# Convective Weather Forecast Accuracy Analysis at Center and Sector Levels

*Yao Wang, NASA Ames Research Center, Moffett Field, California*
*Banavar Sridhar, NASA Ames Research Center, Moffett Field, California*

## Abstract

This paper presents a detailed convective forecast accuracy analysis at center and sector levels. The study is aimed to provide more meaningful forecast verification measures to aviation community, as well as to obtain useful information leading to the improvements in the weather translation capacity models.

In general, the vast majority of forecast verification efforts over past decades have been on the calculation of traditional standard verification measure scores over forecast and observation data analyses onto grids. These verification measures based on the binary classification have been applied in quality assurance of weather forecast products at the national level for many years. Our research focuses on the forecast at the center and sector levels. We calculate the standard forecast verification measure scores for en-route air traffic centers and sectors first, followed by conducting the forecast validation analysis and related verification measures for weather intensities and locations at centers and sectors levels. An approach to improve the prediction of sector weather coverage by multiple sector forecasts is then developed. The weather severe intensity assessment was carried out by using the correlations between forecast and actual weather observation airspace coverage. The weather forecast accuracy on horizontal location was assessed by examining the forecast errors. The improvement in prediction of weather coverage was determined by the correlation between actual sector weather coverage and prediction.

The analysis was accomplished by using observed and forecasted Convective Weather Avoidance Model (CWAM) data collected from June to September in 2007. CWAM zero-minute forecast data with aircraft avoidance probability of 60% and 80% are used as the actual weather observation. All forecast measurements are based on 30-minute, 60-minute, 90-minute, and 120-minute forecasts with the same avoidance probabilities.

The forecast accuracy analysis for times under one-hour showed that the errors in intensity and location for center forecast are relatively low. For example, 1-hour forecast intensity and horizontal location errors for ZDC center were about 0.12 and 0.13. However, the correlation between sector 1-hour forecast and actual weather coverage was weak, for sector ZDC32, about 32% of the total variation of observation weather intensity was unexplained by forecast; the sector horizontal location error was about 0.10.

The paper also introduces an approach to estimate the sector three-dimensional actual weather coverage by using multiple sector forecasts, which turned out to produce better predictions. Using Multiple Linear Regression (MLR) model for this approach, the correlations between actual observation and the multiple sector forecast model prediction improved by several percents at 95% confidence level in comparison with single sector forecast.

## 1. Introduction

Flight delays have been a serious problem in the National Airspace System (NAS) for many years. Statistics shows that approximately 70% of the delays are attributed to weather and up to two third of weather delays could be avoidable [1]. Severe convective weather accounts for a significant fraction of these delays. Because of the significant economic losses caused by these delays, there is a need to improve the air traffic flow management (TFM) operations.

TFM manages air traffic flow to balance the air traffic demand against en route airspace capacity that has been reduced by convective weather. During severe weather, TFM may perform advanced airspace management planning. For example, the air traffic controllers may ask flights to hold or to change routes to stay clear of weather to maintain safety. Currently, most weather support to TFM is done manually, with

weather displays that must be interpreted by air traffic managers and controllers. Since the inaccurate estimate of the reduction of sector capacities often produces unnecessary delays, it is vital to develop automated decision support tools based on the reliable models to translate weather forecast information into its impacts on airspace.

Efforts have been made during the past several years to characterize the dynamic capacity during convective weather based on short-term and long-term forecasts. Several algorithms translating weather forecast information into appropriate TFM constraints, such as sector dynamic capacity [2-7], have been developed. These algorithms are based on concepts, ranging from simple weather precipitation intensity coverage in sector to more complex methods utilizing weather intensity and location information in sector, air traffic patterns, and traffic flows. Given the uncertainty associated with forecast of the weather intensity and location, one might wonder if more sophisticated models would always produce better estimates of dynamic sector capacity than simpler ones. In other words, forecast uncertainty bounds must be taken into accounts in estimates and predictions.

The weather forecast uncertainty directly affects the accuracy of the predicted dynamic sector capacity. The forecast verification efforts over the past few decades have been focused on the calculation of traditional standard verification measure scores over forecast and observation data analyses onto grids [8-10]. These verification measures are based on binary classification and have been applied in quality assurance of weather forecast products at the national level for many years. These measures are often difficult to interpret in meaningful physical terms, and are insufficient to determine the accuracy of forecast on weather intensity and location at air traffic center and sector levels.

In this paper, the standard forecast verification measure scores computed using the binary classification method for en-route air traffic centers and sectors are presented. The enhanced measures to evaluate the accuracy and errors of deterministic weather forecast upon the weather severity, weather location, and timing of weather intensity at center and sector levels were discussed. An approach to improve the prediction of sector weather coverage by using multiple sector forecasts was analyzed. The paper is organized as follows. A brief review of the Convective Weather Avoidance Model (CWAM) weather forecast and the descriptions of forecast verification mechanics and evaluation measures used for forecast evaluation are provided in section 2. The results using these forecast evaluation measures at air traffic centers are presented in section 3. Section 4 reports the results on CWAM forecast evaluations for sectors. An approach for improving sector weather forecast coverage accuracy by using forecast in multiple sectors is presented in section 5. Finally, conclusion remarks are provided in Section 6.

# 2. Forecast Evaluation Methodology

### 2.1 CWAM Forecast Data

CWAM is developed using the Corridor Integrated Weather System (CIWS), a product of MIT Lincoln Laboratory (LL) [11]. CIWS combines data from dozens of weather radars with satellite data, surface observations, and numerical weather models in order to improve the accuracy and timeliness of storm severity information. It provides automated, real-time, high spatial resolution (about 1km), three-dimensional forecast information of storms at a 5-minute update rate. CIWS data includes precipitation measured by the vertically integrated liquid (VIL) and the (storm) echo tops forecast. CIWS is used for air traffic en route control tactical planning with 0-120 minute forecast lead time and in forecast increments of 15 minutes. A study on CIWS accuracy at forecasting regions of storm reflectivity has been published recently [2, 12].

Based on CIWS, LL has developed the Convective Weather Avoidance Model (CWAM). CWAM models the flight deviation behavior around severe weather as a function of reflectivity level and echo tops [13] and translates convective weather information from CIWS data into Weather Avoidance Field (WAF) at each flight altitude level. The WAF provides an estimate of the probability of aircraft deviation around severe weather in en route airspace as a function of horizontal location. For each of the WAF files, the data include high spatial resolution weather polygons of predicted aircraft avoidance with different probability thresholds, such as 60% or 80%, for each of the flight levels from 25000 ft (FL250) up to 45000 ft (FL450), in 1000 ft increments. The CWAM data include forecasted WAFs from 0 to 2 hours with a 15 minutes forecast

interval and an update rate of 5 minutes. The weather information used in this paper is from the initial Convective Weather Avoidance Model, CWAM1. CWAM2 and CWAM3 versions were developed by LL, the performance of the two models varied slightly from CWAM1, especially on horizontal levels [12, 14].

We select a representative CWAM 60% or 80% deviation probability threshold to covert the corresponding WAFs into a deterministic forecast and then use a forecast evaluation approach that is similar to previous studies of deterministic weather forecasts. CWAM weather data from June to September in 2007 is used for this study. CWAM zero-minute forecast data with 60% and 80% aircraft avoidance probabilities are used as the actual weather observation and all forecast measurements are based on same avoidance probability CWAM forecasts at 30-minute, 60-minute, 90-minute, and 120-minute time horizons.

### 2.2 Standard Forecast Verification Measures

The quality assessment on the performance of convective weather forecast products has been ongoing for many years. Over the past decades, a number of studies have been conducted to evaluate the forecast accuracy of the aviation convective weather products using traditional standard forecast verification measures. The standard statistical verification measures are computed from a basic 2X2 contingency table, Table 1, for a dichotomous evaluation (e.g., "Yes"/"No"). The "Yes"/"No" forecast/observation pairs are used to create the statistics in the study.

