

# **The NDFD Thunderstorm Probability Field: An Assessment of Forecast Lead Time and Spatial Displacement**

Prepared By:

Quality Assessment Product Development Team  
NOAA/ESRL/GSD

Steven A. Lack<sup>2</sup>, Andrew F. Loughe<sup>2</sup>, Michael P. Kay<sup>2</sup>,  
Melissa A. Petty<sup>3</sup>, and Jennifer L. Mahoney<sup>1</sup>

September 12, 2012

Affiliations:

- 1 – National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD)
- 2- Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD
- 3 – Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

Corresponding Author:

J.L. Mahoney NOAA/ESRL/GSD, 325 Broadway, Boulder, CO 80303;  
Jennifer.Mahoney@noaa.gov

## Executive Summary

This report summarizes a quality assessment of the National Digital Forecast Database (NDFD) Thunderstorm Probability field in terms of lead-time in forecasting the onset and cessation of significant convective weather (storms with echo tops  $\geq 30$  kft) in the vicinity of 35 major U.S. airports, and in terms of location errors in the forecast placement of hazardous weather. For the 2010/2011 convective seasons (June – September), performance measures were computed and a sensitivity analysis was performed by varying the forecast threshold that defines the likelihood of thunderstorms, the coverage amount within a terminal area, and by varying the temporal length of observed events. For July 2011, an initial investigation of jetway constraints in common AFP boundary regions (FCAA05/08) was also performed. Primary findings from the present study include:

- NDFD convective forecasts in their current form are unable to resolve the finer temporal and spatial scales that are required for forecasts of onset and cessation at the terminal/tactical scale:
  - The scheduled forecast update cycle at individual Weather Forecast Offices (WFOs) limits NDFD's effective use when considering impact along high-traffic arrival and departure routes and for en route traffic within congested corridors.
  - Event coverage over the terminal space is far greater for NDFD forecasts than it is for observed convection. It was determined that this is a characteristic of forecast feature resolution, not the NDFD grid spacing.
- NDFD forecast characteristics do not capture attributes of convection such as growth, decay, and advection:
  - There is no significant movement of forecast convective systems into, out of, or through terminal areas.
  - Forecast objects are not well correlated with observations in either space or time.
- Onset of significant convective weather is generally forecast to occur too early while cessation is typically forecast to occur too late.
- NDFD temporal and spatial accuracy is better for onset than it is for cessation.
- Skill is quite similar for the scattered-and-above and the likely-and-above convective thresholds.
- For  $\pm 3$  hour temporal precision, there is a decrease in skill with increasing lead time for all regions, but this is not as noticeable for the extended lead times (out to 8 hours) due to the persistent nature of coarse-resolution NDFD forecasts.
- Displacement errors are greater at the time of cessation than at the time of onset, and the median displacement error for each of these exceeds 30 nmi.
- Skill varies by region. Displacement errors are higher in the Northeast than in the Southeast according to the predominate form of convection in each region.

# Table of Contents

1	Introduction .....	1
2	TRWG requirements .....	1
3	Data characteristics and constraints .....	3
3.1	NDFD forecasts.....	3
3.2	Observations: Corridor Integrated Weather System.....	4
4	Event definition for onset and cessation.....	5
5	Scoring.....	9
5.1	Creation of contingency tables in event space .....	9
5.2	Lead-time to onset and cessation.....	11
5.3	Spatial displacement error .....	11
6	Results .....	13
6.1	Sensitivity to forecast threshold, coverage threshold, and temporal precision .....	13
6.2	Regional performance .....	17
6.3	Displacement results.....	19
6.4	Displacement results by time of day or season .....	22
6.5	Skill summary with MOC requirements.....	24
7	En-route considerations.....	25
8	Conclusions.....	29
9	Future work.....	30
	Acknowledgments .....	30
	References .....	30

## List of Figures

Figure 3.1: Available forecast lead times (maroon) at 3-h valid-time increments (green boxes along the top) for the operational NDFD product. Since this study focuses on 2, 4, 6, and 8-h lead times, the number of available forecasts is limited by the 3-h valid time increment.....	3
Figure 3.2: Sample of NDFD general thunderstorm probability categories from 02 July 2011 issued at 2247 UTC, with a 4-h lead-time, and valid on 03 July 2011 at 0247 UTC. Broad yellow regions indicate <i>scattered</i> thunderstorm forecasts, while orange indicates <i>likely</i> thunderstorms.....	4
Figure 3.3: CIWS echo top field valid from 03 July 2011 at 0245 UTC, corresponding to the NDFD thunderstorm grid in Figure 3.2. The blue circles identify the terminal areas of interest in this study, with a 75-nmi radius (150-nmi diameter).....	5
Figure 4.1: A broad association of events is illustrated to the left (overlap or near-overlap in time), with an increasing precision constraint applied as the forecast temporal resolution becomes tighter. Brown ellipses represent sensitivity testing of temporal precision carried out for this study, while the purple ellipses indicate constraints that could be tested using high-temporal resolution forecast data. ....	6
Figure 4.2: Forecast events shown in red, with observed events shown in blue, and merged events overlaid in green. Blue lines (non-merged events) are merged together to form the green line (merged events) if the time period between any 2 blue segments is less than 3 hours, which is the minimum length (in time) of a forecast event. The spacing of tick marks along the x-axis in each of these examples is 3 hours.....	7
Figure 4.3: Distribution of non-merged observation event duration (top), merged observation event duration (middle), and NDFD forecast event duration (bottom) for all terminal regions at the 10% coverage level, and for a 4-h lead time. Note the binning of the forecast event into 3-hourly discrete bins as that is the temporal resolution of the product. Also note the median duration of each distribution.....	8
Figure 4.4: The mean duration of observations (in hours) for all terminal regions (left) with the corresponding number of events for the non-merged and merged observations (right). Notice that more merging occurs in the SE U.S. (right-hand figure). Note in these radial plots that stations are grouped by region, starting at the top (noon) position and then moving clockwise through the North, East, South, and West regions. Magnitude of the measured statistic increases radially outward from the center as indicated by the axis labels located along the line representing East.....	9
Figure 5.1: Example of tallying hits (H), misses (M), and false alarms (FA) for onset (O) and cessation (C) using a $\pm 3$ hour temporal window for matching events. Green line is the merged observation event that is made up of smaller events (blue line). Red line is the NDFD forecast event. ....	10

Figure 5.2: Two sample events illustrating the convention used for computing the timing error. The first example (top) shows that the forecast onset was too early by 4:15 (-4.25 h), and the forecast for the cessation of the event was also too early by 1:30 (-1.50 h). The second example (bottom) shows that the forecast onset was too early by 3:15 (-3.25) and the forecast for the cessation of the event was too late by 7:00 (+7.00 h)..... 11

Figure 5.3: An illustration of the center of mass displacement calculation between the NDFD forecast (red objects) and the CIWS observations (blue objects). Note spatial scale differences in the observation and forecast fields. Displacement error is computed for all temporal hits that were recorded within the bounded, terminal space..... 12

Figure 6.1: Radial plot of the probability of detection at the NDFD *likely-and-above* threshold with the coverage threshold set at 10%, and with varying temporal precision,  $\pm 3$  h (blue),  $\pm 6$  h (red), broad (any) association (green). Note in this radial plot that stations are grouped by region, starting at the top (noon) position and then moving clockwise through the North, East, South, and West regions. The magnitude of the POD increases radially outward from the center as indicated by the axis labels located along the line representing East..... 13

Figure 6.2: The median forecast timing error for onset (green) and cessation (red) for the NDFD *likely-and-above* threshold (left) and the NDFD *scattered-and-above* threshold (right) with coverage threshold set at 10% and with a 4-h forecast lead time. .... 14

Figure 6.3: CSI of forecast onset for the 2-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange)..... 15

Figure 6.4: CSI of forecast onset for the 4-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange)..... 16

Figure 6.5: CSI of forecast onset for the 8-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange)..... 16

Figure 6.6: CSI of forecast cessation for the 8-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange)..... 17

Figure 6.7: Bar charts showing regional breakdowns of POD (blue), CSI (orange), and FAR (green) for convective onset by coverage and NDFD probability forecast threshold for the  $\pm 3$  hour temporal precision constraint for the 2-h lead-time (top) and the 4-h lead-time (bottom). The forecast probability threshold and coverage threshold are indicated at the bottom. National averages for each of the computed scores are indicated by the black dots within the graph. .... 18

Figure 6.8: Bar charts showing regional breakdowns of POD (blue), CSI (orange), and FAR (green) for convective onset by coverage and NDFD probability forecast threshold for the  $\pm 3$  hour temporal precision constraint for the 6-h lead-time (top) and the 8-h lead-time (bottom). The forecast probability threshold and coverage threshold are indicated at the bottom. National averages for each of the computed scores are indicated by the black dots within the graph. .... 19

Figure 6.9: Scatter plot of observed fractional coverage (x-axis) and forecast fractional coverage (y-axis) in the terminal space at onset time, with a box plot inset to represent the coverage distribution. The *likely-and-above* threshold for NDFD was used with a coverage threshold for the forecast and observations set to  $\geq 10\%$ , and using a  $\pm 3$ -h temporal precision to define a hit. .... 20

Figure 6.10: Scatter plot of observed fractional coverage (x-axis) and forecast fractional coverage (y-axis) in the terminal space at cessation time, with a box plot inset to represent the coverage distribution. The *likely-and-above* threshold for NDFD was used with a coverage threshold for the forecast and observation set to  $\geq 10\%$  or above and using a  $\pm 3$ -h temporal precision. .... 20

Figure 6.11: Radial plot of the displacement error by terminal region and by lead time using the  $\pm 3$  hour temporal precision constraint, *likely-and-above* NDFD probability threshold, and a 10% coverage threshold for convective onset. .... 21

Figure 6.12: Box plots of displacement error by terminal airspace (top) and by regional breakdown (bottom) for convective onset at a lead-time of 2 hours for the  $\pm 3$ -hour temporal precision constraint. .... 22

