convective weather in the air space (DeLaura et al., 2008).

The CWAM model was developed by analyzing over 2100 flight trajectories through convective weather. Classifying each flight trajectory as a deviation or a non-deviation allowed the creation of a database to predict the probability of deviation based upon the observed weather parameters. Within CWAM the weather data is processed with many different spatial filters to try and simulate the pilot's decision making based upon the severity and coverage of convective weather along the flight path. Statistical analysis was used to select the best predictors of pilot deviation. These variables were combined to create the Weather Avoidance Field (WAF) to provide an estimate of the probability of pilot deviation around convective weather in en route airspace. High WAF probabilities indicate areas where pilots are more likely to deviate. Figure 1 shows the methodology for generation of the WAFs using the CWAM model algorithm.

Two different data sets were used as part of the CWAM, which resulted in two different WAF models, known as CWAM1 and CWAM2. The first CWAM used trajectories over five days in the summer of 2005 from the Indianapolis Enroute Center's airspace (ZID). The second CWAM used trajectories over six days in the summer of 2006 from two additional enroute centers; Washington (ZDC) and Cleveland (ZOB).

Both CWAM1 and CWAM2 models selected the difference between flight altitude and convective storm height as the best indicator of deviation. The convective height is defined as the 18 dBZ echo top. The storm height used at a particular pixel in the model was the 90th percentile echo top in a 16 x 16 km box centered on the pixel.

The CWAM1 and CWAM2 models each selected a different parameter as the second most important predictor. CWAM1 found that the storm intensity, as defined by the spatial coverage of the Vertically Integrated Liquid (VIL) was the second best predictor. The VIL spatial coverage was defined as the percentage of pixels >= VIL level 3 over a 60 x 60 km box centered on the pixel. On the other hand, CWAM2 found that the spatial coverage of the storm height was the second most important. The spatial coverage of storm height was defined as the echo tops >= 30 Kft over a 16 x 16 km box centered on the pixel.

In either case, the performance of the models only varied slightly between the CWAM1 and CWAM2 results. In both models, due to the use of large spatial filters, the WAFs tended to be smoother than the observed weather.

The CWAM1 and CWAM2 model development heavily relied upon statistical analysis tools and multiple spatial filters to generate a correlation between pilot behavior and convective weather. Hand analysis of the specific weather-air traffic scenarios in the CWAM 1 and 2 databases allowed for a subjective performance assessment and gave rise to valuable insight into potential avenues for model improvement. Although the CWAM effectively indicated regions that are likely to be avoided by pilots, there were many instances of deviation over-prediction.

Via our flight-by-flight analysis, we found certain weather scenarios produced high probabilities of deviation in the current CWAM, yet caused no problems for pilots. For example, the stratiform rain left behind after a convective storm can have high echo tops as well as VIL greater than level 3, without being highly turbulent. We also found many cases where aircraft penetrated high deviation probabilities associated with the edges of convective weather. In some instances, due to the large spatial filters used in CWAM, it was possible for an aircraft to penetrate a high deviation probability while not actually encountering significant convective weather. Finally, we noted considerable uncertainty with the deviation predictions from all CWAM results when the flight altitude is near the echo top height (+/- 4000 ft.).

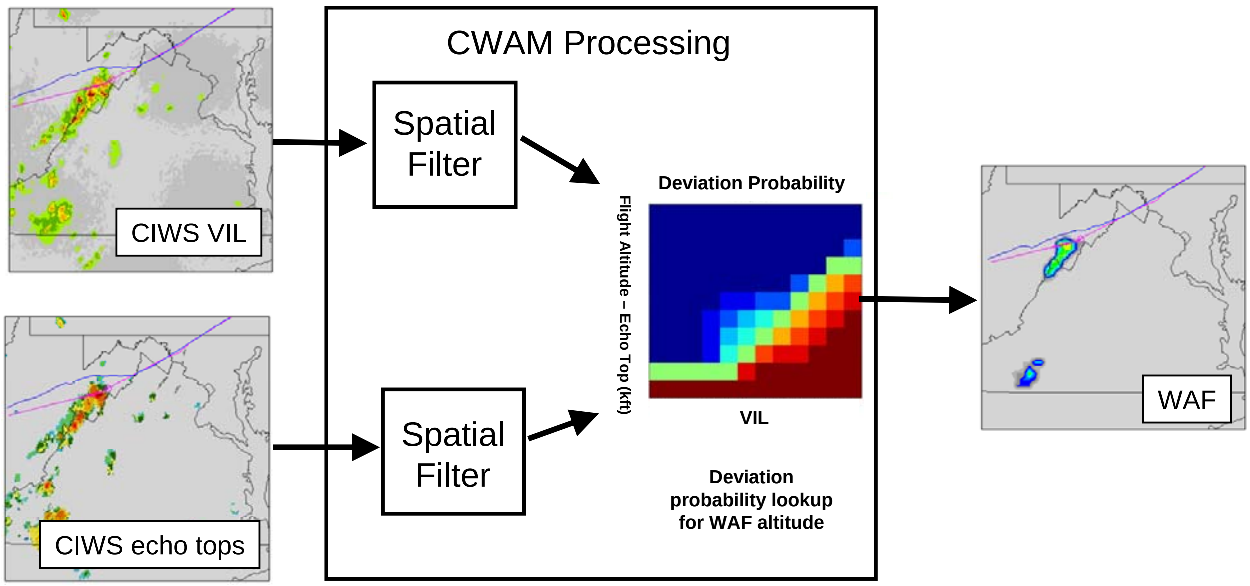

Figure 1. Generation of weather avoidance fields (WAFs) using the Convective Weather Avoidance Model (CWAM).

3. BLOCKAGE ALGORITHM

3.1 Cockpit and ATC Interaction

In trying to predict pilot behavior it is important to understand the decision making that goes on inside the cockpit. Large towering convective clouds are known by pilots to be regions of turbulence that are desirable to avoid. While en route, pilots are often flying above the cloud tops and have visual cues that factor into the decision making. At low altitudes, or at higher altitudes with high cloud tops, pilots are dependent upon the onboard radar to evaluate the storm intensity. Regions of VIP level 3 and above on the pilot's radar are avoided because the pilot believes these are indicative of turbulence.

All of the information on storm convective height and storm intensity is of little use in understanding pilot behavior without taking into consideration the constraints placed upon pilots due to air traffic control. In the NAS, aircraft are monitored by air traffic control to provide a safe flight from origination to destination. The air traffic controllers have a primary responsibility to maintain safe separation between aircraft. To do this controllers provide pilots with vectors that move aircraft along established 'highways in the sky' known as air routes. The system is designed to efficiently move flights along these routes from point to point with the appropriate spacing between planes in the absence of weather. The problems come on days with lots of convective weather. If convective weather is blocking a planned route, pilots, air traffic controllers, and dispatchers have to provide the aircraft with alternate routing, or delay the departure until the hazardous weather has cleared. Air traffic controllers can and do provide pilots with information about weather, but are not required to maintain separation between aircraft and weather.