**Table 1 Basic contingency table for dichotomous ("Yes"/"No") events. Elements in the cells are the pixel counts of forecast-observation pairs**

| | | **Observation** | |
|---|---|---|---|
| | | **Yes** | **No** |
| **Forecast** | **Yes** | YY | YN |
| | **No** | NY | NN |

The method adopts the concept of binary classification in which the forecasts and the actual weather observations are compared pixel by pixel on a high-resolution grid in the specific horizontal region for analysis. All grid pixels in the region were initially assigned either a "Yes" or a "No" value as follows: for a forecast, a "Yes" pixel on the grid designates that it has intersection with any part of the weather forecast polygons, whereas a "No" means otherwise; for an observation, pixels on the grid with a "Yes" value indicate the actual weather observation area, whereas a "No" indicates no weather was observed. A true positive pixel of forecast is assigned if the pixel is a "Yes" pixel for both forecast and observation and a true negative pixel of forecast assigned if the pixel is a "No" pixel for both forecast and observation. The True Positive pixel counts, denoted as YY, are calculated as the total true positive pixels of forecasts in the region. The True Negative pixel counts, NN, are computed as the total true negative pixels of forecasts. The False Positive (YN) is then equal to the number of total "Yes" forecast pixels minus the YY. The False Negative (NY) is the difference between the total observation "Yes" pixels and the YY. The total number of pixels in the region on the grid, denoted as N, is the sum of YY, YN, NY and NN.

Using the dichotomous forecast statistics generated in Table 1, some frequently used forecast verification measures are described briefly below. More information about these measures can be found in Ref 15-18.

The Probability of Detection for true positives (*PODy*) is the proportion of "Yes" observed events that were correctly forecast. $PODy = YY / (YY + NY)$. It has a range of 0 to 1. If $NY = 0$, then the score goes to 1, which is the best value possible.

The Probability of Detection for true negatives (*PODn*) is the proportion of "No" observed events that were correctly forecast. $PODn = NN / (NN + YN)$. Its values also range from 0 to 1. If $YN = 0$, then the score goes to 1, the best one can expect.

The False Alarm Ratio (*FAR*) is the proportion of "Yes" values incorrectly forecasted. $FAR = YN / (YY + YN)$. It has a range from 0 to 1. If $YY = 0$, then the score goes 1, which is the worst value possible. For a perfect forecast, $PODy=1$ and $FAR=0$ (or $PODn=1$).

The Critical Success Index (*CSI*) is the proportion of true positives that were either forecast or observed. $CSI = YY / (YY + NY + YN)$. Its values range from 0 to 1 with a value of 1 indicating a

perfect forecast score. The *CSI* is more complete measure than PODy, PODn, or FAR. It depends on both false positives and false negatives, namely the *CSI* is sensitive to the accuracy of both forecast weather intensity and locations.

The Gilbert Skill Score (*Gilbert*) is the *CSI* corrected for the number of true positives expected by random chance. $Gilbert = (YY - C) / (YY - C + YN + NY)$, where C is the number of true positives expected by chance. The probability of true positives expected by random chance is $(YY + YN)*(YY + NY) / N^2$. So that, $C = (YY + YN)*(YY + NY) / N$.

The True Skill Statistic (*TSS*) is a measure of the ability of the forecast to discriminate between "Yes" and "No" observations. $TSS = PODy + PODn - 1$.

The Heidke Skill Score (Heidke) is the percent correct (true positives and true negatives of forecast) adjusted by the number expected by random chance. $Heidke = (YY + NN - B) / (N - B)$, where $B$ is the number of true positives and true negatives expected by chance. $B= [(YY + YN)*(YY + NY) + (NN + YN)*(NN + NY)] / N$.

Bias Rate (*BR*) is defined as the ratio of the number of "Yes" forecasts to the average number of "Yes" forecasts and "Yes" observations: $BR = (YY + YN) / (2*YY + NY+YN)$. The BR value ranges from 0 to 2, it is a measure of over- (>1) or under-forecasting (<1). An unbiased forecast has a *BR* value of 1.

*FWC* is the ratio of the forecast domain area where convection is expected to occur and the region area. $FWC = (YY + YN) / N$. This measure does not depend on the observations.

*OWC* is the ratio of the weather observation area and the region area. $OWC = (YY + NY) / N$. This measure does not depend on the forecast.

### 2.3 Forecast Evaluation Measures for Intensity and Location

The forecast evaluation developed in this paper focuses on the capacity prediction accuracy for centers and sectors during severe weather. The existing models for translating en route convective weather forecasts into weather impacted airspace capacities rely on the accuracy of forecast of the weather intensities and weather locations [5-6, 19-22, 24]. The standard verification measures give statistical tools to evaluate the forecast and compare different forecast products. However, the measures do not provide quantitative values on the accuracy of forecast intensity and location in terms of errors. The forecast evaluation methods described in this subsection are used to verify the accuracy of forecast of the weather intensity and location quantitatively at center and sector levels.

For many years, sector weather coverage has been used as a measure of weather intensity to estimate reduction in airspace capacity. The evaluation measures of convective weather observation coverage (*OWC*), weather forecast coverage (*FWC*), and Forecast Bias rate (*BR*) are used to score the forecast accuracy on weather intensity qualitatively. Higher values of weather coverage indicate worse weather in a sector. To evaluate the forecast intensity accuracies quantitatively, the correlation between *FWC* and *OWC* is used to calculate the forecast intensity errors. The correlation might provide not only the knowledge on the forecast intensity accuracies in the airspaces but also the prediction accuracy of the airspace capacity using forecast weather coverage method.

To study the impact of forecast location accuracy on airspace sector capacity predictions alone, we can examine the cases in which forecast weather intensity is assumed to be as same as the actual weather intensity in the airspace. Under this condition ($BR = 1$), YN is as same as NY. Forecast location error rate for 1-mile at least can be estimated by YN/(YY+YN) or approximately defined by ((YN+ NY)/2.0)/(YY + (YN+ NY)/2.0) = (1-*CSI*)/(1+*CSI*) if BR is close to 1.

Since comparison of high resolution (1-mile) forecast pixels to observation pixels can be affected by small variations in weather location that are not of concern to air traffic management operation, a revised YY, denoted as $YY_{20}$, was used to evaluate the forecast accuracy for the horizontal location. Based on the same forecast and observation weather "Yes" /"No" grids, the true positive counts, $YY_{20}$, of binary classification were re-calculated using a distance threshold of 20 miles between the forecast and observation "Yes" pixels in a given airspace (a center or a sector). A true positive forecast was assigned to a "Yes" forecast pixel if any observation "Yes" pixels in the same sector or center could be

detected within 20 miles of its location. Likewise, a true positive observation was assigned to a "Yes" observation pixel if its distance to any forecast pixels in the same sector or center is within 20 miles. The true positive counts, $YY_{20}$, is defined as the minimum of the total true positive pixels of forecasts and the total true positive pixels of observations in a center or a sector to avoid double counting. Obviously, YY is never greater than $YY_{20}$. The corresponding false positive $YN_{20}$ is equal to the number of total "Yes" forecast pixels in the sector or center minus $YY_{20}$. The false negative $NY_{20}$ is the difference between the total observation "Yes" pixels in the sector or center and $YY_{20}$.

The forecast Location Error Rate, *LER*, is defined as

$$LER = (YN_{20} + NY_{20}) / (2*YY_{20} + YN_{20} + NY_{20}) \quad (1)$$

When the airspace *FB* is 1, then $YN_{20} = NY_{20}$ and $LER = YN_{20} / (YY_{20} + YN_{20}) = NY_{20} / (YY_{20} + NY_{20})$. So the *LER*s give the errors on the forecasted horizontal locations if *FB* is close to 1. For example, given the sector *FB* of 1, a *LER* of 0.2 means about 20% of the weather forecast in the airspace occurred at locations more than 20 miles from the actual observed weather. The *LER* value ranges from 0 to 1. A *LER* value of 0 indicates that forecasts match actual observations within 20 miles and a *LER* value of 1 means that the all forecasts are more than 20 miles from weather observations. The *LER* values give the inaccuracy of the forecast horizontal location to the observation in the airspace. For evaluations of the forecast location accuracies using airspace *FLE*, only those data where the *BR* values are in the range of 0.9 and 1.1 were collected and analyzed in this study. This *BR* requirement is to ensure that the forecast weather intensity is close to the actual weather intensity in the airspace so that *LER* would depend on the location accuracy only.