Figure 6.13: Box plots of displacement error for daytime (yellow) and nighttime (dark gray) for convective onset at a lead time of 2 hours for the  $\pm 3$ -hour temporal precision constraint. .... 23

Figure 6.14: Box plots of displacement error by month for convective onset with a lead time of 2 hours for the  $\pm 3$ -hour temporal precision constraint. .... 23

Figure 7.1: FCAA05 (western boundary) and FCAA08 (southern boundary) in green with jetways overlaid in gray. This region defines the en-route region of interest (shown in blue). .... 25

Figure 7.2: Jetway J29 with a 20-nmi buffer shown in red in the domain of interest (blue). .... 26

Figure 7.3: A plot of the convective coverage (top) and the number of jetways impacted (bottom) during 17-26 July 2011. The observations are represented by a bold black line. The 2-h forecast (red), 4-h (blue), 6-h (green), 8-h (cyan) are also shown. .... 27

Figure 7.4: CIWS echo top observation and the 8-h lead-time NDFD forecast convection valid at 2045 UTC on 25 July 2011. 14 jetways were impacted from the observations whereas only 32 jetways were impacted by the forecast. There are 42 total jetways in this domain. .... 28

## List of Tables

Table 1. Adapted draft of onset and cessation requirements for both the near-term and MOC for convection within a 150 nmi diameter of a core airport, and with a forecast probability greater than or equal to <i>likely</i> from the NDFD Thunderstorm Probability field..	2
Table 2: Summary performance measures used in this report: Probability of Detection (POD), Critical Success Index (CSI, threat score), and False Alarm Ratio (FAR), which utilize a tally of hits (H), misses (M), and false alarms (FA). .....	10
Table 3: Summary statistics for all terminal regions for the NDFD forecast product for the 2010 and 2011 convective seasons compared to the MOC requirements at the <i>likely-and-above</i> threshold, a 10% coverage threshold of convective onset or cessation, and a $\pm 3$ -h temporal precision constraint (top) and an idealized $\pm 1$ -h temporal precision constraint (bottom). .....	24

## **1 Introduction**

The Federal Aviation Administration (FAA) and National Weather Service (NWS) established a joint Traffic Flow Management Weather Requirements Working Group (TRWG) in August 2010 to achieve three main objectives: (1) baseline current NWS weather support capabilities; (2) develop firm requirements for near-term services for the Next Generation Air Transportation System (NextGen) and for a Middle Operating Capability (MOC); and (3) develop a plan for implementing solutions to meet each of the established requirements. In support of this effort, the NOAA Earth System Research Laboratory (ESRL) Forecast Impact and Quality Assessment Section (FIQAS) is developing concepts and capabilities to provide performance measures for NWS forecasts that may be used for FAA Traffic Flow Management (TFM) decision making. These measures are used to baseline current forecasting capabilities and are provided prior to implementing the forecast services that will support NextGen MOC.

As part of this initial effort, and following requirements set forth by the TRWG, an in-depth assessment has been conducted of a National Digital Forecast Database (NDFD) convective weather forecast product that is being considered for traffic flow planning in terminal areas. Motivation for this assessment includes developing measures of forecast accuracy relative to the onset and cessation of convective weather events within the terminal area, and a requirement to provide an assessment of aviation-specific concerns centered on the Northeast U.S. This study covers two convective seasons (June–September 2010 and 2011) and measures the forecast lead time to onset and cessation of convection, the displacement of forecasts relative to the observations, and it provides results regarding traffic influences surrounding specific jetways. Skill scores and performance measures are presented in the context of TRWG requirements for terminal space within a 75-nmi radius of Operational Evolution Partnership (OEP) 35 airports ([http://aspmhelp.faa.gov/index.php/OEP\\_35](http://aspmhelp.faa.gov/index.php/OEP_35)) over the CONUS.

## **2 TRWG requirements**

The TRWG has established NextGen performance metrics for a variety of aviation impact variables that are expected to meet both near-term and MOC requirements. Table 1 presents a draft version of requirements for terminal-area convection metrics. The statistics of interest are probability of detection (POD), false alarm ratio (FAR), timing error with respect to time of onset and cessation, and the spatial displacement error at the time of onset and cessation. Of interest in this particular study is the 2, 4, 6, and 8-h lead times to onset and cessation of significant convection in the terminal area, although longer lead times can certainly be considered in future studies.

**Table 1. Adapted draft of onset and cessation requirements for both the near-term and MOC for convection within a 150 nmi diameter of a core airport, and with a forecast probability greater than or equal to *likely* from the NDFD Thunderstorm Probability field.**

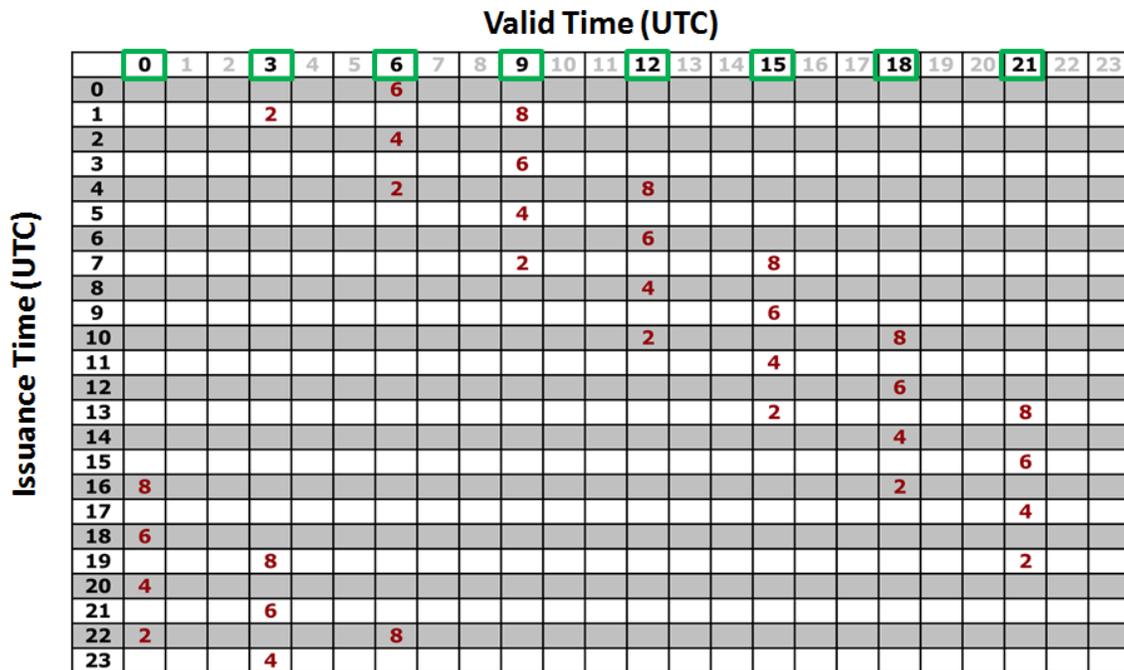
			POD		FAR		Timing		Location	
			Near Term	MOC	Near Term	MOC	Near Term	MOC	Near Term	MOC
<b>Thunderstorms for Core Airports with: Probability <math>\geq</math> LKLY ; Area Diameter <math>\leq</math> 150 NM</b>	<b>Time of Onset</b>	2-h	$\geq 80\%$	$\geq 85\%$	$\leq 20\%$	$\leq 15\%$	$\pm 15$ min	$\pm 10$ min	$\leq 3$ mi	$\leq 3$ mi
		4-h	$\geq 75\%$	$\geq 80\%$	$\leq 25\%$	$\leq 20\%$	$\pm 30$ min	$\pm 20$ min	$\leq 3$ mi	$\leq 3$ mi
		6-h	$\geq 70\%$	$\geq 75\%$	$\leq 35\%$	$\leq 25\%$	$\pm 45$ min	$\pm 30$ min	$\leq 3$ mi	$\leq 3$ mi
		8-h	$\geq 65\%$	$\geq 75\%$	$\leq 40\%$	$\leq 30\%$	$\pm 60$ min	$\pm 45$ min	$\leq 3$ mi	$\leq 3$ mi
			POD		FAR		Timing		Location	
			Near Term	MOC	Near Term	MOC	Near Term	MOC	Near Term	MOC
	<b>Time of Cessation</b>	2-h	$\geq 80\%$	$\geq 85\%$	$\leq 20\%$	$\leq 15\%$	$\pm 15$ min	$\pm 10$ min	$\leq 3$ mi	$\leq 3$ mi
		4-h	$\geq 75\%$	$\geq 80\%$	$\leq 25\%$	$\leq 20\%$	$\pm 30$ min	$\pm 20$ min	$\leq 3$ mi	$\leq 3$ mi
		6-h	$\geq 70\%$	$\geq 75\%$	$\leq 35\%$	$\leq 25\%$	$\pm 45$ min	$\pm 30$ min	$\leq 3$ mi	$\leq 3$ mi
		8-h	$\geq 65\%$	$\geq 75\%$	$\leq 40\%$	$\leq 30\%$	$\pm 60$ min	$\pm 45$ min	$\leq 3$ mi	$\leq 3$ mi