In theory, the air traffic controller's responsibility is to move the plane from point A to point B along the air routes. However, in practice, the pilot is allowed to make minor deviations off of the centerline while traveling between the two points. The RAPT algorithm defines a route width as roughly 40 kilometers, or 20 kilometers on either side of the center line. This was estimated from discussions with air traffic controllers, pilots and clear air observations of aircraft trajectories.

Using tools to plot flight trajectories over weather, analysts have observed aircraft flying in close proximity to convective weather. It is speculated that these pilots are avoiding the heavy weather along the planned trajectory, but are staying within the acceptable bounds of deviation along a jet route. This concept is the primary building block of the blockage algorithm.

3.2 Computational Methodology

The blockage algorithm is a directional approach to spatial filtering of the weather. It begins by extracting from the weather grid a rectangle that is 55 km along the defined direction of travel and 40 km wide (Figure 2). The 55 km length is chosen to simulate the amount of time needed to coordinate a fairly large deviation around weather, including interaction with Air Traffic Control. For instance, an en route plane traveling at 650 km/hr would traverse the 55 km in approximately five minutes. This is approximately the time it would take a pilot to observe visually (or on radar) a hazard that is going to require action, convey that measure to ATC, and have ATC approve and convey the approval back to the cockpit. The 40 km width is chosen to simulate the amount of deviation typically possible without any consequence of ATC coordination. The 55 km long by 40 km wide rectangle is referred to as the route segment.

Once the route segment is extracted, the algorithm begins looking for the widest path along the 55 km length of the segment that encounters unavoidable weather. This is achieved by incrementally looping on the weather variable value until it is impossible to traverse the path without encountering the present value, then stepping back one increment. For instance, if the rectangle contains levels 1-4 precip, but it is possible to traverse the path without encountering level 3 and 4 precip, then level 2 becomes the unavoidable weather and level 3+ becomes the avoidable weather (Fig. 2). The algorithm then looks at the width of the gaps between avoidable weather regions. The gaps would be regions where the aircraft could fly between avoidable weather cells or, as in Fig. 2, between the avoidable (level 3+) cell and the edge of the 40 km wide route segment. Once all of the gaps are determined, the algorithm assigns the flight trajectory to the widest gap between the avoidable weather regions.

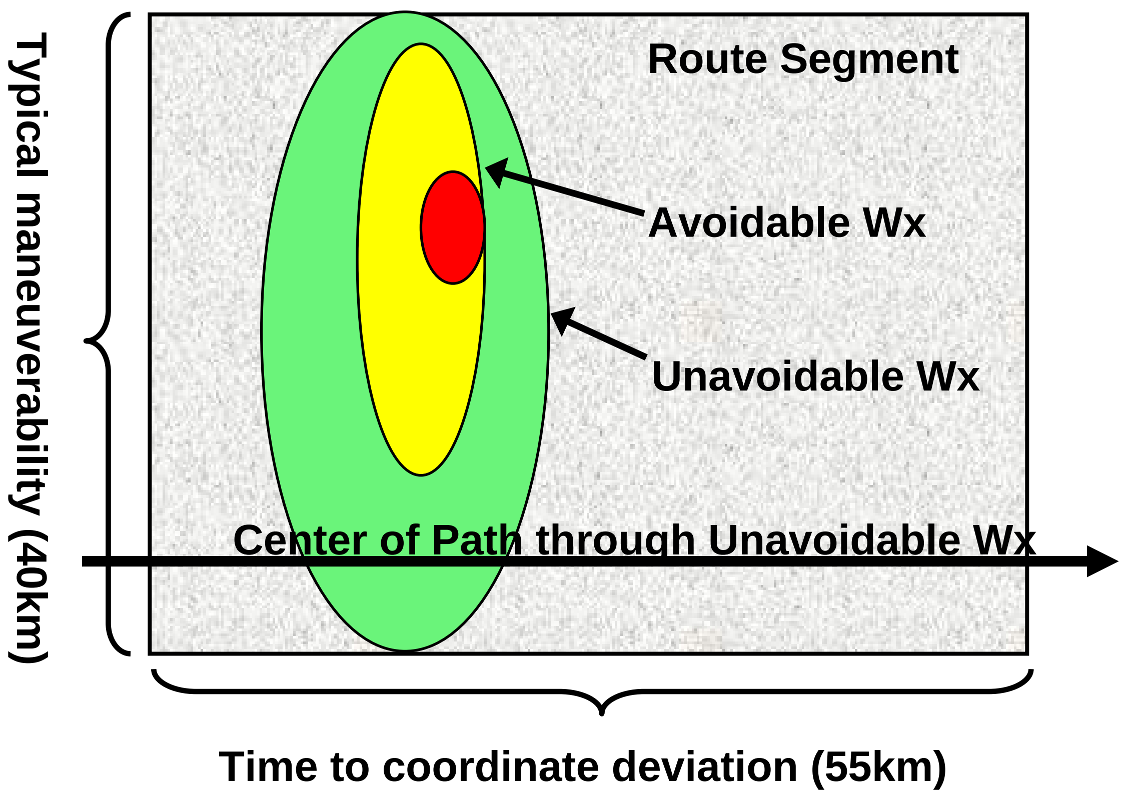

Figure 2. Finding the center of the path through the unavoidable weather in the blockage algorithm.

Once a preferred path has been determined through the route segment the algorithm does a distance weighted mean of all of the weather within the segment centered on the path. The weighting scheme used is a 1/R where R is the distance from the preferred path to the weather pixel. Only pixels greater than the unavoidable weather are used in the weighting scheme. Thus, the blockage value assigned to the segment will be at a minimum the value of the unavoidable weather. However, if there is a significant amount of weather within the segment that is greater than the unavoidable weather, the blockage value will be greater than the unavoidable weather. This weighting scheme is designed to reflect the impact of the unavoidable weather while taking into account the possibility that there is so much convective weather that a pilot may chose to avoid this segment just due to the area coverage of heavier weather.

Route blockage is also directional. The route segment around a particular pixel can be oriented in any arbitrary direction. We typically test for blockage in the E-W, N-S, NE-SW, and NW-SE directions. Depending on the structure and orientation of the weather cells, the blockage pattern in the different directions can be very different. This directional characteristic of route blockage captures what is actually observed in air traffic flow around weather.

3.3 Applying Blockage to VIL and Echo tops

The notion of route blockage provides a traffic flow orientated mapping between the weather and ultimately route capacity by using the blockage algorithm as a spatial filter prior to the CWAM statistical analysis. In this study we focused on characterizing the actual and forecasted weather, with the blockage algorithm on VIL and echo tops at multiple angles and averaged into one non-directional result at each pixel. An alternative approach for different purposes might select from the multiple blockage orientations in such a way as to match the predominant flow in today's NAS route structure.