### 2.4 Correlation Analysis

Correlation analysis between forecast and actual weather coverage was used to evaluate the center and sector forecast accuracy on weather intensity. The strength of the linear relationship between the weather forecast and actual weather observation is described by the Pearson correlation coefficient $r$ and 95% confidence lower and upper bounds. The confidence bounds of the correlation are based on an asymptotic normal distribution of $0.5*\log((1+r)/(1-r))$ assumption, with an approximate variance equal to $1/(n-3)$. The $r_l$ and $r_u$ denote lower and upper bounds for a 95% confidence interval, respectively. These bounds would be accurate for large samples when the two variables have multivariate normal distributions. A Pearson correlation greater than .8 is generally considered as strong whereas a correlation of less than .5 is generally treated as weak. As an example, if $r = .80$, then $r^2 = .64$, which means that 64% of the total variation in the actual weather can be explained by the forecast using a linear relationship. The other 36% of the total variation of actual weather remains unexplained by forecast. In this paper, $1- r^2$ has been used to measure the weather intensity error for the forecast quantitatively.

The Spearman's rank correlation coefficient $\rho$ was also used for evaluating the nonlinear relationship between forecast and observation. The 95% confidence interval (*CI*) around the sample Spearman correlation was constructed and derived from 5000 "non-parametric" bootstrap samples of observed frequency distribution of the estimates.

## 3. Evaluations of CWAM Forecast for Centers

The forecast analysis is based on the CWAM data collected from the five centers in northeast region of US. The five centers are Chicago (ZAU), Washington (ZDC), Indianapolis (ZID), New York (ZNY), and Cleveland center (ZOB).

### 3.1 Standard Forecast Verification Measures

The average summaries of standard verification measures are computed using CWAM data on ten bad weather days (6/12, 7/09, 7/10, 7/11, 7/17, 7/18, 7/19, 7/19 and 8/9) during the convective season of 2007. The verification measure scores are listed in Table 2 and Table 3 for 60-minute and 120-minute forecasts respectively. The measures are calculated using the same 60% avoidance probability at FL300 on the grid of 1X1 mile pixels.

These 2 tables indict two general trends. First, there is over-forecasting in all cases (*BR*>1). Second, for any particular center, both the forecast weather coverage (*FWC*) and Bias Rate (*BR*) show little change in different time horizons. For all other verification measures, the scores in Table 3 are worse than that in Table 2, meaning that short term forecast

are more accurate as compared to long term ones. That is, we have a better estimate about what the weather will be in 1 hour than that in 2 hours.

Table 4 displays the standard verification measure average scores computed using the 120-minute forecast with the 80% aircraft deviation threshold. Compared to Table 3, the scores are worse.

**Table 2 Verification statistical table for 60-minute forecast**

| Center | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU** | .735 | .930 | .434 | .473 | .418 | .666 | .560 | 1.17 | .169 |
| **ZDC** | .653 | .963 | .506 | .387 | .361 | .618 | .509 | 1.15 | .074 |
| **ZID** | .643 | .948 | .528 | .370 | .335 | .593 | .477 | 1.17 | .097 |
| **ZNY** | .635 | .928 | .499 | .386 | .337 | .565 | .479 | 1.14 | .132 |
| **ZOB** | .655 | .952 | .560 | .351 | .324 | .608 | .465 | 1.22 | .083 |

**Table 3 Verification statistical table for 120-minute forecast**

| Center | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU** | .614 | .889 | .541 | .349 | .278 | .504 | .406 | 1.17 | .191 |
| **ZDC** | .477 | .947 | .635 | .247 | .216 | .425 | .336 | 1.13 | .079 |
| **ZID** | .435 | .931 | .651 | .233 | .192 | .367 | .296 | 1.11 | .100 |
| **ZNY** | .462 | .913 | .598 | .266 | .211 | .376 | .321 | 1.07 | .129 |
| **ZOB** | .478 | .936 | .674 | .232 | .200 | .416 | .311 | 1.21 | .090 |

**Table 4 Verification statistical summaries for 120-minute forecast (80%)**

| Center | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU** | .566 | .908 | .629 | .282 | .233 | .475 | .351 | 1.24 | .141 |
| **ZDC** | .435 | .952 | .733 | .184 | .162 | .389 | .266 | 1.24 | .061 |
| **ZID** | .392 | .931 | .764 | .162 | .132 | .325 | .218 | 1.24 | .085 |
| **ZNY** | .445 | .911 | .696 | .213 | .171 | .357 | .269 | 1.18 | .114 |
| **ZOB** | .443 | .942 | .768 | .170 | .149 | .386 | .240 | 1.31 | .073 |

**Table 5 Verification statistical summary for ZDC using different pixel size grids**

| Pixel Size | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* | *OWC* |
|---|---|---|---|---|---|---|---|---|---|---|
| **1X1 Mile** | .477 | .947 | .635 | .247 | .216 | .425 | .336 | 1.13 | .079 | .063 |
| **10X10 Mile** | .536 | .931 | .579 | .297 | .253 | .469 | .386 | 1.12 | .108 | .088 |
| **20X20 Mile** | .585 | .918 | .532 | .341 | .285 | .504 | .425 | 1.11 | .137 | .113 |

Table 5 lists the standard verification measures for ZDC center using 120-minute forecast with 60% threshold on the grids with different pixel sizes. For large pixel size (10X10 or 20X20 miles) case, the "Yes" /"No" pixels are assigned using the score box definitions from Figure 2-3 in Ref. 2: a large "Yes" pixel (e.g. for 20X20 mile grid) is assigned if the numbers of 1X1 mile "Yes" pixels in the large size pixel are greater than 16 (about 4% of the total 1X1 mile pixels); a "No" means otherwise. As the pixel size increases, the standard verification measure scores improve. Since the *OWC* are higher for the larger size grid, the actual observation coverage accuracy for the forecast verification is somewhat reduced if the resolution of CWAM observation is about 1 mile. For the subsequent comparisons investigated in this study, we limit our CWAM forecast evaluations on the 1-mile scale.

### 3.2 Forecast Accuracy on Weather Intensity

The results of weather intensity forecast accuracy for ZDC are described in this subsection.

#### 3.2.1 Center Bias Rate

Figure 1 displays the ZDC center *BR* histograms at FL300 (CWAM with avoidance probability of 60% and 80%) for 30-minute, 60-minute, 90-minute, and 120-minute forecasts. The center *BR* for longer forecast times are spread more widely from 1, indicating deterioration in forecast intensity accuracy. It also shows that the forecast intensity accuracy for 60% is better than that for 80%.

#### 3.2.2 Correlation between *FWC* and *OWC*

The correlations between forecast and observed weather coverage of ZDC at FL300 using CWAM with avoidance probability of 60% and 80% for different forecast time horizons are listed in Table 6 and Table 7, respectively. The Pearson correlations are displayed by the correlation coefficient $r$, 95% confidence lower, $r_l$, and upper bounds, $r_u$, and 1- $r^2$. The Spearman's rank correlations $\rho$ and 95% confidence intervals of the rank correlations are also listed in these tables.