### 3 Data characteristics and constraints

#### 3.1 NDFD forecasts

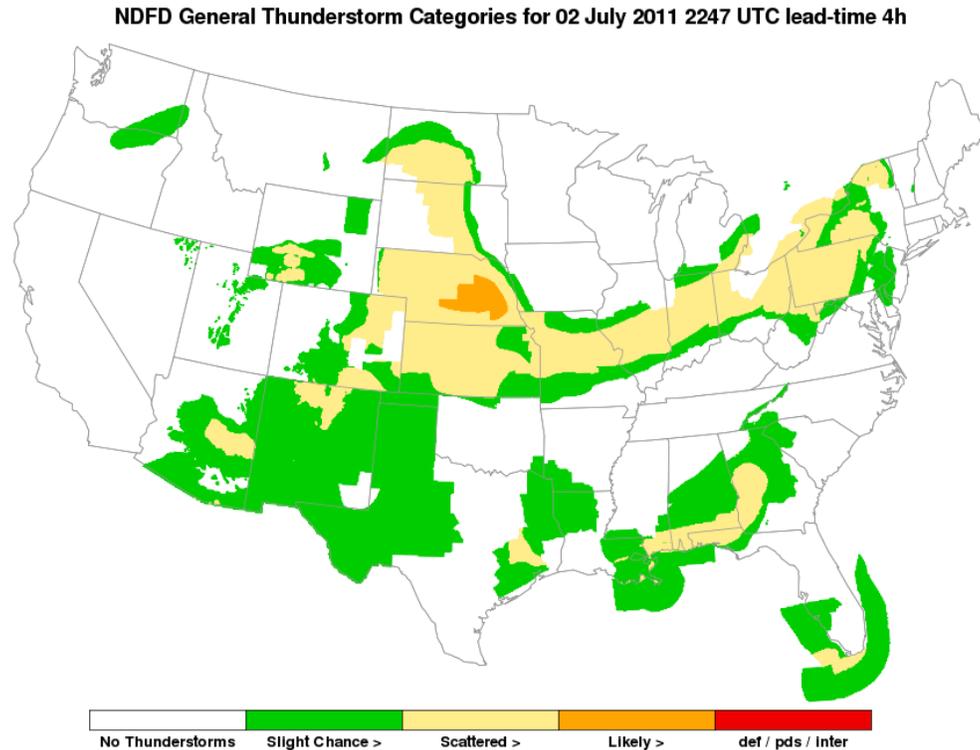
Data used in this study are from the NDFD Forecast Thunderstorm Probability field that is disseminated on a 5-km output grid (Glahn et al. 2003). This field was utilized by first applying a threshold for the *scattered-and-above* level ( $\geq 25\%$  probability) and for *likely-and-above* ( $\geq 55\%$  probability). As outlined by the TRWG, greater consideration is given to data at the higher threshold of likely-and-above.

Careful consideration was given to the availability of operational NDFD data. Although the product is generated hourly, forecasts are only valid at 3-h increments (e.g., at 00, 03, 06,... UTC). Since the lead-times of interest for this study are 2, 4, 6, and 8 hours, the 3-h valid time increment restricts the number of forecasts that can be incorporated into the analysis (see Figure 3.1). It is important to note that although NDFD is valid in 3-h increments, the true temporal resolution of the forecast data is dictated by scheduled Weather Forecast Office (WFO) update cycles, which tend to be every 6 h. This is a critical point that will affect many of the results presented in this paper.



**Figure 3.1:** Available forecast lead times (maroon) at 3-h valid-time increments (green boxes along the top) for the operational NDFD product. Since this study focuses on 2, 4, 6, and 8-h lead times, the number of available forecasts is limited by the 3-h valid time increment.

Although the grid spacing of the NDFD thunderstorm field is 5 km, the effective resolution of the product is much coarser. Figure 3.2 shows an example of a thresholded NDFD thunderstorm probability field. It is important to note that the structure of the NDFD forecast field is typically similar to the scale of a WFO County Warning Area (CWA), rather than a 5-km grid box, although *likely* regions of thunderstorm activity (orange) are on a noticeably finer scale.



**Figure 3.2:** Sample of NDFD general thunderstorm probability categories from 02 July 2011 issued at 2247 UTC, with a 4-h lead-time, and valid on 03 July 2011 at 0247 UTC. Broad yellow regions indicate *scattered* thunderstorm forecasts, while orange indicates *likely* thunderstorms.

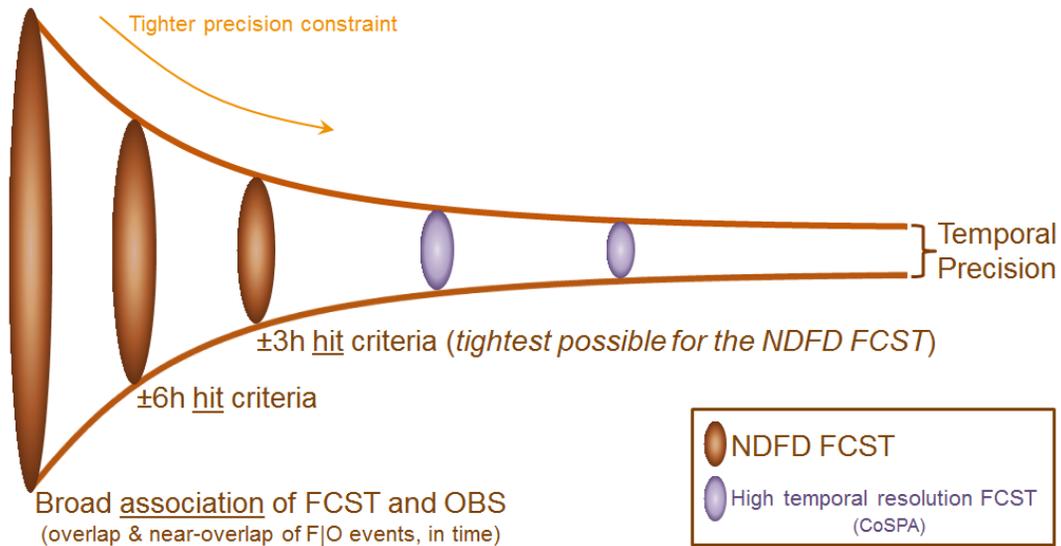
### 3.2 Observations: Corridor Integrated Weather System

The observations used in this study are from the Corridor Integrated Weather System (CIWS) echo top field (Evans et al. 2006). The CIWS analysis field is composed of both NEXRAD and FAA radars that are used to create a CONUS mosaic of both VIL and echo top parameters, with a number of overlays available. The CIWS analysis field is issued every 2.5 min at a spatial resolution of 1 km. The echo top field with a threshold of 30 kft was used in this study instead of the Vertically Integrated Liquid (VIL) field. There are two primary reasons for this decision. First, it was determined that the structure observed when thresholding the echo top field at 30 kft is closer to the structure of forecast objects provided by the NDFD. If VIP-level 3 data were instead utilized, the observation field would be strikingly different from the forecast field, since these observations at this threshold possess much smaller coverage than the likely-and-above forecast field. Second, when using a 30-



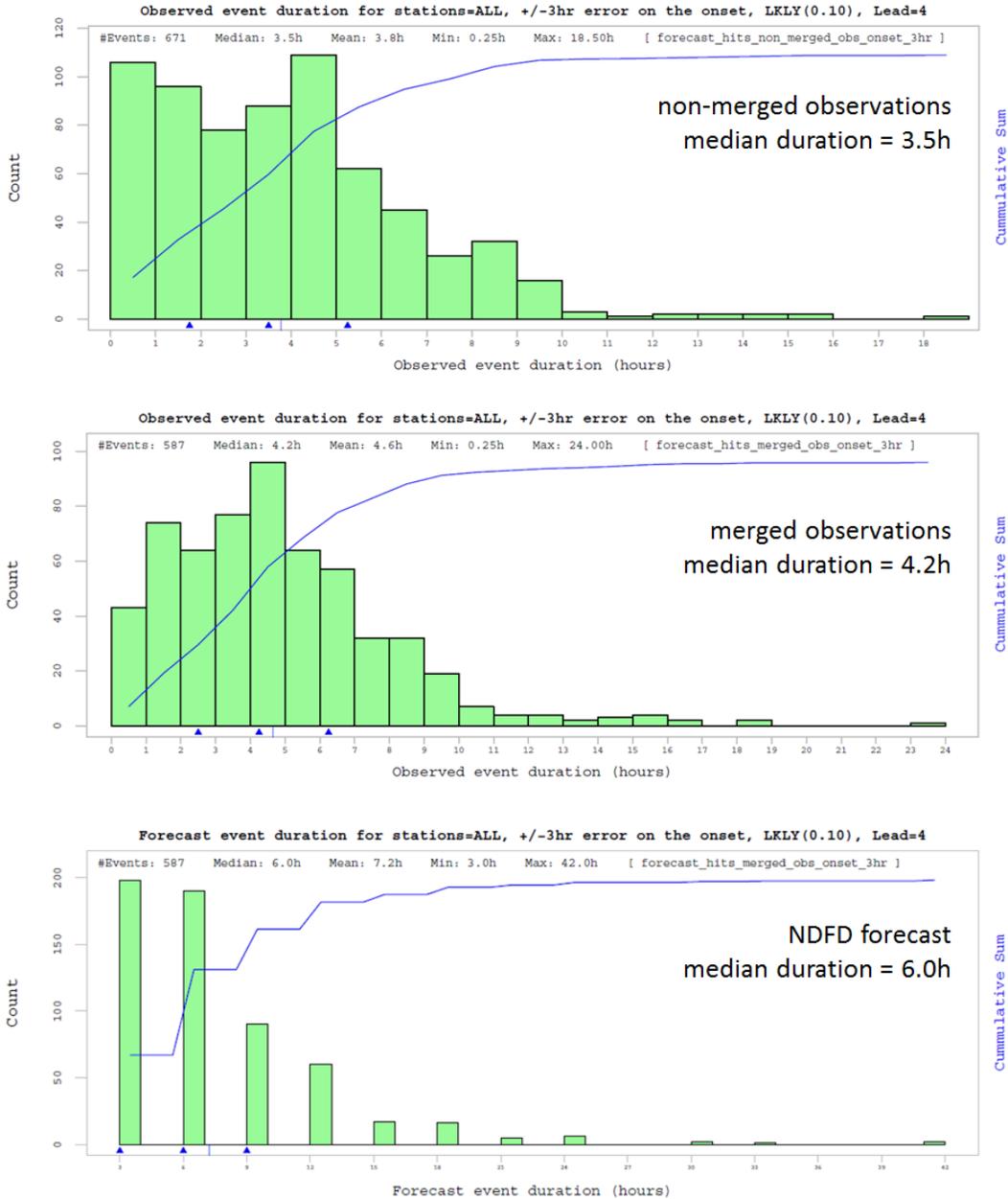
Although there is some sensitivity to what constitutes an impactful terminal weather event as it relates to the orientation of jetways and designated arrival and departure routes, fixed coverage thresholds have been chosen for thunderstorm forecasts and observations. Three coverage levels were part of a sensitivity analysis performed for this study; namely, areal coverage  $>0\%$ ,  $\geq 5\%$ , and  $\geq 10\%$ . The first time a forecast or observation field is measured to have exceeded a particular threshold value it is recorded as the onset time for an event within the terminal space. The preceding drop below that same threshold value is recorded as the cessation time of the event.