Applying the blockage to the VIL and echo top fields essentially allows for an additional spatial filter to be incorporated into the Convective Weather Avoidance Model. Figure 3 shows the two dimensional histograms from the original CWAM1 versus the new blockage-based spatial filters, dubbed CWAM3. The left column figures show the probability of deviation, from 0 to 1. The right column figures show the total count of pilot weather encounters (deviations and non-deviations) for each bin on a scale of 0 to 180 for CWAM1 and 0 to 100 for the blockage algorithm-based CWAM3. (Notice that the vertical scale of flight altitude – echo tops is plotted from positive to negative for CWAM1 in the top row and from negative to positive for CWAM3 in the bottom row.)

The new CWAM3 method demonstrates a much stronger delineation between deviations and non-deviations, especially at the crucial range of +/-4000 feet in the difference between flight altitude and echo top height, where large uncertainty was noted. Also, this new model has moved the center of the highest concentration of encounters into the region where the discrimination between deviations and non-deviations is most distinct.

It is important to understand why applying the blockage algorithm improves the results of the CWAM analysis in regions where flights are near the echo top height. High altitude pilots generally use echo tops, or cloud tops with their growing towers and anvils, as the best visual cue of turbulence and convective weather. Observations show that they try to provide a wider margin of deviation around the towering tops to avoid the turbulence, even in the surrounding clear air. When the aircraft is not above the clouds, the echo tops do not provide much correlation with pilot behavior. For pilots flying within the clouds, the reliance on the onboard weather radar will dominate their decision making. In this case, pilots associate the level 3+ weather with turbulence. Thus, while flying in instrument conditions, pilots will use radar to find the widest gaps through the level 3 weather, much like the blockage algorithm begins by finding gaps through the avoidable weather. Capturing the appropriate VIL avoidance behavior for flights at and below the echo top height is the main reason the blockage algorithm filtering step improves the CWAM analysis.

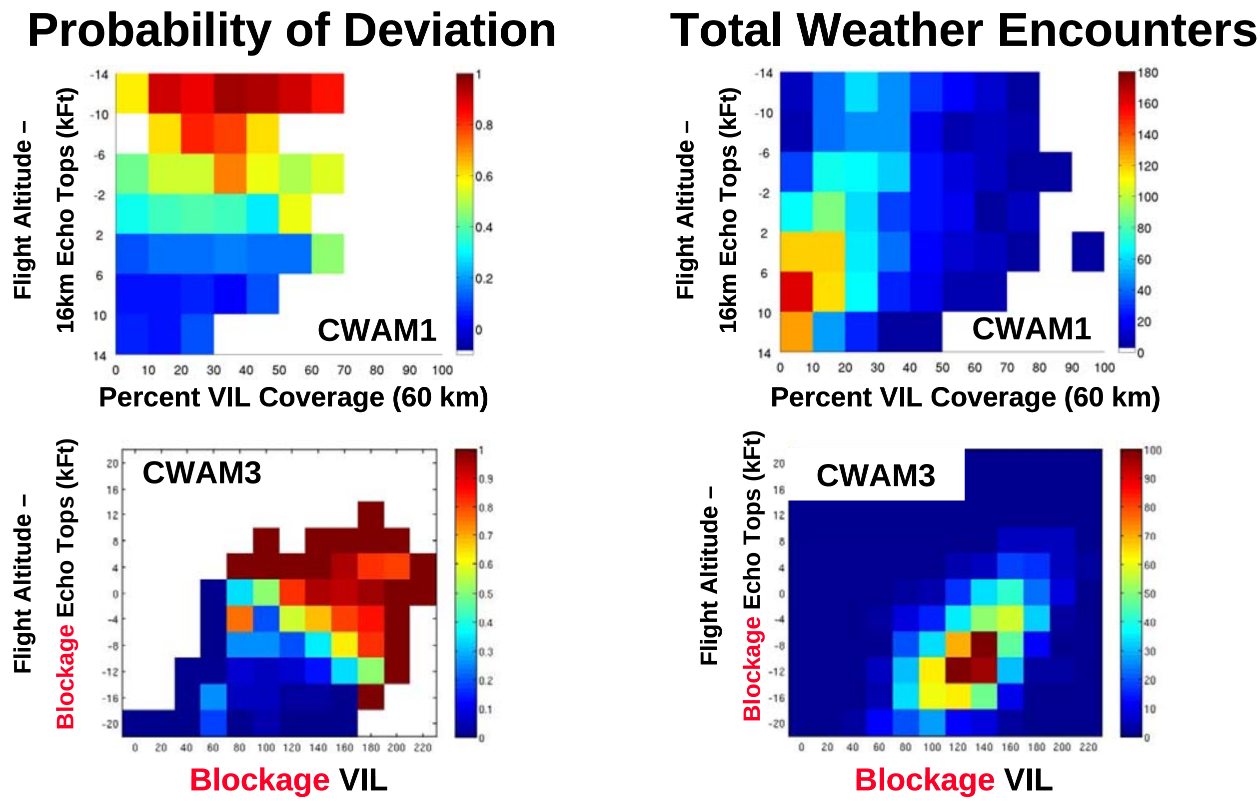

Figure 3. Two dimensional histograms showing the results of the original Convective Weather Avoidance Model (a and c) versus the results of applying the blockage algorithm to the VIL and echo top fields (b and d). The probability of deviation is shown on the left column, while the total weather encounters is shown on the right column.

3.4 Comparison of Blockage to Traditional Spatial Filters

The benefits of the blockage algorithm when applied to the VIL and echo tops can be observed when compared to some traditional spatial filters used on weather radar data. Figures 4, 5, and 6 compare the blockage algorithm with a mean filter and an area coverage filter applied to three different convective weather scenarios. Each of these figures contains the VIL results in the top row, and the echo tops results in the bottom row. The original VIL and echo tops are shown in the first column, a mean filter in the second column, an area coverage filter in the third column, and the blockage algorithm in the fourth column. For both VIL and echo tops the mean is a 41 x 41km kernel centered on the pixel. For area coverage, the VIL is a 61 x 61km kernel of level 3 and greater pixels and for echo tops the kernel is 61 x 61km for tops greater than or equal to 30,000 feet. Figure 4 is from 18:54 UTC on July 10, 2007 over the upper Midwest for a growing convective line moving towards Chicago. Figure 5 is 18:59 UTC on July 10th over New York and western New England, illustrating a case of isolated convective cells. Figure 6 is a complex of disorganized convective cells over Kentucky at 17:05 UTC on July 10, 2007.

Each weather scenario demonstrates different advantages of the blockage algorithm compared with other techniques. In Figure 4 the developing convective line is very strong and has several areas of intense precipitation with high echo tops. These regions are strong enough to impact aviation for aircraft traversing the line on a southeast - northwest course. Comparing the mean VIL and the blockage VIL demonstrates that the intensity of the convective cells is maintained with the blockage VIL, while the mean VIL at this scale smears out and reduces the intensity of the convection. In fact, the blockage calculation maintains the level 5 intensity of the convective line over northern Illinois, while the mean VIL reduces the intensity to level 2 – even though the 41 x 41 scale of the mean filter is comparable to the 55 x 40 scale of the blockage route segments. This difference is very important when determining what is an acceptable flight route or a route requiring a deviation within automated decision support tools.