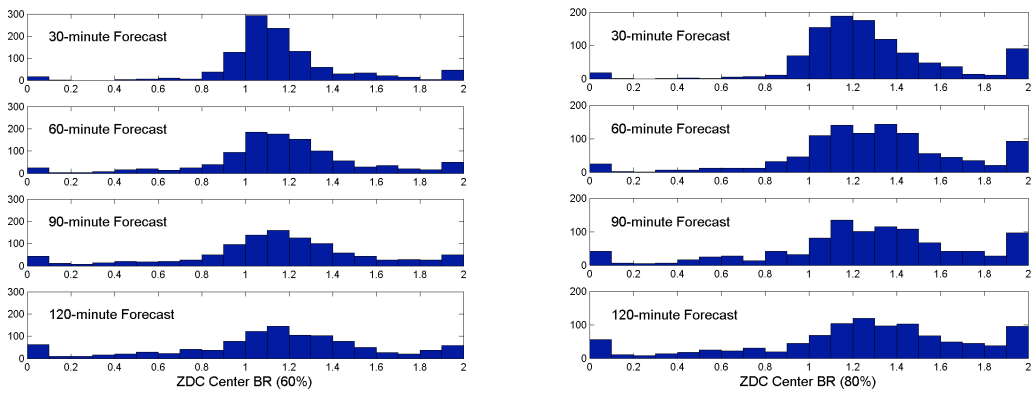


**Figure 1 Center ZDC *BR* histograms for different CWAM forecast times with 60% and 80% avoidance probability, respectively**

**Table 6 Correlations between the center ZDC observed and forecast weather coverage at FL300 (avoidance probability of 60%)**

| Forecast | $r$ | $r_l$ | $r_u$ | 1- $r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **30 min.** | .977 | .974 | .980 | .045 | .966 | .005 |
| **60 min.** | .937 | .929 | .944 | .122 | .923 | .010 |
| **90 min.** | .887 | .873 | .899 | .214 | .872 | .016 |
| **120 min.** | .815 | .794 | .835 | .335 | .814 | .021 |

**Table 7 Correlations between the center ZDC observed and forecast weather coverage at FL300 (avoidance probability of 80%)**

| Forecast | $r$ | $r_l$ | $r_u$ | 1- $r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **30 min.** | .956 | .950 | .961 | .087 | .942 | .007 |
| **60 min.** | .909 | .898 | .919 | .173 | .896 | .012 |
| **90 min.** | .858 | .841 | .873 | .264 | .842 | .018 |
| **120 min.** | .788 | .764 | .810 | .379 | .792 | .022 |

Table 6 shows that about 5%, 12%, 21%, and 34% of the total variation of actual weather intensity (using 1- $r^2$) are unexplained for 30-minute, 60-minute, 90-minute, and 120-minute forecasts with a 60% aircraft avoidance probability, respectively. The correlation, even though all strong (>0.8), tends to be significantly weaker (at a 95% confidence level) as forecast time horizons increase. The strong correlations reflect a strong degree of linear relationship between forecast and observation weather coverage.

Table 7 shows that about 9%, 17%, 26%, and 38% forecast intensity errors for the forecast times of 30-minutes, 60-minutes, 90-minutes, and 120-minute using 80% avoidance probability, respectively. From the two tables, one can observe that longer term forecasts produce weaker correlations between forecast and observation coverage, also the lower avoidance probabilities are associated with the better correlations, even though all the correlations are seemingly strong. This also holds for other centers we studied.

### 3.3 Forecast Accuracy on Horizontal Weather Positions

#### 3.3.1 Center *CSI*

Figure 2 displays the ZDC center *CSI* histograms at FL300 for CWAM forecasts with avoidance probability of 60% and 80% and 30-minute, 60-minute, 90-minute, and 120-minute time intervals. When the *CSI* value is closer to 1, it indicates better forecast intensity and location accuracies qualitatively. This figure clearly reveals the same pattern as shown in section 3.2.2, i.e., with longer-term forecasts producing worse *CSI* scores, worse forecast in weather intensity and location accuracy. These conclusions apply to the forecasts of other centers as well.

#### 3.3.2 Center Forecast Location Error Rate

Table 8 lists the average values of ZDC center Forecast Location Error Rate, denoted as *LER*, at FL300 for forecasts of 30-minute, 60-minute, 90-minute, and 120-minute time intervals with 60% and 80% aircraft avoidance probabilities, respectively. As mentioned before, the data were filtered by the requirement of $0.9<BR<1.1$.

The table shows that, on average, about 9%, 13%, 18%, and 25% of forecast locations were off by more than 20 miles from actual observations for forecast times of 30-minute, 60-minute, 90-minute, and 120-minute, respectively, using 60% aircraft avoidance probability. It also reveals that the drift between forecast and actual weather is higher for 80% avoidance probability than that for the 60% case.

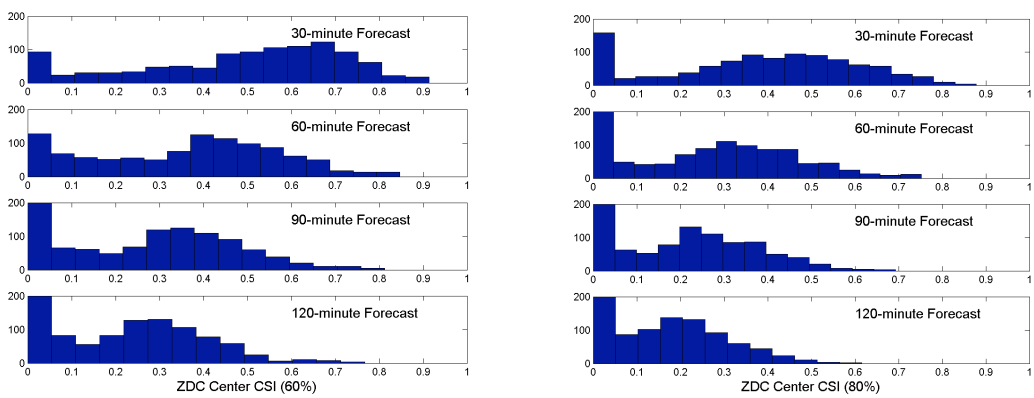


**Figure 2 Center ZDC *CSI* histograms for different CWAM forecast times with 60% and 80% avoidance probability respectively**

**Table 8 Average of center ZDC *LER* (20 miles apart at least)**

| | 60% Prob. | 80% Prob. |
|---|---|---|
| 30-minute | 0.088 | 0.111 |
| 60-minute | 0.130 | 0.195 |

| | | |
|---|---|---|
| 90-minute | 0.183 | 0.244 |
| 120-minute | 0.253 | 0.332 |

# 4. Forecast Evaluation at Sector-level

The forecast analysis for several high-altitude sectors that have high demand and are impacted by high-frequency weather events is described in the follow subsections. The list of sectors follows: Chicago Center sector 83 (ZAU83), Washington Center sector 32 (ZDC32), Indianapolis Center sector 66(ZID66), New York Center sector 42 (ZNY42), and Cleveland Center sector 66 (ZOB66).

### 4.1 Standard Forecast Verification Measures

The analysis is conducted using 60% avoidance probability CWAM data at FL300. The result is a summary of verification measure scores based on the data collected in August, 2007. The measures for 60-minute and 120-minute forecasts are calculated on the 1X1 mile grid and listed in Table 9 and Table 10 respectively.

The results show the same pattern reported in section 3.1 for centers, i.e., for any particular sector, 60-minute and 120-minute forecasts produce very similar average values of the forecast weather coverage (*FWC*) and Bias Rate (*BR*), and over-forecasting for all sectors. For all other measures, the scores in Table 10 are worse than that in Table 9. The forecast quality is getting worse as the forecast time getting longer.

Table 11 shows the average verification measure scores computed using the CWAM 120-minute forecast with 80% deviation threshold at FL300 flight level. Compared with Table 10, the scores are worse in most cases.

Table 12 lists the verification measures for ZDC32 from the same 120-minute CWAM forecast with 60% threshold using grids with different pixel sizes. As the pixel size increases, the *OWC* and *CSI* measure scores increase significantly.