Once event information for each terminal is recorded for all hours of the day, forecast and observed events are matched in time so that appropriate metrics can be computed. In practice, the matching method is sensitive to the timing constraints inherent in the NDFD thunderstorm forecasts (3-h valid time increments). It is important to take into consideration the lower temporal resolution of the NDFD forecasts when matching these data to high-temporal resolution CIWS observations. Therefore, when matching these particular forecasts to the observations, the tightest possible timing precision constraint that can be appropriately considered for a successful match is  $\pm 3$  hours. If a higher temporal resolution forecast product were available, this constraint could be correspondingly tightened, and this would facilitate a higher degree of timing accuracy for the onset and cessation of hazardous weather events. Figure 4.1 highlights the issues related to this temporal matching constraint.



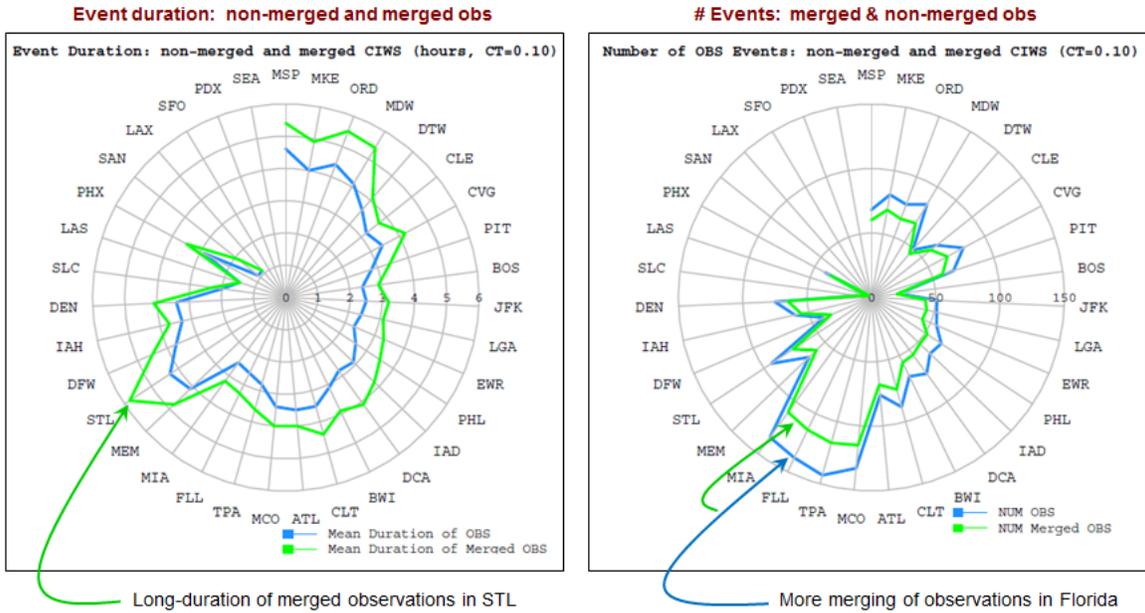
**Figure 4.1:** A broad association of events is illustrated to the left (overlap or near-overlap in time), with an increasing precision constraint applied as the forecast temporal resolution becomes tighter. Brown ellipses represent sensitivity testing of temporal precision carried out for this study, while the purple ellipses indicate constraints that could be tested using high-temporal resolution forecast data.





**Figure 4.3:** Distribution of non-merged observation event duration (top), merged observation event duration (middle), and NDFD forecast event duration (bottom) for all terminal regions at the 10% coverage level, and for a 4-h lead time. Note the binning of the forecast event into 3-hourly discrete bins as that is the temporal resolution of the product. Also note the median duration of each distribution.

When examining the duration of events by station (Figure 4.4), it is apparent that the SE U.S. terminals experience more merging of observed events than the rest of the country. This is most likely due to the nature of air mass storms that primarily occur within this region. Note in these radial plots that the stations are grouped geographically beginning at the 12 o'clock position that represents North, then moving clockwise through stations located in the East, South, and West.



**Figure 4.4:** The mean duration of observations (in hours) for all terminal regions (left) with the corresponding number of events for the non-merged and merged observations (right). Notice that more merging occurs in the SE U.S. (right-hand figure). Note in these radial plots that stations are grouped by region, starting at the top (noon) position and then moving clockwise through the North, East, South, and West regions. Magnitude of the measured statistic increases radially outward from the center as indicated by the axis labels located along the line representing East.

## 5 Scoring

### 5.1 Creation of contingency tables in event space

Once the observed events are appropriately merged-in-time, the forecast events can be matched to the observed events to create a tally of hits, misses and false alarms. Choosing a hit criterion based on the timing window outlined in Figure 4.1 allows for a contingency table to be populated for computing skill scores and performance measures. Examples of hits, misses, and false alarms for events with a  $\pm 3$  hour timing precision are shown in Figure 5.1. The first example at the top gives a tally of a miss on onset ( $M_o$ ), a miss on cessation ( $M_c$ ), a false alarm for forecast onset ( $FA_o$ ) and a false alarm for forecast cessation ( $FA_c$ ). The forecast and observed events were broadly associated in this case because they nearly overlap-- the forecast event was within 3 hours of the start or end of the observed event. The second example results in a hit on the onset ( $H_o$ ) and a hit on the cessation ( $H_c$ ), as the forecast time of onset and cessation are each within  $\pm 3$  hours of the observed times for onset and cessation. Onset for the third event was forecast to occur more than 3 hours too early, but cessation is well forecast, so there is a false alarm on forecast onset ( $FA_o$ ) and a hit registered for the event cessation ( $H_c$ ), and a miss on the observed onset ( $M_o$ ). Finally, although the forecast and observed events nicely overlap as shown in the fourth event (bottom), the timing for onset and cessation is significantly off, and it results in a miss and a false alarm on both onset and

cessation ( $M_o, FA_o, M_c, FA_c$ ). For a collection of events gathered over two full seasons, skill scores are computed that do not rely on the correct negative count, as it is difficult to define a correct negative in the context of the event space. For this particular study, probability of detection (POD), false alarm ratio (FAR), and the critical success index (CSI, threat score) are computed in event space, and the equation for each is given in Table 2.



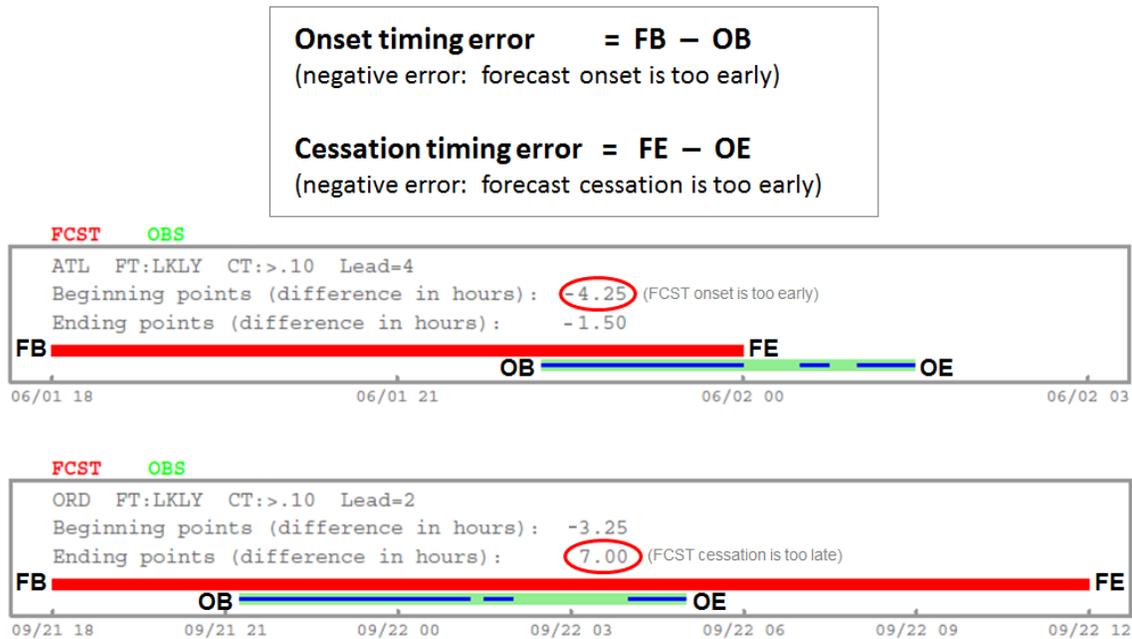
**Figure 5.1:** Example of tallying hits (H), misses (M), and false alarms (FA) for onset (O) and cessation (C) using a  $\pm 3$  hour temporal window for matching events. Green line is the merged observation event that is made up of smaller events (blue line). Red line is the NDFD forecast event.

**Table 2:** Summary performance measures used in this report: Probability of Detection (POD), Critical Success Index (CSI, threat score), and False Alarm Ratio (FAR), which utilize a tally of hits (H), misses (M), and false alarms (FA).

Statistic	Formula	Description
POD	$H / (H + M)$	Proportion of observed events that were correctly detected
CSI	$H / (FA + H + M)$	Proportion of observed and/or forecast events that were correctly detected
FAR	$FA / (FA + H)$	Proportion of forecast events that actually did not occur

## 5.2 Lead-time to onset and cessation

Conventions for computing the timing error for the onset and cessation are established for this study by consistently subtracting the observed time of an event from that of the forecast time of the event. This convention results in a *negative* timing error whenever the forecast is *too early* in predicting either the onset or cessation. Subsequently, a *positive* timing error refers to those situations where the forecast is *too late* on either the onset or cessation of an event. An example of this convention for two events is shown in Figure 5.2.