Figure 5 depicts a collection of small, isolated convective cells near Lake Erie, in central NY and in western Massachusetts. These types of storms do not typically require large pilot deviations or significant rerouting. In most instances, pilots are able to maneuver around these cells using the visual cues of cloud structure and onboard radar. A comparison of the mean and blockage panels shows that isolated cells are eliminated or significantly reduced in the blockage algorithm, while maintained to some degree in the mean calculation.

When the population of cells grows large enough that the gaps between cells fill in and pilots are no longer able to perform minimal maneuvers around the cells, then they will produce impacts on ATC. The blockage algorithm maintains the cells in northeastern NY both in the VIL and echo tops as potentially aviation impacting.

Figure 6 is a case over Kentucky of a fairly large disorganized complex of level 3+ precipitation, with fairly low echo tops. This type of weather does not have a significant impact on ATC for the high altitude aircraft due to the low tops. However, measuring the performance of the forecast algorithms requires that the intensity of the precipitation be correctly captured in the VIL forecast. In this case, there is a significant difference between the mean VIL and the blockage VIL. The mean smears out the intensity of the precipitation, reducing most to level 2. However, the blockage algorithm maintains the intensity of the level 4 weather in the core of the complex where the coverage of level 3+ is highest. In many ways the blockage looks very similar to the area coverage, but it retains more details and does not need wholesale recalibration to relate to hazard level and/or pilot deviation potential.

The blockage algorithm as a spatial filter is a very powerful tool for analyzing the weather and the deterministic forecasts. Unlike the other filters, the most important aspect of the blockage approach is that it takes into account the decision making process of pilots who are constrained to a route structure. The blockage algorithm reduces significantly or eliminates the weather that does not have an impact on ATC operations, yet maintains the intensity and spatial scale of weather that is of importance to traffic flow management.

10 July 2007 18:54 UTC

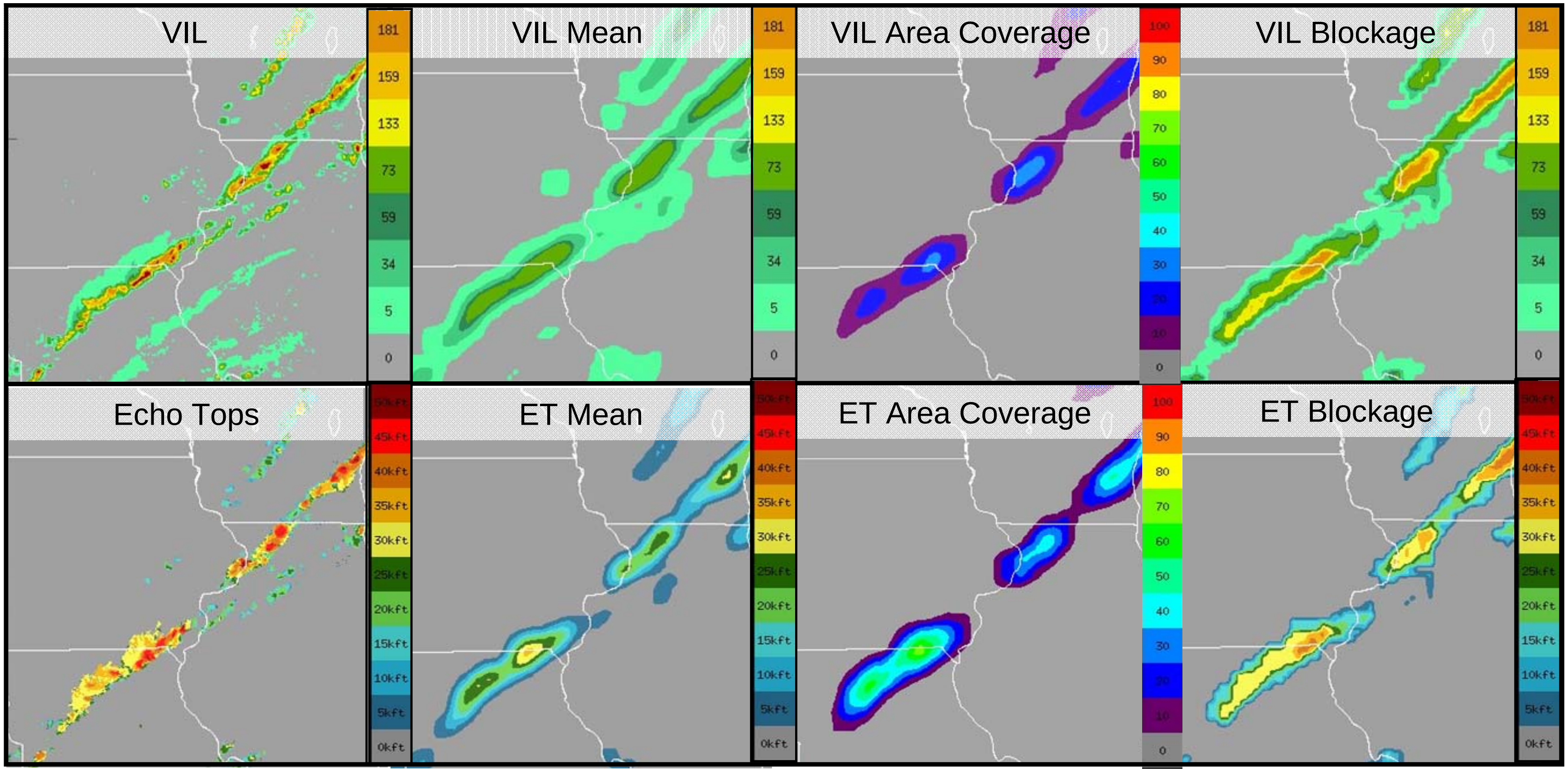

Figure 4. (a) VIL, (b) 41km VIL Mean, (c) 61km VIL area coverage > level 3, (d) Blockage VIL, (e) echo tops, (f) 41km echo tops Mean, (g) 61km echo tops area coverage > 30kft, (h) and Blockage Echo tops on July 10, 2007 at 18:54 UTC for a convective line forming over the upper Midwest.