**Table 9 Verification statistical table for 60-minute forecast**

| Sector | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU83** | .733 | .798 | .403 | .498 | .340 | .533 | .468 | 1.13 | .366 |
| **ZDC32** | .618 | .911 | .448 | .412 | .346 | .531 | .476 | 1.09 | .174 |
| **ZID66** | .590 | .893 | .511 | .359 | .285 | .484 | .406 | 1.10 | .204 |
| **ZNY42** | .657 | .821 | .388 | .478 | .319 | .480 | .436 | 1.03 | .339 |
| **ZOB66** | .605 | .882 | .457 | .394 | .298 | .489 | .426 | 1.08 | .229 |

**Table 10 Verification statistical table for 120-minute forecast**

| Sector | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU83** | .601 | .745 | .478 | .391 | .209 | .347 | .303 | 1.09 | .369 |
| **ZDC32** | .461 | .891 | .528 | .290 | .215 | .354 | .317 | 1.01 | .169 |
| **ZID66** | .450 | .845 | .581 | .244 | .156 | .297 | .239 | 1.04 | .204 |
| **ZNY42** | .569 | .764 | .449 | .390 | .198 | .335 | .293 | 1.03 | .339 |
| **ZOB66** | .475 | .837 | .531 | .283 | .169 | .313 | .262 | 1.03 | .229 |

**Table 11 Verification statistical table for 120 minute forecast (80%)**

| Sector | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|
| **ZAU83** | .580 | .733 | .531 | .355 | .178 | .314 | .266 | 1.16 | .367 |
| **ZDC32** | .434 | .900 | .586 | .249 | .190 | .336 | .289 | 1.04 | .139 |
| **ZID66** | .405 | .862 | .640 | .200 | .127 | .269 | .198 | 1.09 | .176 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **ZNY42** | .586 | .744 | .453 | .396 | .206 | .331 | .303 | 1.04 | .360 |
| **ZOB66** | .472 | .828 | .567 | .273 | .164 | .301 | .254 | 1.07 | .228 |

**Table 12 Verification statistical table for ZDC32 using different pixel size grids**

| Size | *PODy* | *PODn* | *FAR* | *CSI* | *Gilbert* | *TSS* | *Heidke* | *BR* | *OWC* | *FWC* |
|---|---|---|---|---|---|---|---|---|---|---|
| **1 Mile** | .461 | .891 | .528 | .290 | .215 | .354 | .317 | 1.01 | .167 | .169 |
| **10 Mile** | .548 | .781 | .456 | .369 | .221 | .331 | .332 | 1.02 | .279 | .284 |
| **20 Mile** | .592 | .730 | .426 | .409 | .211 | .322 | .340 | 1.03 | .356 | .376 |

### 4.2 Forecast Accuracy on Weather Intensity

#### 4.2.1 Sector Bias Rate

Figure 3 displays the ZDC32 sector *BR* histograms at FL300 (CWAM avoidance probability of 60% and 80%) for 30-minute, 60-minute, 90-minute, and 120-minute forecasts. The figure indicates the worsened intensity forecast accuracies for longer forecast time horizons. It also illustrates that the forecast intensity accuracy for 60% avoidance probability is better than that for 80% avoidance probability. These patterns also hold for other sectors.

#### 4.2.2 Correlation between *FWC* and *OWC*

The correlations between forecast and observation weather coverage of ZDC32 sector at FL300 using CWAM with 60% and 80% avoidance probability for different forecast times are listed in Table 13 and Table 14, respectively.

Once again, the same trend is observed in these tables, namely that longer-term forecasts and higher avoidance probabilities produce a lower correlation between forecast and observation coverage. However, when compared with Washington Center (ZDC) in Table 6 and 7, both the linear and nonlinear correlations are poorer. For sector ZDC32, about 32% and 39% of the total variation of actual weather intensity are unexplained by the 1-hour forecast using 60% and 80% aircraft avoidance probabilities, respectively.

The correlations between 60-minute forecasts and observation weather coverage using CWAM 60% and 80% avoidance probabilities for several other high-demands and high-altitude sectors are listed in Table 6 and 7, respectively. Poor forecast intensity accuracy occurred for these sectors as well. This underscores the challenge to provide accurate and useful weather forecasts for smaller areas.

### 4.3 CWAM Forecast Accuracy on Weather Horizontal Positions for Sectors

#### 4.3.1 Sector *CSI*

Figure 4 displays the ZDC32 sector *CSI* histograms at FL300 for CWAM (60% and 80% avoidance probability) forecast times of 30-minute, 60-minute, 90-minute, and 120-minute. The figure shows again that as the time horizon extends out, the forecasts of weather intensity and location accuracy become worse. The same conclusions are true for other sectors as well.

#### 4.3.2 Sector Forecast Horizontal Location Error Rate

Table 17 lists the average of ZDC32 sector forecast Location Error Rate (*LER*) values for CWAM (avoidance probability of 60% and 80% at FL300) forecast times of 30-minute, 60-minute, 90-minute, and 120-minute. As stated before, the data were filtered to include only those having sector *FB* values between 0.9 and 1.1 only.

The table illustrates that on average, about 7%, 10%, 14%, and 15% of forecast locations differ from the actual observations by more than 20 miles in ZDC32 for forecast times of 30-minutes, 60-minutes, 90-minutes, and 120-minutes, respectively, using CWAM avoidance probabilities of 60%. The average forecast location errors are smaller than the corresponding errors in Table 8 for ZDC center. This is understandable because the location errors of forecast and actual weather are limited by the sector boundary.

The average forecast location error *LER* of several other sectors using 60-minute forecast and observation data with the sector *FB* values between

0.9 and 1.1 are listed in Table 18 for 60% and 80% aircraft avoidance probabilities, respectively. The table shows forecast locations differed from the actual weather by more than 20 miles ranged from 8 to 14% for 60% avoidance probability forecast.

### 4.4 Quality Assessment of CWAM Forecast Times

To illustrate the differences between the convective forecast times and actual weather times, sector weather intensities were traced by both 120-minute forecast weather coverage (*FWC*) and the actual observation (*OWC*) at FL300. Figures 5 and 6 show the time series of the weather coverage of sector ZAU83, ZDC32, ZID66, and ZOB66 with CWAM flight avoidance probabilities of 60% and 80% respectively. In these figures, Eastern Daylight Time (EDT) is displayed in minutes on the x-axis, ranging from 0 to 1440; the sector weather intensities for forecast *FWC* (red lines) and observed weather *OWC* (blue line) are on the y-axis, ranging from 0 to 1.

The time distributions in the figures show the forecasts differ from the actual weather, sometimes the two have a time-lagged correlation; and for the cases examined, one tendency seems clear, i.e., the forecast is always more severe than the actual weather is.

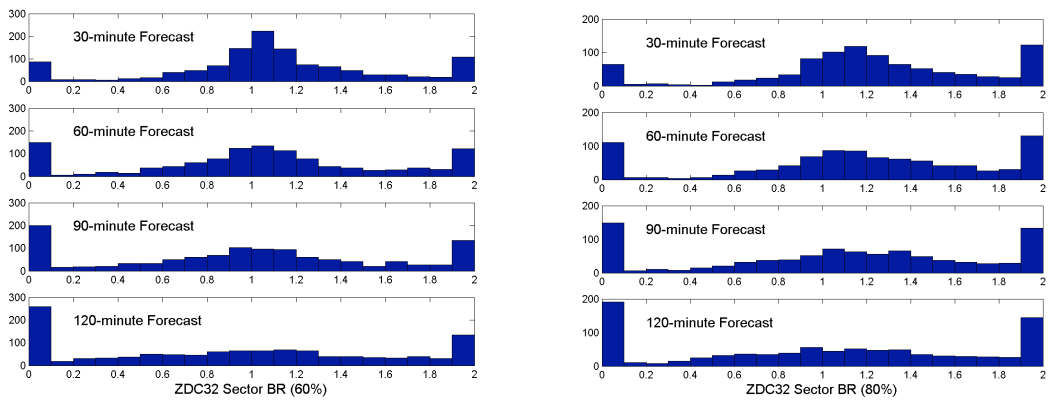


**Figure 3 Sector ZDC32 *BR* histograms for different CWAM forecast times with 60% and 80% avoidance probability, respectively**.