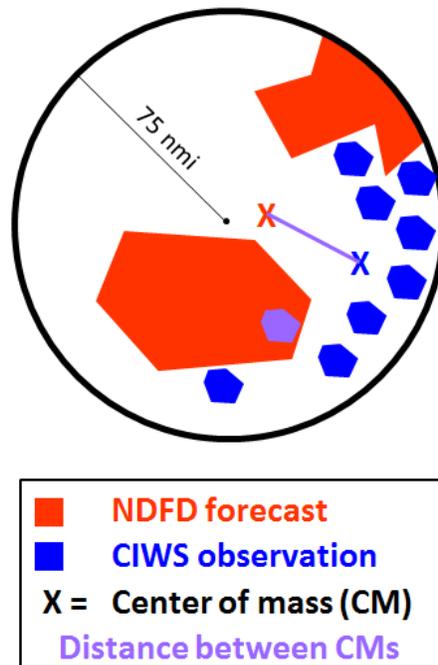
**Figure 5.2:** Two sample events illustrating the convention used for computing the timing error. The first example (top) shows that the forecast onset was too early by 4:15 (-4.25 h), and the forecast for the cessation of the event was also too early by 1:30 (-1.50 h). The second example (bottom) shows that the forecast onset was too early by 3:15 (-3.25) and the forecast for the cessation of the event was too late by 7:00 (+7.00 h).

## 5.3 Spatial displacement error

The final requirement to complete this portion of the study is to establish a suitable convention for calculating the spatial displacement error of forecast convection within the terminal space. Due to differences in the structure of the forecast and observation fields, a simple center of mass calculation was employed for the study. Advanced spatial displacement methods using object techniques may also be applied, but could produce potentially confusing results due to the discrepancy in object size distributions between these particular forecasts and observations (Lack

et al. 2010b). Advanced techniques would also be suitable for use on forecasts of simulated reflectivity that would have inherent spatial similarities to the CIWS observation field.

The first step in assessing spatial errors is to eliminate the temporal displacement that is present in the forecasts for both the onset and cessation of an event. To accomplish this, the forecast onset (cessation) is matched to the observed onset (cessation) in all cases that are defined as a “hit” on the event within the chosen timing precision constraint. Hits for onset and cessation are independent of each other, i.e., a forecast event can register a hit on the onset without necessarily registering a hit on the cessation. An example of a hit is shown in the second illustration of Figure 5.1. For each hit within a terminal space, the center of mass is computed for both the forecast and the observation, and then the displacement is calculated. For hits on event onset, the spatial state of the forecast (observation) at the time of the forecasted (observed) onset of the event is used when computing the center of mass. Likewise, for cessation, the spatial state of the forecast (observation) at the time of the forecasted (observed) cessation of the event is used when computing the center of mass. The center of mass displacement is computed by taking the difference between a weighted mean of the objects identified in the observation field and the forecast field, bounded by the 75-nmi radius defined for the terminal space, as illustrated in Figure 5.3.

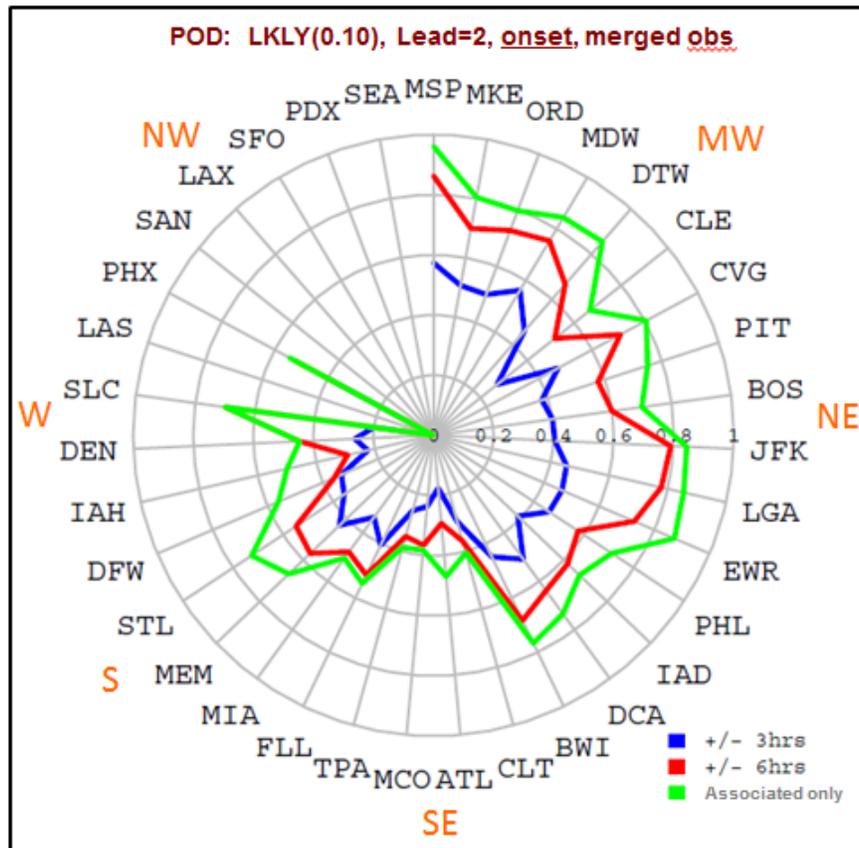


**Figure 5.3:** An illustration of the center of mass displacement calculation between the NDFD forecast (red objects) and the CIWS observations (blue objects). Note spatial scale differences in the observation and forecast fields. Displacement error is computed for all temporal hits that were recorded within the bounded, terminal space.

## 6 Results

### 6.1 Sensitivity to forecast threshold, coverage threshold, and temporal precision

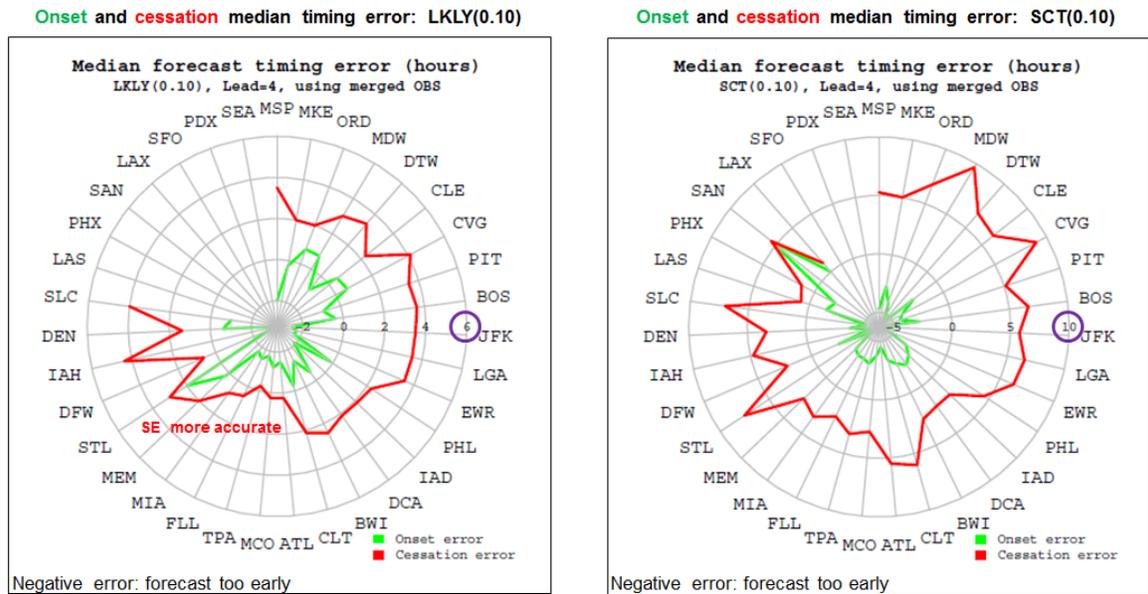
Multiple coverage thresholds were examined for this assessment. Terminal thunderstorm events were defined by varying the percent coverage threshold for both the forecast and observation field within a 75-nmi radius of a terminal. Any coverage  $>0\%$ ,  $\geq 5\%$ , and  $\geq 10\%$  were the thresholds used for determining the onset and cessation of events within the terminal space. In this report, emphasis was placed on the  $\geq 10\%$  threshold, as this is the threshold at which blockage of major departure and arrival routes within a terminal airspace begins to occur. The skill of the NDFD forecast was also evaluated for the scattered-and-above and likely-and-above probability thresholds. In addition to coverage and threshold sensitivities, tests were applied to the temporal precision constraint utilized to define a hit for calculating skill scores. The temporal constraints used for this testing include:  $\pm 3$  hour (to match NDFD valid frequency),  $\pm 6$  hour, and any near-association.



**Figure 6.1:** Radial plot of the probability of detection at the NDFD *likely-and-above* threshold with the coverage threshold set at 10%, and with varying temporal precision,  $\pm 3$  h (blue),  $\pm 6$  h (red), broad (any) association (green). Note in this radial plot that stations are grouped by region, starting at the top (noon) position and then moving clockwise through the North, East, South, and West regions. The magnitude of the POD increases radially outward from the center as indicated by the axis labels located along the line representing East.