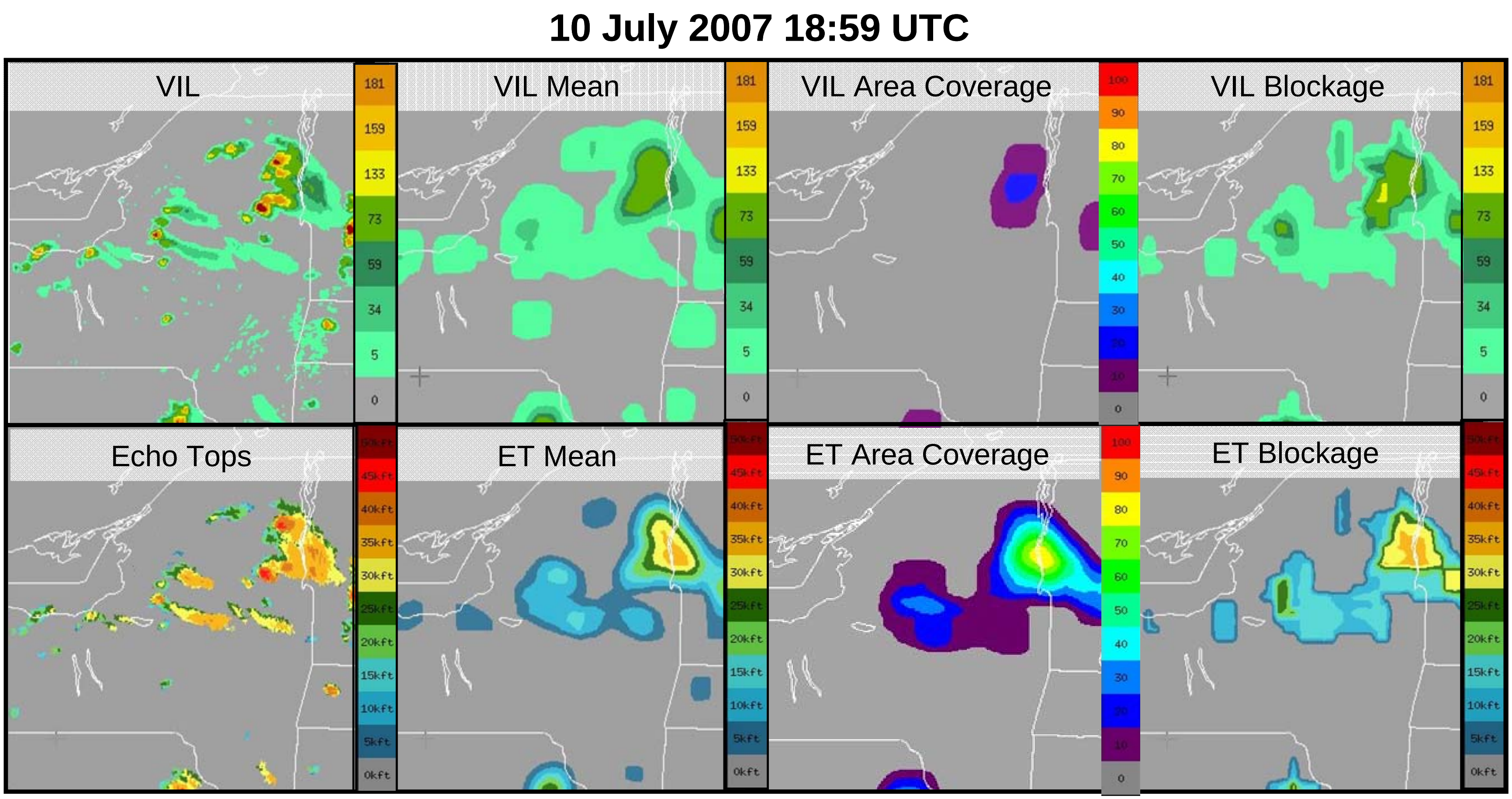

Figure 5. (a) VIL, (b) 41km VIL Mean, (c) 61km VIL area coverage > level 3, (d) Blockage VIL, (e) echo tops, (f) 41km echo tops Mean, (g) 61km echo tops area coverage > 30kft, (h) and Blockage echo tops on July 10, 2007 at 18:54 UTC for a collection of isolated convective cells forming over New York and western New England.

10 July 2007 17:05 UTC

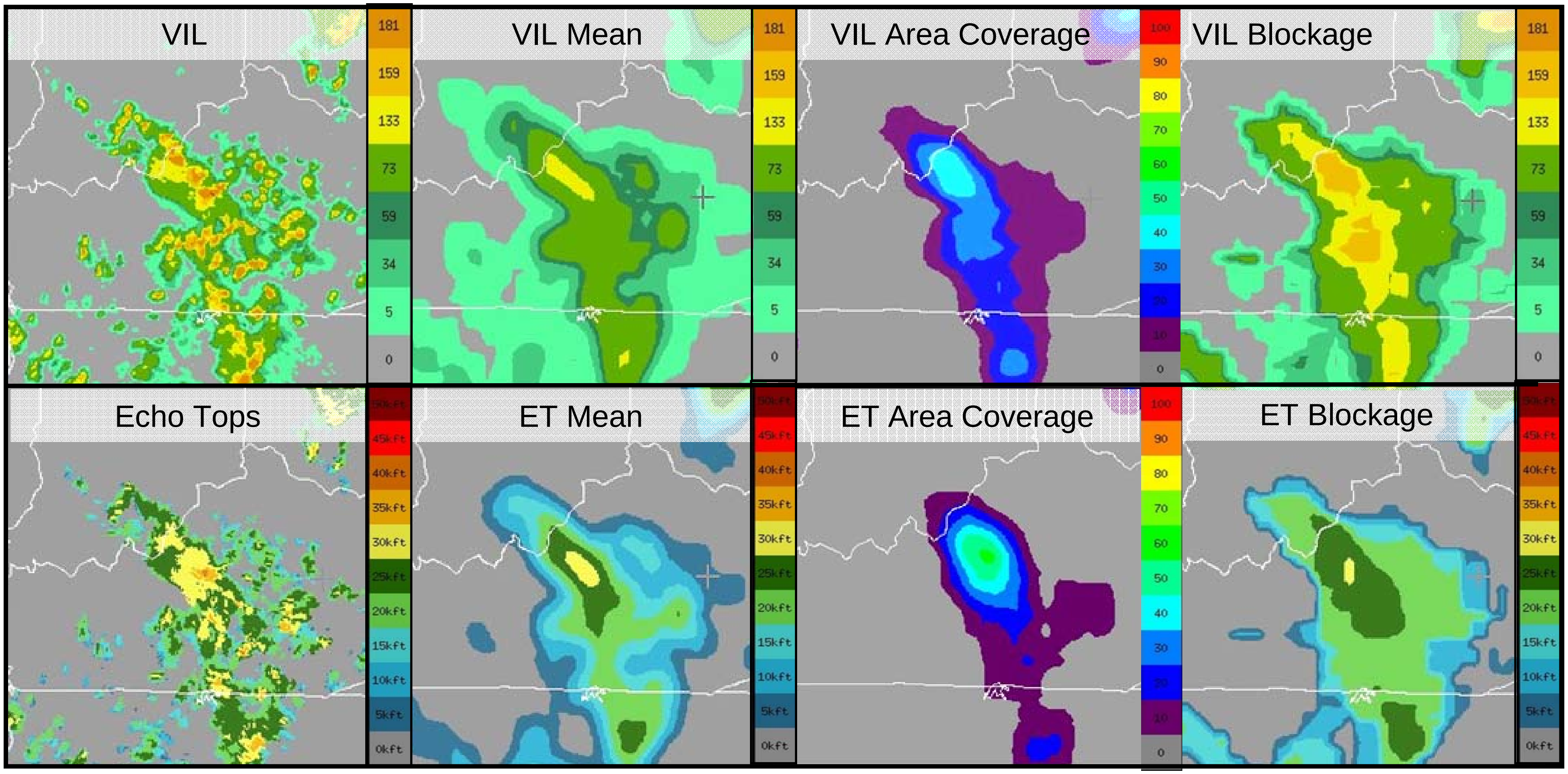

Figure 6. (a) VIL, (b) 41km VIL Mean, (c) 61km VIL area coverage > level 3, (d) Blockage VIL, (e) echo tops, (f) 41km echo tops Mean, (g) 61km echo tops area coverage > 30kft, (h) and Blockage echo tops on July 10, 2007 at 17:05 UTC for a disorganized complex of convective cells over Kentucky.