**Table 13 Correlations between the sector ZDC32 observed and forecast weather coverage (aircraft avoidance probability of 60%)**

| Forecast | $r$ | $r_l$ | $r_u$ | $1-r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **30 min.** | .915 | .905 | .923 | .164 | .845 | .019 |
| **60 min.** | .822 | .803 | .840 | .324 | .726 | .031 |
| **90 min.** | .726 | .698 | .752 | .473 | .609 | .043 |
| **120 min.** | .603 | .565 | .638 | .637 | .481 | .048 |

**Table 14 Correlations between the sector ZDC32 observed and forecast weather coverage (aircraft avoidance probability of 80%)**

| Forecast | $r$ | $r_l$ | $r_u$ | $1-r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **30 min.** | .885 | .870 | .898 | .217 | .796 | .027 |
| **60 min.** | .783 | .757 | .807 | .386 | .668 | .039 |
| **90 min.** | .682 | .647 | .715 | .534 | .554 | .048 |
| **120 min.** | .539 | .492 | .583 | .710 | .402 | .057 |

**Table 15 Correlations between observed and 60-minute forecast weather coverage for sectors at FL300 (avoidance probability of 60%)**

| Forecast | $r$ | $r_l$ | $r_u$ | $1-r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **ZAU83** | .827 | .818 | .837 | .316 | .660 | .023 |
| **ZID66** | .724 | .709 | .739 | .475 | .627 | .023 |
| **ZID87** | .784 | .772 | .795 | .386 | .621 | .022 |
| **ZNY42** | .706 | .689 | .722 | .502 | .551 | .028 |
| **ZOB66** | .794 | .783 | .804 | .370 | .631 | .022 |
| **ZTL50** | .710 | .696 | .723 | .496 | .598 | .020 |

**Table 16 Correlations between observed and forecast weather coverage for sectors at FL300 (avoidance probability of 80%)**

| Forecast | $r$ | $r_l$ | $r_u$ | $1-r^2$ | *Spearman's* $\rho$ | *95% CI* |
|---|---|---|---|---|---|---|
| **ZAU83** | .762 | .748 | .776 | .419 | .541 | .030 |
| **ZID66** | .588 | .564 | .612 | .654 | .447 | .034 |
| **ZID87** | .601 | .579 | .623 | .639 | .428 | .032 |
| **ZNY42** | .525 | .496 | .552 | .725 | .402 | .037 |
| **ZOB66** | .663 | .644 | .681 | .561 | .489 | .031 |
| **ZTL50** | .530 | .508 | .552 | .719 | .436 | .028 |

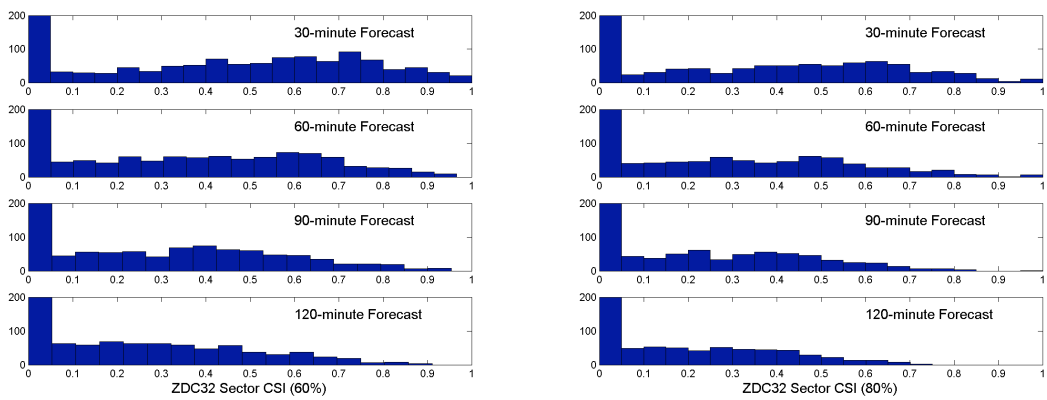


**Figure 4 Sector ZDC *CSI* histograms for different CWAM forecast times with 60% and 80% avoidance probability, respectively.**

**Table 17 Average of Sector ZDC32 *LER* (20 miles apart at least)**

| | Prob. Of 60% | Prob. Of 80% |
|---|---|---|
| 30-minute | 0.068 | 0.081 |
| 60-minute | 0.101 | 0.110 |
| 90-minute | 0.143 | 0.148 |
| 120-minute | 0.150 | 0.193 |

**Table 18 Sector ZDC32 location error rates for 20 miles apart at least (0.9<FB<1.1)**

| | Prob. Of 60% | Prob. Of 80% |
|---|---|---|
| ZAU83 | 0.078 | 0.078 |
| ZID66 | 0.141 | 0.160 |
| ZID87 | 0.087 | 0.137 |
| ZNY42 | 0.101 | 0.102 |
| ZOB66 | 0.108 | 0.118 |
| ZTL50 | 0.135 | 0.207 |

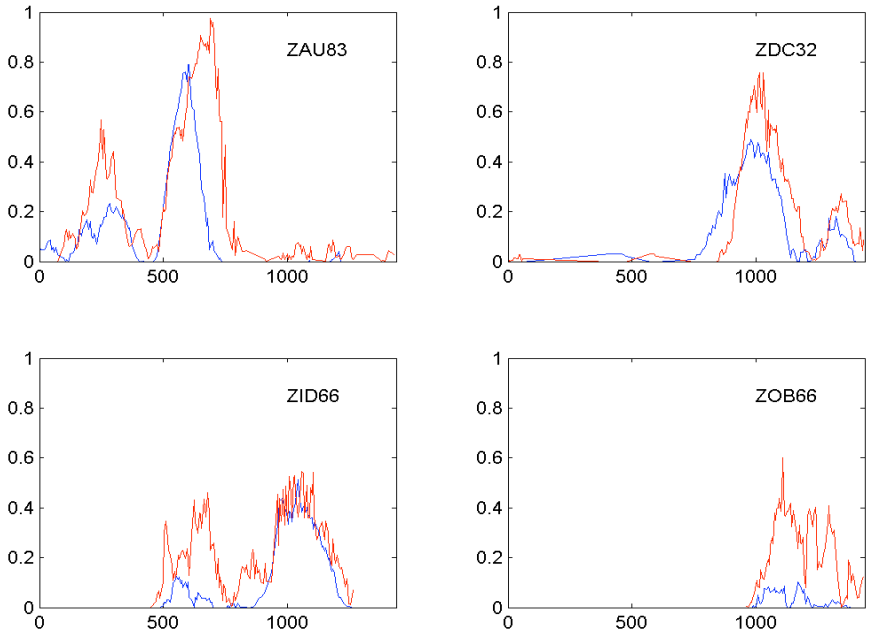


**Figure 5 Sector *FWC* and *OWC* time series distributions of sector ZAU83, ZDC32, ZID66, and ZOB66 for CWAM avoidance probability of 60% on July 17, 2007.**