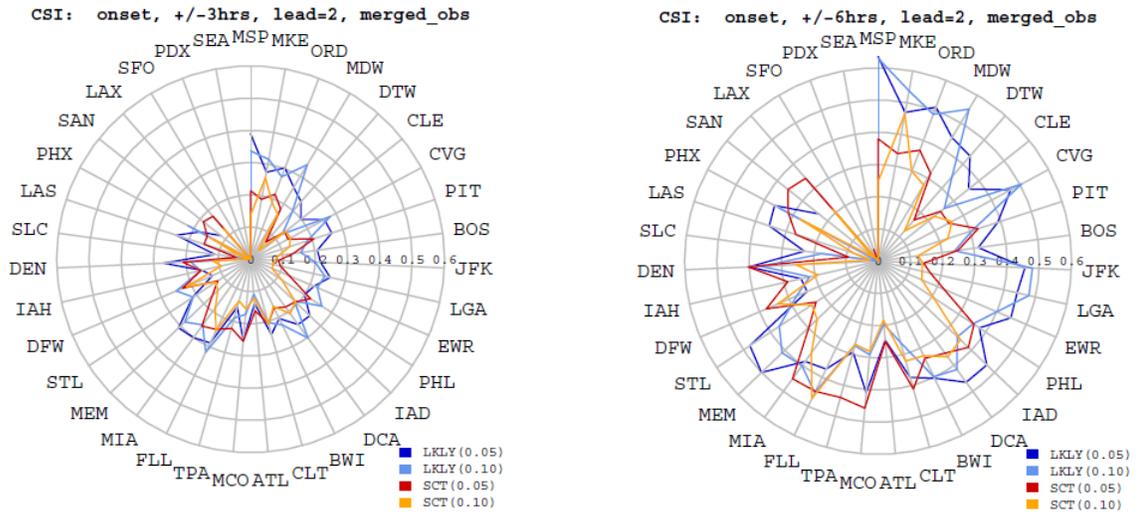
Sensitivity to these choices is summarized in Figure 6.1, which shows the impact on results for the POD by varying the temporal precision for matching forecast and observed events. POD increases as the temporal precision is relaxed; that is, as the timing window for a match is widened. This result is intuitive, since by relaxing the timing constraint, more forecast events are likely to achieve a hit. Note that results in these radial plots are grouped geographically. Sites from the upper Midwest are located in the upper right-hand corner, while Northwest airports are grouped together in the upper left-hand corner. For most of the sites in the NW there are no POD results for any of the temporal precision categories being tested. This is due to the lack of events at the likely-and-above threshold with  $\geq 10\%$  coverage in the NW region. The POD is lowest in the SE U.S., with the highest POD in the NE U.S., which is not surprising due to substantial differences in the mode of convection for these regions. The NE United States typically experiences frontally-forced convection that is more predictable than the scattered, air mass convection common in the SE U.S.

Figure 6.2 shows the sensitivity in the timing error with varying NDFD thunderstorm probability (likely vs. scattered). Interestingly, the magnitude of the median timing errors only changes slightly with varying forecast thunderstorm probability. Of note are the negative timing errors for onset (events forecast to occur too soon) and the positive error for cessation (events forecast to occur too late) for most terminal locations. This result is not surprising, in that most coarse forecasts capture entire events only by predicting the onset to occur too soon and the cessation to occur too late. The cessation has a greater timing error because it is difficult to accurately forecast the dissipation and decay of convection.

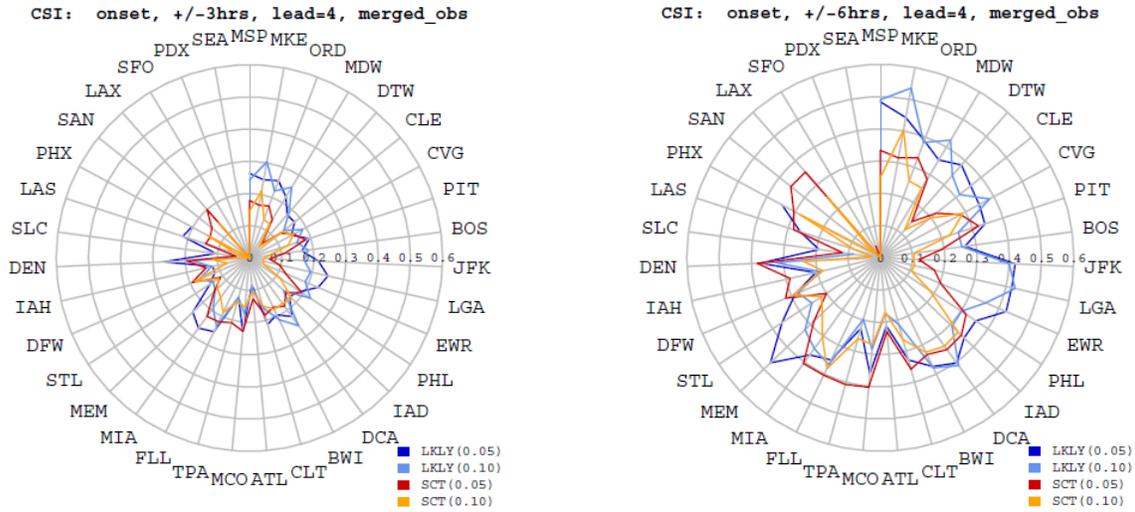


**Figure 6.2:** The median forecast timing error for onset (green) and cessation (red) for the NDFD *likely-and-above* threshold (left) and the NDFD *scattered-and-above* threshold (right) with coverage threshold set at 10% and with a 4-h forecast lead time.

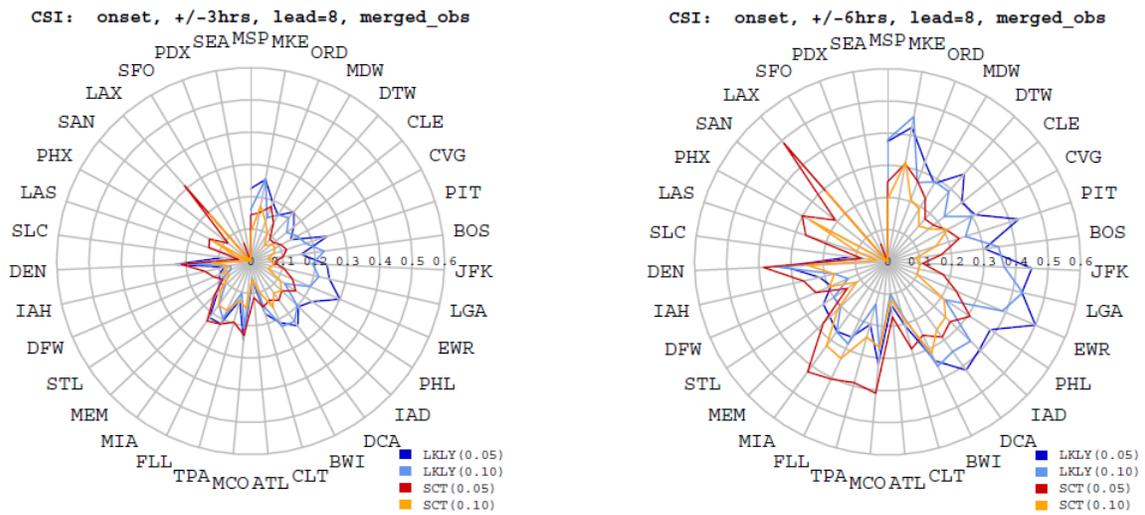
Figure 6.3 - Figure 6.6 illustrate the difference in forecast skill (CSI) with varying forecast lead time and varying temporal precision used to define a hit, also with varying coverage threshold and forecast probability threshold. Of note in these figures is lower skill for the scattered-and-above NDFD probability threshold compared to likely-and-above at both  $\geq 5\%$  and  $\geq 10\%$  coverage. The  $\pm 6$ -h temporal precision constraint scores higher than the  $\pm 3$ -h constraint, by about a factor of 2. Interestingly, there is little sensitivity in scores when stratified by lead-time. The 2-h lead and the 4-h lead CSI scores are virtually the same for all terminal regions with only a slight difference in results from the 8-h lead-time (Figure 6.5). This is due to the coarseness of the NDFD forecast in both time and space. There is typically little difference in the forecast product as time moves forward. Finally, Figure 6.5 and Figure 6.6 compare the skill of the 8-h lead-time for onset and cessation, respectively. This result supports the earlier statement that onset is typically easier to forecast than cessation.



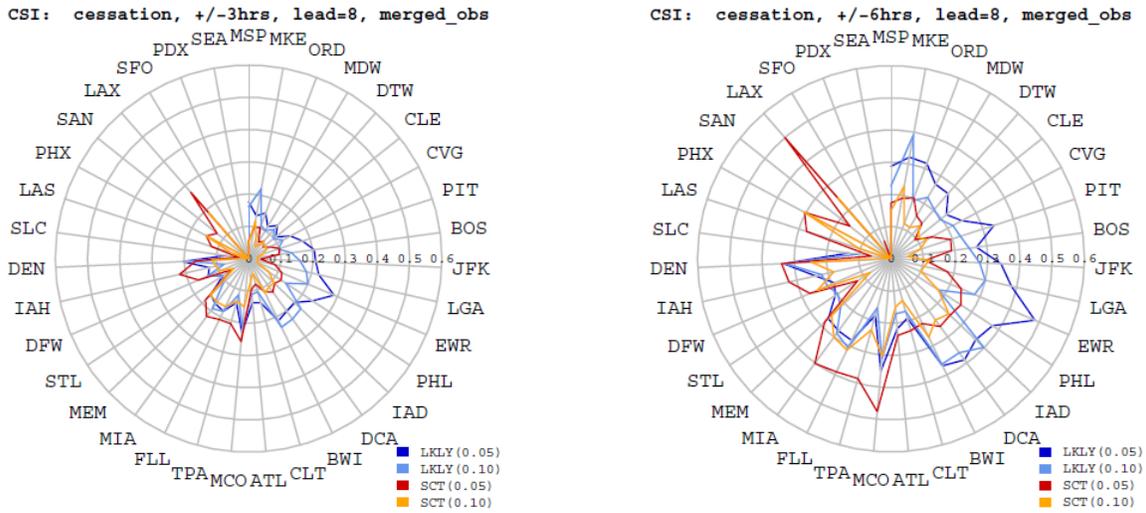
**Figure 6.3:** CSI of forecast onset for the 2-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange).