4. NEW FORECAST SCORING METRIC

The challenge of assessing the performance of convective weather forecast products has been ongoing for many years. Traditional methods of scoring the weather forecasts have used very specific localized comparisons of precipitation levels or applied very broad measures to large areas. Comparing high resolution (1km) data pixel-to-pixel from a system such as CIWS produces results which can be affected by small variations in the weather which may not be of concern to ATM operations. Applying broad area coverage or mean estimates, as shown in the previous section, can appropriately de-emphasize insignificant small storms, but can also grossly underestimate very important features that are of concern to air traffic management.

Techniques to probe the forecasts and truth using flight paths have long been a goal of the research community (Brasunas and Merritt, 1983). More recently researchers have focused on comparing a forecasted capacity estimate of sectors within air traffic control system with the actual capacity (Klein et al., 2008). However, the usefulness of such methods may be limited for classifying the real time forecast performance due to the required *a priori* knowledge of the route and sector structure. Scoring the performance of a convective storm forecast not impacting an ATC route would not be possible, yet knowledge of that performance would be important if the storm were moving into managed ATC air space.

The goal of this work has been to produce a new way of scoring the convective weather forecasts that is appropriate for traffic flow management operations. The new blockage model has been shown to simulate the decision making process pilots use to avoid convective weather in coordination with air traffic control in the en route airspace. We now apply this technique to the forecasted weather, as well as the eventual truth verification of those forecasts, as a proposed new scoring metric and compare it with other scoring methods.

Figure 7 is comparison between two traditional scoring techniques and the new blockage technique. The forecasted VIL values are shown on the x-axis and the verification or truth VIL values are shown on the y-axis. The data used in the comparison is a five hour period on July 10, 2007. This was a very active convective day with multiple weather types occurring over the domain. The VIL in this analysis is the digitized VIL used within the CIWS system. Values range from 0 to 255, with level 3 ranging from 133 and 159. The 15, 20, 60, and 120 minute forecasts are shown from left to right in the columns. The forecast score is represented with a probability distribution for each forecasted value on the x-axis according to the color scale shown. Therefore, the sum of the vertical columns will equal one.

One very simple scoring method is to compare each bin's forecasted value with the corresponding verification "truth" value. In Figure 7, the first row is a score of the pixel-to-pixel or binary VIL. This scoring method shows that the forecast performance decreases rapidly with time until there is little forecast skill at the 120 min. But there are also a disproportionate number of verification pixels that fall on the zero line of the forecasted VIL, meaning there is a high likelihood that forecasted weather at 60 or 120 minutes will be matched with no weather in the truth data. Unfortunately, this is a misrepresentation of the forecast quality because is it highly likely that a slight displacement in the forecast is responsible for this error. In fact, aviation impacting weather may be in close proximity to these zero verification pixels. Scoring forecasts at too high a resolution (e.g., 1 km in the case of CIWS) causes the forecast quality to drop sharply with increasing forecast lead time and does not represent the value of these forecasts to ATC operational users.

Another simple method of scoring the forecast performance is to apply a mean filter to the image. This type of spatial filtering of the data will smooth the forecast and the truth, producing much better results when scored. The middle row in Figure 7 shows a 41 km mean applied to the VIL. The 41 x 41 km kernel for the mean was chosen to correspond to roughly the width of the air routes as used within RAPT and the blockage algorithm. The performance of the forecast does improve with this method, but the amount of weather greater than level 3 is significantly reduced due to the smoothing and the forecast values are biased low relative to the truth (i.e., more values below the diagonal line).

The results of the forecast performance comparison using the blockage algorithm as a spatial filter are shown in the last row of Figure 7. At 15 and 30 minutes a tight unbiased correlation is observed between the forecasted blockage VIL and the verification blockage VIL. Forecast times from 60 minutes out to 120 minutes also demonstrate a broader but significant, largely unbiased correlation. In fact, at the 60 minute time horizon, a forecast of level 4 (159 digitized units) blockage VIL is unlikely to verify as anything less than level 2. Also significant is the large population of strong blockage scores that verify even at longer lead times.

Through many years of interaction with ATC users, MIT Lincoln Laboratory developers have recognized that small forecast displacement errors have little impact on the ability of ATC managers to use the CIWS forecast for strategic planning purposes. Working with the operational users, a scoring method was devised for the Integrated Terminal Weather System (ITWS) and also used in CIWS that performs large kernel searches for similar weather to account for acceptable displacement errors in the forecast (Theriault et al., 2001). These techniques use a 19 x 19 km search box to score the forecasts at 30 and 60 minutes and a 39 x 39 km box for 120 minutes. For each pixel containing level 3+ precipitation, the algorithm searches over the designated neighborhood for a minimum number of pixels with equal or greater value. For instance, for a pixel with level 3+ VIL to be classified as a hit, at least 16 level 3+ VIL pixels over the 19 x 19 km search box must be present. If there are less than 16 truth pixels, the pixel is classified as a false alarm. Instances where level 3+ weather exists in the verification but not in the forecast are classified as misses.

A typical hit/miss/false alarm image is shown in Figure 8 for the CIWS system on July 10, 2007 using level 2 weather as the scoring threshold. Regions of green indicate correct forecasts of level 2+ weather, red indicates false alarms, and blue indicates missed forecasts. The image in the upper left panel represents the most stringent pixel-by-pixel binary scoring method. The image in the upper right represents the CIWS “user scores”, calculated according to the method described in the previous paragraph. CIWS creates the final score known as a Critical Success Index (CSI) from this forecast “user score”. The CSI is the ratio of hits to the sum of hits, misses, and false alarms. The CSI score is computed over very large regions associated with geographically important “home” areas for ATC operations (e.g., surrounding each major terminal). The typical CSI “home” for CIWS is 560 x 560km. Over time, CIWS users have become very comfortable with how the CIWS CSI scores can be interpreted into operational meaningful results.

For the image in the bottom panel of Figure 8 we have replaced the scoring kernels used by CIWS with the non-directional blockage algorithm discussed in Section 2. It is quite reassuring that the panel using the blockage algorithm has many of the same characteristics as the traditional CIWS user and binary forecast scores. The missed forecast of a newly developing line over Illinois and Wisconsin is clearly evident, while the excellent forecasting of the storms over Ohio is also visible. The major difference between the images is that the small details visible in the traditional scoring methods have been eliminated in the blockage method. Also of interest is the ability of the blockage scoring to show the error in the forecast associated with the edges of the weather. The most common errors in CIWS forecasts are due to missing new growth or decay, or due to small variations in the location of the leading edge of the storms. The CIWS user scores do not capture the error in the leading edges very well, while the blockage algorithm actually does. In fact, the results using the blockage algorithm look very similar to the binary scores minus the small details that are insignificant to ATC operations.