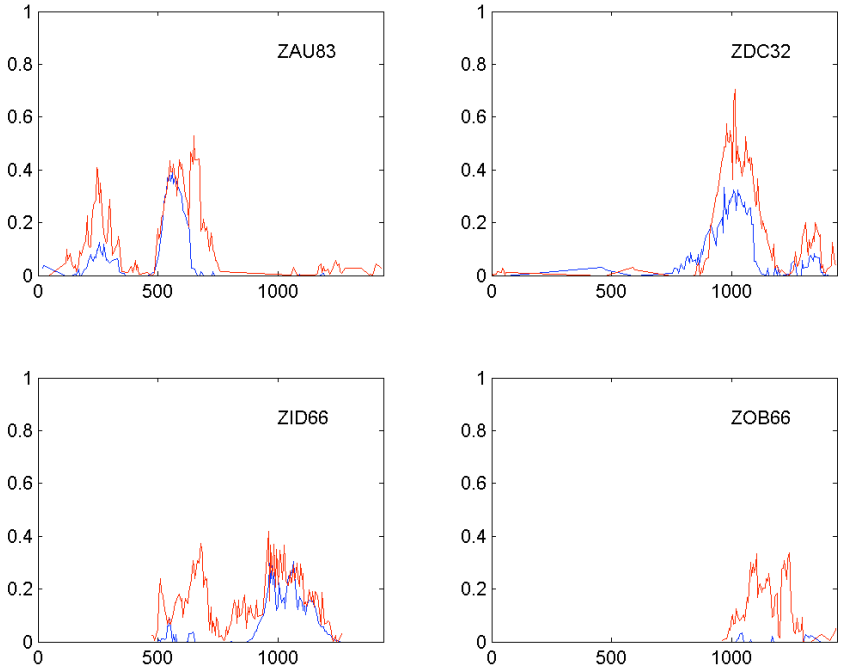


**Figure 6 Sector *FWC* and *OWC* time series distributions of sector ZAU83, ZDC32, ZID66, and ZOB66 for CWAM avoidance probability of 80% on July 17, 2007.**

# 5. Sector Observed Weather Coverage Predicted by Multiple Sector Forecasts

### 5.1 Sector Weather Coverage Index

Airspace convective weather coverage has been used by traffic controllers to estimate the traffic reduction for many years. The 3D convective weather coverage was originally proposed by taking into account the forecasted echo tops at different flight levels in CWAM [23]. The CWAM weather area at a flight level with a given aircraft deviation probability threshold in a sector was used for the weather coverage at that altitude. At each level, the ratio of the coverage in a sector to the sector area, $S_w$, can be calculated. The 3-D sector convective Weather Coverage Index (*WCI*) is then the average of weather coverage for all flight levels inside a sector:

$$\boldsymbol{WCI} = \sum_{k=1}^{nFL} \boldsymbol{w_k} \times \boldsymbol{S_w} \qquad (2)$$

Here $\boldsymbol{w_k} = 1/\boldsymbol{nFL}$ is a constant, and *nFL* is the total number of flight levels separated by 1000 ft in the sector. The *WCI*'s range is from 0 to 1, where 0 means no weather in the sector and the value of 1 indicates a complete coverage by weather.

The *WCI* can be used as a model to translate weather forecast into air traffic and sector capacity. The errors of sector capacity predictions are mainly caused by the inaccuracy of forecast which can be estimated by the correlations between weather forecast coverage, $WCI_f$, and observation coverage, $WCI_o$, for a sector [24]. Table 19 and 20 list the Pearson correlations for different forecast time horizons and sectors with aircraft avoidance probability of 60% and 80%. The results correspond to our previous findings that longer-term forecasts have lower *WCI* correlations, and that lower avoidance probabilities have WCI higher correlation. The correlation between $WCI_f$ and $WCI_o$ ranged from 0.6 to 0.7 for 2 hour predictions with 60% aircraft avoidance probability. That means more than 50% of 120-minute forecasts may not have the precision to be really useful. Therefore, in order to use *WCI* weather translation models in TFM operations one has to improve the prediction for the sector observation weather coverage, especially for 120-minute horizons.

**Table 19 Correlations between observed and forecast *WCI* for sectors (avoidance probability of 60%)**

| Forecast | *30-minute* | *60-minute* | *90-minute* | *120-minute* |
|---|---|---|---|---|
| **ZAU83** | .91 | .86 | .80 | .72 |
| **ZDC32** | .88 | .81 | .72 | .61 |
| **ZID66** | .86 | .79 | .70 | .58 |
| **ZID87** | .91 | .86 | .78 | .69 |
| **ZOB66** | .90 | .85 | .80 | .71 |
| **ZTL50** | .84 | .74 | .62 | .48 |

**Table 20 Correlations between observed and forecast *WCI* for sectors (avoidance probability of 80%)**

| Forecast | *30-minute* | *60-minute* | *90-minute* | *120-minute* |
|---|---|---|---|---|
| **ZAU83** | .85 | .80 | .74 | .66 |
| **ZDC32** | .78 | .71 | .62 | .51 |
| **ZID66** | .76 | .68 | .59 | .48 |
| **ZID87** | .84 | .78 | .70 | .59 |
| **ZOB66** | .83 | .76 | .70 | .60 |
| **ZTL50** | .70 | .59 | .45 | .32 |

### 5.2 An Approach to Improve Sector Observed *WCI* Prediction

The forecast evaluation of airspace intensity errors from section 3 and 4 indicates that the correlation between weather forecast and observation coverage for a center is much better than that for an individual sector. Hence the prediction of a sector actual observation $\boldsymbol{WCI_o}$ could be improved by using not only the sector forecast coverage but also to take into account of the forecast coverage for adjacent sectors.

The sector observed actual 3D weather coverage was determined by the Multiple Linear Regression (MLR) model using multiple sectors ($n$) 3D forecast coverage as,

$$\mathrm{WCI_O} = \sum_{p=1}^{n} \alpha_p \mathrm{WCI_{f}}_p + \beta \qquad (3)$$

Here $\alpha_p$ and $\beta$ can be derived by MLR on training data that include both of the sector $\boldsymbol{WCI_o}$ and $n$ sector $\boldsymbol{WCI_f}$. Consequently, using $\boldsymbol{WCI_f}$ from the test data, the correlation between sector $\boldsymbol{WCI_o}$ and the prediction of $\sum_{p=1}^{n} \alpha_p \mathbf{WCI_{f}}_p + \beta$ can be compared with the correlations between $\boldsymbol{WCI_o}$ and $\boldsymbol{WCI_f}$ for the same sector using the same test data. To validate the results, tenfold cross-validation and paired t-test were used in this paper to compare the two prediction methods.

In tenfold cross-validation the data was first partitioned into 10 equally (or nearly equally) sized segments or folds. Subsequently, 10 iterations of training and validation were performed such that within each iteration a different fold of the data was held out for validation test while the remaining 9 folds were used for MLR process training to compute $\alpha_p$ and $\beta$ . The constructed MLR model and using single sector forecast were then applied to make predictions about the sector actual weather coverage. The accuracy of each was defined by the correlation analysis between the predictions and the observations in the test fold. The paired statistical hypothesis tests, paired Student's t-test, were performed for the assessment of the statistical significances for the difference of two methods obtained through cross-validation.

To illustrate this approach, the results for ZDC32 are shown below. The $\boldsymbol{WCI_o}$ of ZDC32 and 2-hour forecast $\boldsymbol{WCI_f}$ of ZDC32 with 8 adjacent sectors (ZDC04, ZDC12, ZDC16, ZDC36, ZDC52, ZTL33, ZTL42, and ZTL43) were collected using CWAM data with 60% aircraft avoidance probability from July to August in 2007. Tenfold group cross-validation was used for the MLR model validation. The data were randomly divided into 10 groups with similar sample sizes. In cross-validation, a MLR model was constructed, each time by dropping a different group of the data and applying the MLR to the remaining 9 groups of data. The one being dropped was used to test the MLR model. At the end of this procedure, 10 predictions assembled from the dropped cases are compared with the observed target, ZDC32 $\boldsymbol{WCI_o}$, and the correlations between the MLR predictions and the targets were computed. The correlation between ZDC32 $\boldsymbol{WCI_f}$ and ZDC32 $\boldsymbol{WCI_o}$ was also calculated for the 10 individual testing groups (labeled as "Single" in the following tables). For a tenfold cross-validation, the correlation results are listed in Table 21.