**Figure 6.4:** CSI of forecast onset for the 4-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange).



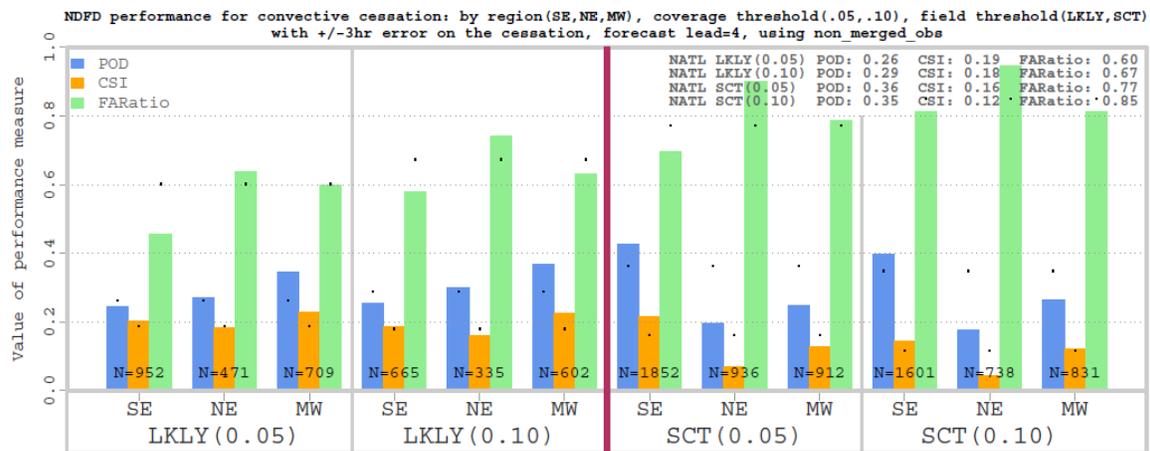
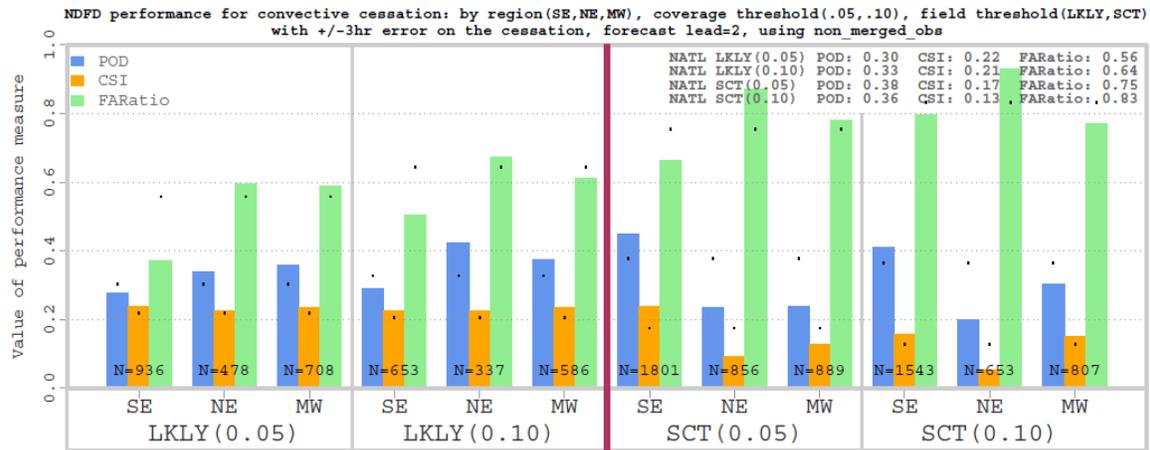
**Figure 6.5:** CSI of forecast onset for the 8-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange).



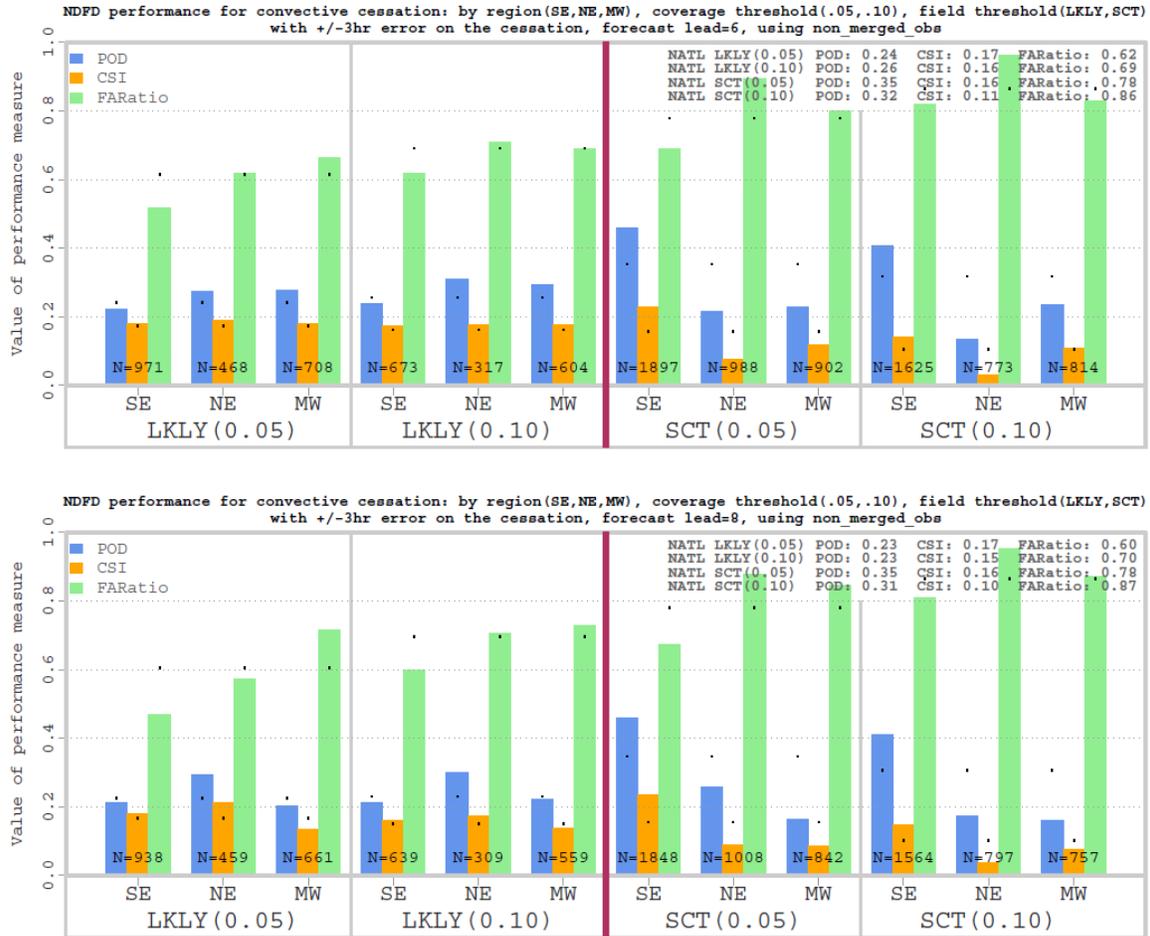
**Figure 6.6:** CSI of forecast cessation for the 8-h lead-time using the temporal precision constraint of  $\pm 3$  h (left) and  $\pm 6$  h (right) for the NDFD *likely-and-above* threshold using 5% coverage (blue) and 10% coverage (cyan) and the NDFD *scattered-and-above* threshold using 5% coverage (red) and 10% coverage (orange).

## 6.2 Regional performance

Figure 6.7 and Figure 6.8 summarize the standard categorical statistics (POD, CSI, and FAR) that were shown (for CSI) in the previous radial plots. The results are stratified by region (SE, NE, and MW) for the  $\pm 3$  h temporal constraint. Again, this temporal constraint was chosen as a primary stratification given the characteristic temporal resolution of NDFD forecast valid times. Included in Figure 6.7 and Figure 6.8 are summary statistics for all the terminals combined by region. These findings mirror those illustrated previously, in that there is little variability in results between lead times, thresholds, and even regions. Black dots on these figures indicate the national average for the respective statistic, region, probability threshold, and coverage threshold.



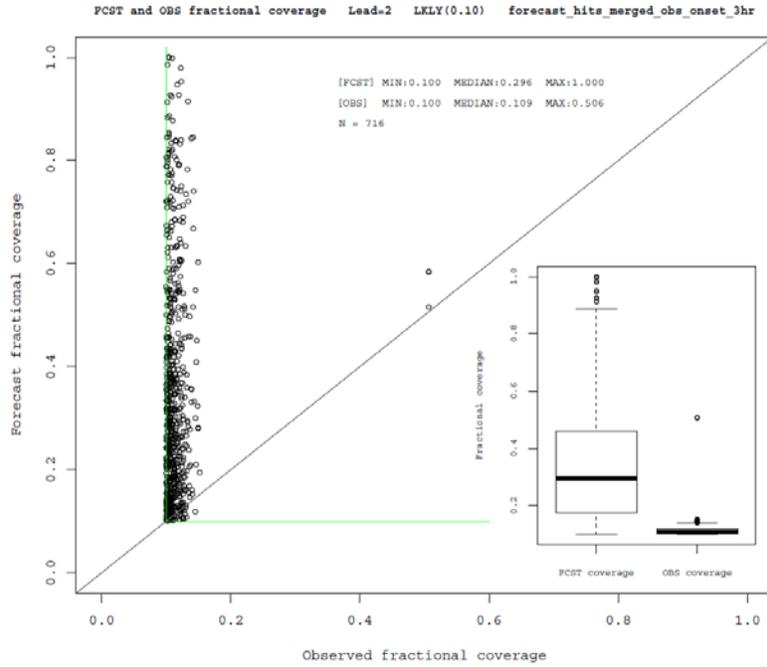
**Figure 6.7:** Bar charts showing regional breakdowns of POD (blue), CSI (orange), and FAR (green) for convective onset by coverage and NDFD probability forecast threshold for the  $\pm 3$  hour temporal precision constraint for the 2-h lead-time (top) and the 4-h lead-time (bottom). The forecast probability threshold and coverage threshold are indicated at the bottom. National averages for each of the computed scores are indicated by the black dots within the graph.