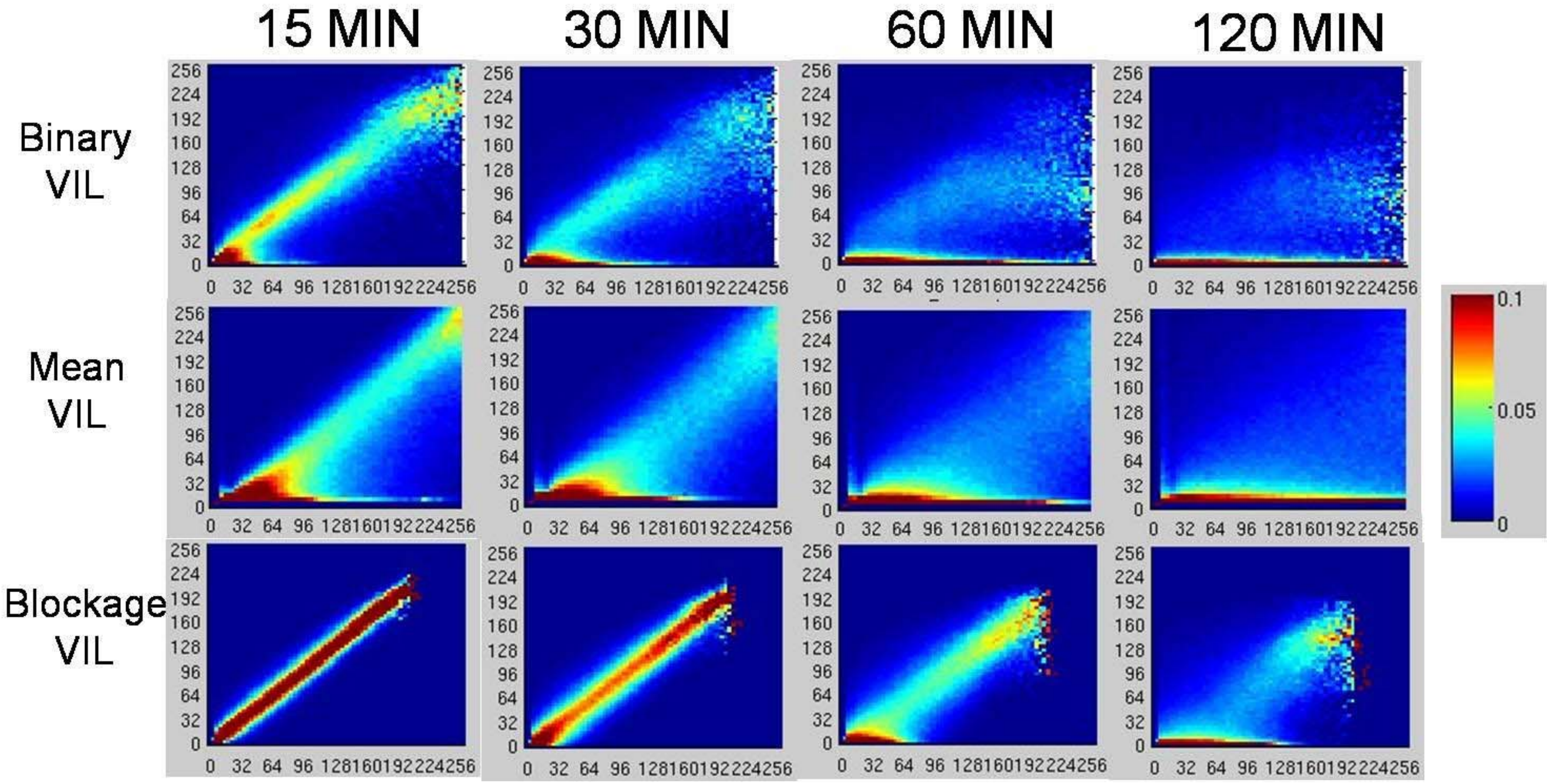

Figure 7. Comparison of probability distribution of 15 minute (1st column), 30 minute (2nd column), 60 minute (3rd column) and 120 minute (4th column) forecasts for three different scoring metrics. The binary VIL comparison (1st row) is a pixel to pixel score of the forecast, the mean VIL (2nd row) is a 41x41km mean applied to the forecast and truth, the blockage VIL (3rd row) is applying the blockage algorithm to both the forecast and truth. The forecasted values are on the x-axis, the truth are on the y-axis.