As expected the results show that the correlations between the ZDC32 observation weather coverage and nine sector 2-hour forecast MLR predictions are all better than the ones using single ZDC32 2-hour forecast for all 10 cases. Using paired sample t-test to test a null hypothesis that the sample comes from the same normal distribution, resulted in a p-value is about 0.000008. This test rejects the null hypothesis at the default 0.05 significance level, namely the MLR prediction of the actual weather coverage is notably better than that by using the ZDC32 forecast along at least at the 95% confidence level. To ensure the randomness of 10 group selection, the process was repeated for 100 times. All p-values are much less than 0.05 with an average of 0.00016; the repeated experiments for different tenfold cross-validation produced the similar results, which certifies our previous conclusion, even with 50% errors from exact p-value [25]. However the correlation using adjacent sector forecast coverage MLR model is still not strong; the improvement for the correlation is only about 6%.

**Table 21 Tenfold correlation result of MLR and single forecast for ZDC32**

| Group | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MLR | .605 | .693 | .728 | .661 | .669 | .717 | .618 | .658 | .685 | .629 | 0.666 |
| Single | .574 | .637 | .672 | .629 | .608 | .690 | .568 | .611 | .647 | .610 | 0.625 |

## 6. Concluding Remarks

This paper presents standard forecast verification measure scores for weather intensities and locations at center and sector levels. It also presents an approach to improve the prediction of sector weather coverage by multiple sector forecasts. The weather intensity (weather severity) assessment was carried out by using the correlations between forecast and actual weather observation airspace coverage. The weather forecast horizontal location accuracy was evaluated by forecast location errors. The improvement on prediction of weather coverage was determined by the correlation between actual sector weather coverage and prediction. The forecast evaluation and analysis was accomplished by using actual and forecast Convective Weather Avoidance Model data from June through September of 2007.

For forecast times under one-hour, the errors in forecast of weather intensity and location in centers are relatively low. The study showed that the 1-hours forecast intensity and horizontal location errors were about 0.12 and 0.13 for 60% aircraft avoidance probability forecast. We found that the correlations between sector 1-hour forecast and actual weather coverage were not strong, the sector forecast intensity errors ranging from 0.3 to 0.5 for 60% avoidance probability and 0.4 to 0.7 for 80%; the horizontal location errors were relatively small, under 0.15 for 60% avoidance probability and under 0.2 for 80%.

The paper also introduces an approach to estimate the sector three-dimensional actual weather coverage by using multiple sector forecasts which turned out to produce better prediction. Using MLR model for this approach, the correlations between actual observation and the model prediction can be significantly improved by several percents at 95% confidence level.

## Acknowledgments

The authors would like to thank those who helped data collection and the conversion of the CIWS data into CWAM data. In particular, the assistance of Mohamad Refai is noted.

## References

1. Pace, D., "*ATM-Weather Integration Plan Overview"*, FAA AJP-B, Aviation Weather Office, May 27, 2009.
2. Evans J., Weber, M., Wolfson, M., Clark, D. and Newell, O., *"Roadmap for Weather Integration into Traffic Flow Management Modernization (TFM-M)", Project Report ATC-347,* Lincoln Laboratory, MIT, July 24, 2009.
3. Martin, B., Evans J. and DeLaura, R., *"Exploration of A model Relating Route Availability in En Route Airspace To Actual Weather Coverage Parameters", 86th AMS Annual Meeting,* Atlanta, GA, 2006.
4. Song, L., Greenbaum, D. and Wanke, C., *"The Impact of Severe Weather on Sector Capacity", 8th USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)*, Napa, California, USA, June-July 2009.
5. Song, L., Wanke, C. and Greenbaum, D., *"Methodologies for Estimating the Impact of Severe Weather on Airspace Capacity", 8th AIAA Aviation Technology, Integration, and Operations Conference,* Anchorage, Alaska, AIAA Paper 2008-8917, Sept. 14-19, 2008.
6. Klein, A., Cook, L., Wood, B., and Simenauer, D., "Airspace Capacity Estimation Using Flows and Weather-Impacted Traffic Index", Integrated Communications, Navigation and Surveillance Conference, 2008.
7. Klein, A., Kavoussi, S. and Lee, R., *"Weather Forecast Accuracy: Study of Impact on Airport Capacity and Estimation*

*of Avoidable Costs", 8th USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)*, Napa, California, USA, June-July 2009.

8. National Research Council (U.S.). Committee for a Workshop on Weather Forecasting Accuracy for FAA Air Traffic Control, "Weather Forecasting Accuracy for FAA Traffic Flow Management: a workshop report", 2003.
9. Brown, B. and Mahoney, J., *"Quality Assessment of National Convective Forecast Product". Quality Assessment Report provided to the Technical Review Panel of the Aviation Weather Technology Transfer Board,* 2000.
10. Seseske, S., Kay, M., Madine, S., Hart, J., Mahoney, J., and Brown, B., *"Quality Assessment Report: National Convective Weather Forecast 2 (NCWF-2)"*, March 2006.
11. Evans, J. and Ducot, E., *"Corridor Integrated Weather System",* MIT Lincoln Laboratory Journal, Volume 16, Number 1, 2006.
12. Matthews, M., Wolfson, M., Delaura, R., Evans, J., Reiche, C., Balakrishnan, H., and Michalek, D., *"Measuring The Uncertainty of Weather Forecasts Specific to Air Traffic Management Operations", The 89th American Meteorological Society Annual Meeting,* January 2009.
13. DeLaura, R., Robinson, M., Pawlak, M., and Evans, J., *"Modeling Convective Weather Avoidance In En route Airspace", MIT Lincoln Laboratory Project report NASA/A-6,* June 2006
14. DeLaura, R., and Evans, J., *"An Exploratory Study of Modeling En route Pilot Convective Storm Flight Deviation Behavior", 13th Conference on Aviation, Range and Aerospace Meteorology,* New Orleans, LA, January 20-24, 2008.
15. Mahoney, J. L., J. K. Henderson, B. G. Brown, J. E. Hart, A. Loughe, C. Fischer, and B. Sigren, "*The Real-Time Verification System (RTVS) and its application to aviation weather forecasts*", *10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 2002.
16. Wilks, D.S., "*Statistical Methods in the Atmospheric Science",* Academic Press, 467 pp., 1995.
17. Doswell, C.A., Davies-Jones, R., and Keller, D. L., "*On summary measures of skill in rare event forecasting based on contingency tables*", *Wea. and Forec.*, **5**, 576-585, 1990.
18. Hudson, H.R. and Foss, F.P., "*The Collaborative Convective Forecast Product from the Aviation Weather Center's Perspective*", *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Oregon, WA, Amer. Meteor. Soc., 73-76, 2002.
19. Martin, B., Evans J. and DeLaura, R., *"*Model Estimates of Traffic Reduction in Storm Impacted En Route Airspace*", 7th AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference,* Belfast, Northern Ireland, 2007.
20. Song, L., Greenbaum, D. and Wanke, C., *"The Impact of Severe Weather on Sector Capacity", 8th USA/Europe ATM R&D Seminar*, Napa, California, USA, June 29 – July 2, 2009.
21. Krozel, J., Mitchell, S., Polishchuk, V., and Prete, J., *"Capacity Estimation for Airspaces with Convective Weather Constraints", AIAA Guidance, Navigation, and Control Conf., Hilton Head, SC,* 2007.
22. Klein, A., Cook, L., and Wood, B., *"Airspace Availablility Estimation for Traffic Flow Management Using the Scanning Method", 27th Digital Avionics Systems (DASC) Conference, St. Paul*, MN, Oct. 2008.
23. Krozel, J. and Doble, N. A., *"*Simulation of the National Airspace System in Inclement Weather*", AIAA Modeling and Simulation Technologies Conference,* Hilton Head, SC, Aug. 20-23, 2007.
24. Wang, Y. and Grabbe, S., *"Comparison of Sector Capacity Weather Translation Models", NASA/TM-2010-216401, 2010.*

25. Menke, J. and Martinez, T., *"Using Permutations Instead of Student's t Distribution for p-values in Paired-Difference Algorithm Comparisons", IEEE, 2004.*