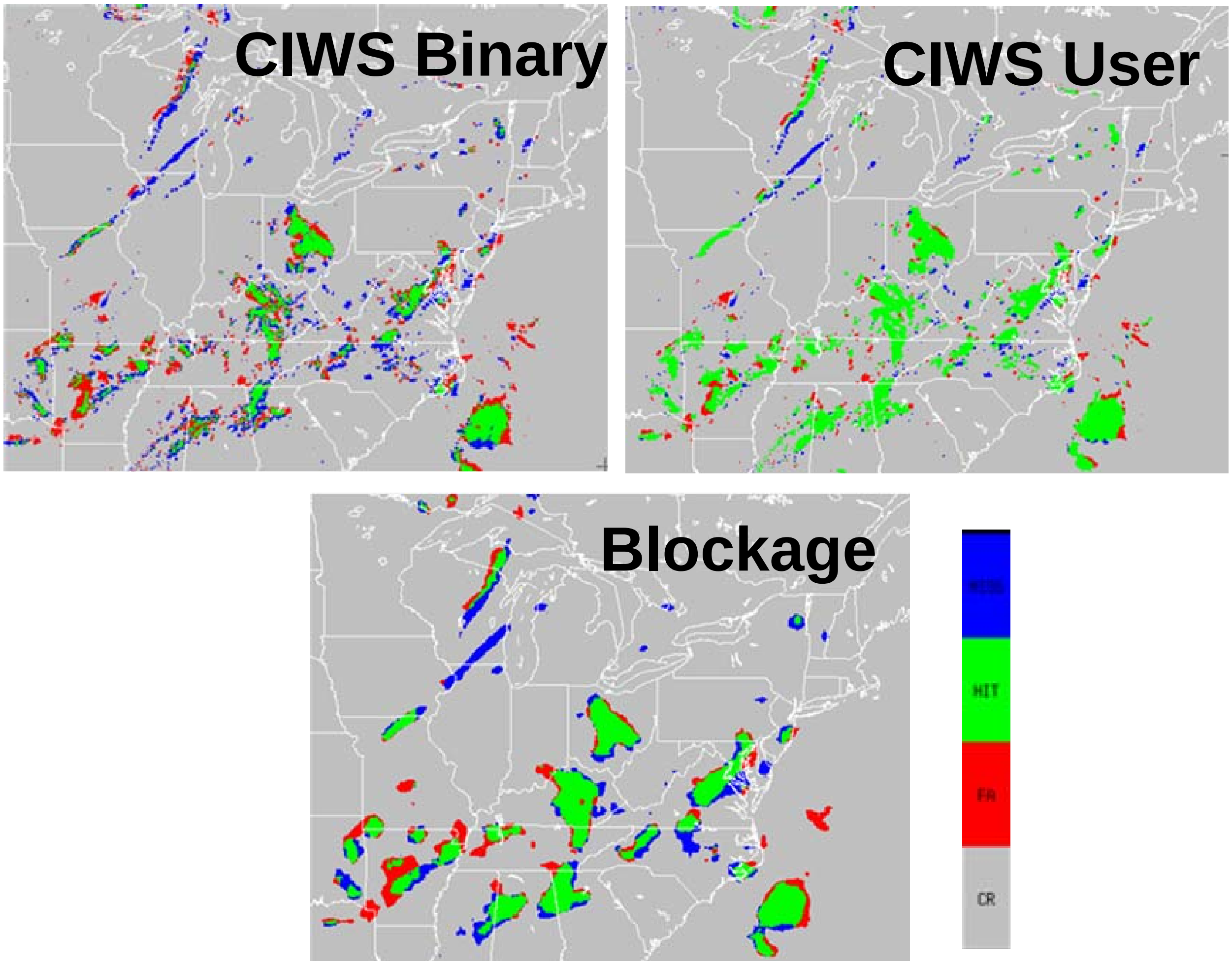

Figure 8. Comparison of current binary CSI scores (left panel), user CSI scores (middle panel) with blockage CSI scores (center panel) for the 60 minute forecast of VIL issued at 1715 UTC on 10 July 2007. Scores were computed for level 2 VIL (0.76 kg m^{-2}). Pixels of green are hits, blue are misses, and red are false alarms.

5. CONCLUSIONS AND FUTURE WORK

This paper has presented the concepts of a new forecast scoring technique using a traffic flow management oriented operational approach. The new method, known as the blockage algorithm, takes into account the nominal width of air routes in the NAS and information on pilot behavior to spatially filter the weather information. The algorithm looks for the best possible path through a rectangle centered on the region of interest that is 40 kilometers wide and 55 kilometers long. The method is an approach that takes into account the direction of travel of the aircraft. Along with finding the best possible path, the algorithm does a distance weighting of all pixels within the rectangle to account for the amount of convective weather present near the aircraft's flight path.

Using the data base assembled by MIT Lincoln Laboratory in conjunction with NASA from the Convective Weather Avoidance Model work, the algorithm's performance was validated as an excellent discriminator of pilot deviation due to the convective weather along planned flight routes. The blockage algorithm also demonstrated a significant improvement in reducing the amount of over-warning present in the current Weather Avoidance Fields that are used within the Route Availability Planning Tool.

Comparison of some traditional scoring methods and the blockage algorithm has been performed and the results show a significant improvement. This new technique shows excellent correlation between the CIWS forecasted VIL product and the verification at 15 and 30 min while still demonstrating utility and overall lack of bias at the 60 and 120 min time horizons. We have also shown how this new technique could be used in place of the current CIWS user scoring technique that has been widely accepted by ATC users.

Future work will continue to analyze the performance of the Convective Weather Avoidance Model and focus on increasing the size of the CWAM data set. Currently, the data base consists of 1,665 convective weather encounters, with roughly one-third of these as documented deviations due to convective weather. Additional data is being analyzed from the summers of 2007 and 2008 and will be used in validating the models.

With the development of a new scoring technique that is highly correlated with the pilot's decision making model, the focus of this effort can now shift to providing a real-time prediction of the forecast error at each time horizon. The deterministic forecast map along with the estimated error at each pixel provides the full probabilistic forecast information needed for NextGen automated trajectory planning operations. We will be analyzing the performance of the forecast algorithms using this new technique in many different weather scenarios and over many different forecast time horizons (including the more strategic 2-8 hr forecasts.)

6. REFERENCES

Brasunas, J. C. and Merritt, M. W., **Short-Term Prediction of High Reflectivity Contours for Aviation Safety,** 9th Conference Aerospace and Aeronautical Meteorology, Omaha, NE, *Amer. Meteor. Soc.*, 1983.

DeLaura, R. A., Evans, J. E., **An Exploratory Study of Modeling Enroute Pilot Convective Storm Flight Deviation Behavior**, 12th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), Atlanta, GA, *Amer. Meteor. Soc.*, 2006. [Paper]

DeLaura, R. A., 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. [Paper]

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 (ARAM), New Orleans, LA, *Amer. Meteor. Soc.*, 2008. [Paper]

Dupree, W. J., Wolfson, M. M., Johnson Jr., R. J., Boldi, R. A., Mann, E. B., Calden, K. T., Wilson, C. A., Bieringer, P. E., Martin, B. D., Iskenderian, H., **FAA Tactical Weather Forecasting in the United States National Airspace,** World Weather Research Program Symposium on Nowcasting and Very Short Term Forecasts, Toulouse, France, 2005. [Paper]

Evans, J. E., Ducot, E. R., **Corridor Integrated Weather System**, MIT Lincoln Laboratory Journal, Volume 16, Number 1, 2006. [Journal Article]

Kay, M., Mahoney, J., Hart, J., **An Analysis of CCFP Forecast Performance for the 2005 Convective Season** 12th Conference on Aviation, Range, and Aerospace Meteorology (ARAM), Atlanta, GA, *Amer. Meteor. Soc.*, 2006.

Klein, A., Cook, L., Wood, B., **Airspace Availability Estimates for Traffic Flow Management Using the Scanning Method,** 27th Digital Avionics Systems Conference (DASC), St. Paul, MN, 2008.

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

Martin, B. D., Evans, J. E., DeLaura, R. A., **Results of an Exploratory Study to Develop a Model for Route Availability in en Route Airspace as a Function of Actual Weather Coverage and Type**, Project Report NASA/A-7, MIT Lincoln Laboratory, Lexington, MA, 2006. [Report]

Martin, B. D., **Model Estimates of Traffic Reduction in Storm Impacted En Route**, 7th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Belfast, Ireland, 2007.

Theriault, K. E., Wolfson, M. M., Dupree, W. J., Forman, B. E., Hallowell, R. G., Johnson Jr., R. J., Moore, M. P., **TCWF Algorithm Assessment – Memphis 2000**, Project Report ATC-297, MIT Lincoln Laboratory, Lexington, MA, 2001. [Report]

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

Wolfson, M. M. Forman, B. E., Calden, K. T., Dupree, W. J., Johnson Jr., R. J., Boldi, R., Wilson, C. A., Bieringer, P., Mann, E. B., Morgan, J., **Tactical 0-2 Hour Convective Weather Forecasts for FAA**, 11th Conf. on Aviation, Range and Aerospace Meteorology, Hyannis, MA, 2004. [Paper]