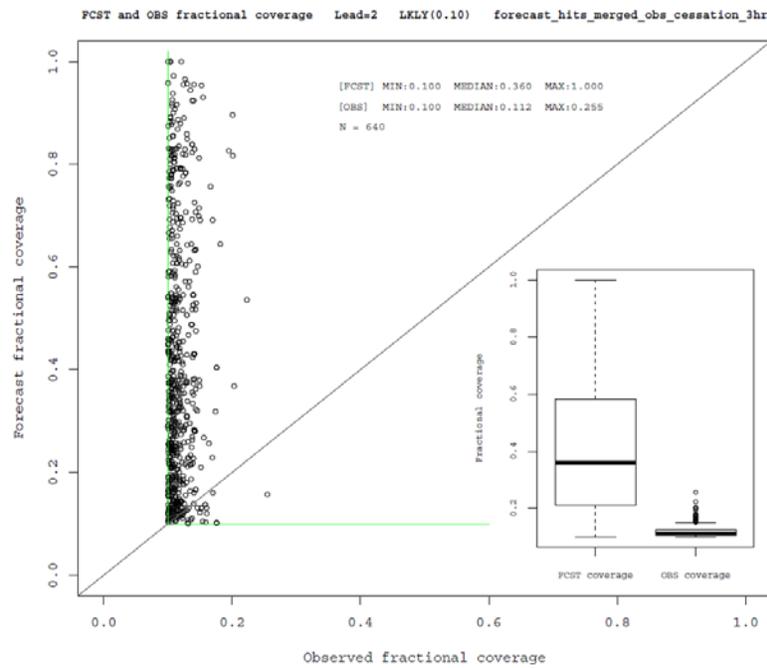
**Figure 6.8:** Bar charts showing regional breakdowns of POD (blue), CSI (orange), and FAR (green) for convective onset by coverage and NDFD probability forecast threshold for the  $\pm 3$  hour temporal precision constraint for the 6-h lead-time (top) and the 8-h lead-time (bottom). The forecast probability threshold and coverage threshold are indicated at the bottom. National averages for each of the computed scores are indicated by the black dots within the graph.

### 6.3 Displacement results

Just as sensitivity analyses were conducted for the temporal aspects of this work, similar studies were performed on the sensitivity of errors for the spatial displacement. Figure 6.9 presents a scatter plot of the observed coverage and the forecast coverage at the time of onset for the  $\pm 3$ -hour temporal precision constraint used to define a hit, while Figure 6.10 presents the same results for event cessation. The first plot provides a comparison of the forecast coverage distribution to the observed coverage at the onset time. In these results, the forecast and observation coverage thresholds were set to 10%. The observation coverage for onset is near 10% in most cases, while the forecast coverage varies from 10% to 100% at the event onset time. Variation in forecast coverage from 0% to 100% in a single time step for a given terminal indicates that there is convective initiation occurring in the spatial and temporal event space. This is a characteristic of the broad temporal and spatial scales found inherent in the NDFD forecast product.

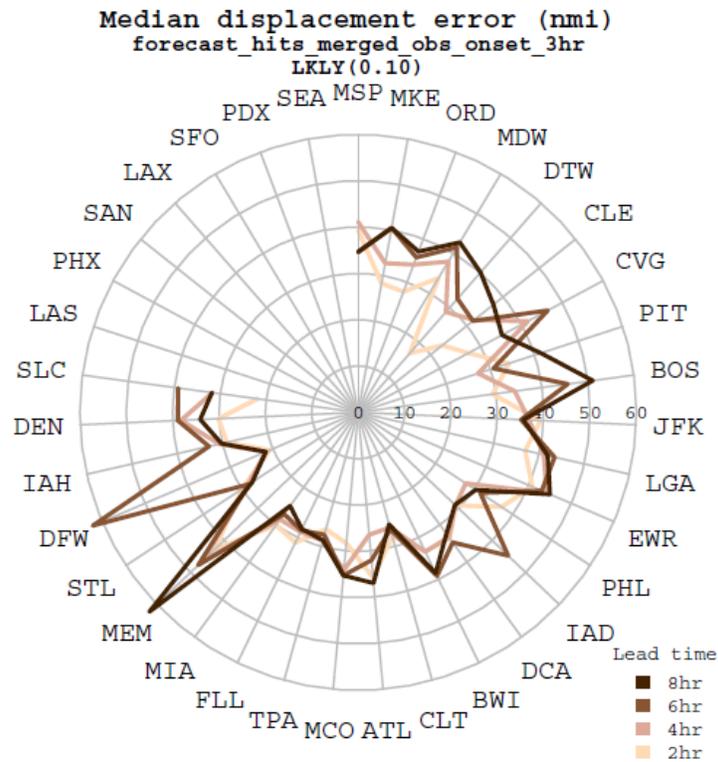


**Figure 6.9:** Scatter plot of observed fractional coverage (x-axis) and forecast fractional coverage (y-axis) in the terminal space at onset time, with a box plot inset to represent the coverage distribution. The *likely-and-above* threshold for NDFD was used with a coverage threshold for the forecast and observations set to  $\geq 10\%$ , and using a  $\pm 3$ -h temporal precision to define a hit.



**Figure 6.10:** Scatter plot of observed fractional coverage (x-axis) and forecast fractional coverage (y-axis) in the terminal space at cessation time, with a box plot inset to represent the coverage distribution. The *likely-and-above* threshold for NDFD was used with a coverage threshold for the forecast and observation set to  $\geq 10\%$  or above and using a  $\pm 3$ -h temporal precision.

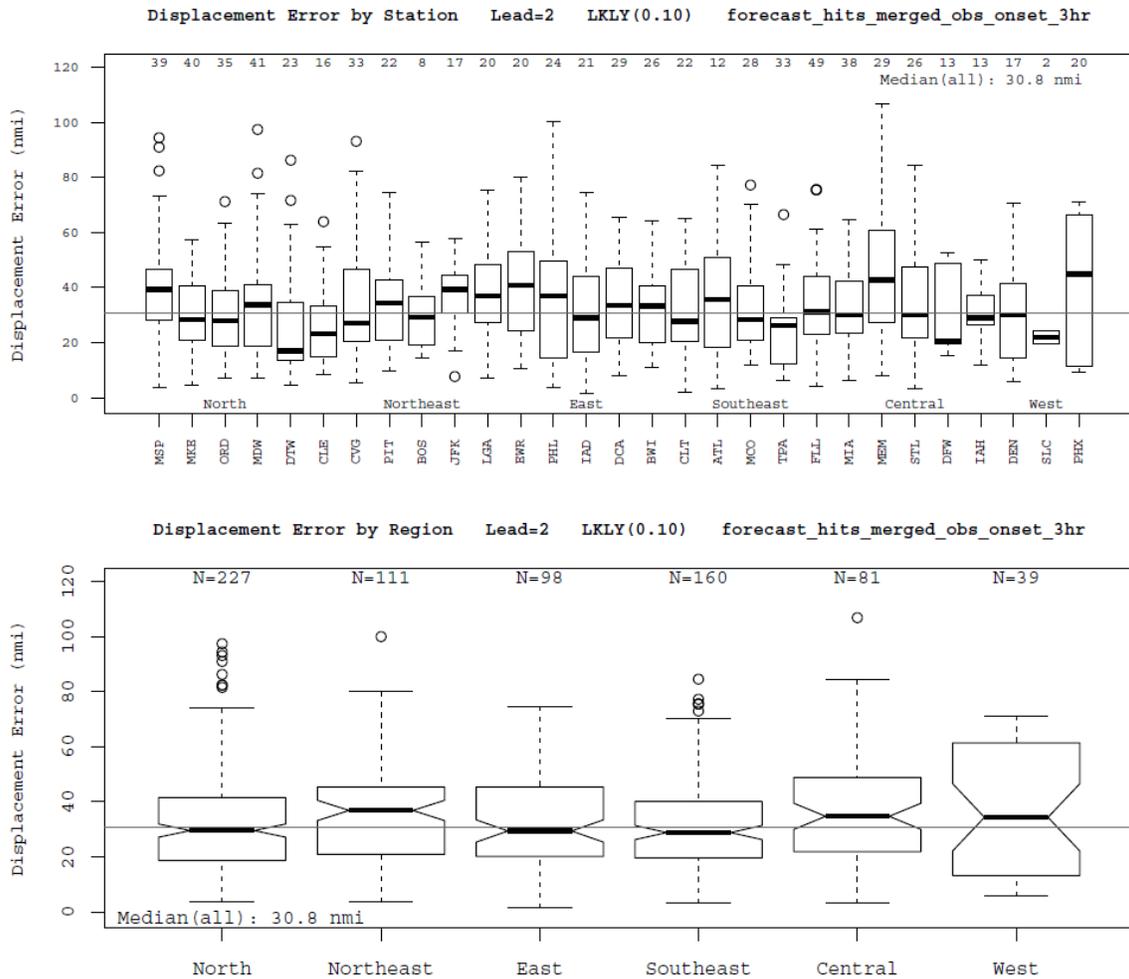
Figure 6.11 shows the median displacement error (in nmi) for each of the terminal spaces, with darker line colors representing increasing forecast lead time. Similar to the temporal displacement and event skill results, there is not much variation in skill by forecast lead time. Again, this is due to the low temporal and spatial resolution of NDFD forecasts. Errors exceeding 30 nmi are typical in most terminal spaces. Displacement errors tend to be lower in the SE U.S. where air mass convection dominates. This result is due to broad areas of forecasts over the terminal area capturing widely scattered convection in a given region. The center of mass for widely scattered convection is likely to be located near the center of the terminal region in this framework, while a broad forecast that covers most of the domain will also have a center of mass located near the center of a terminal area. This results in stronger agreement.



**Figure 6.11:** Radial plot of the displacement error by terminal region and by lead time using the  $\pm 3$  hour temporal precision constraint, *likely-and-above* NDFD probability threshold, and a 10% coverage threshold for convective onset.

Figure 6.12 - Figure 6.14 present additional results on the sensitivity of displacement errors as were summarized in the previous radial plot. When examining forecast performance by sub-region (Figure 6.12, bottom), it is apparent that the Southeast performs better than the Northeast in terms of errors in spatial displacement. Again, this is due to the nature of convection found in these particular regions. The Northeast requires a more accurate forecast than the coarse NDFD can provide, as convection in that region is dominated by transient fronts, whereas convection in the Southeast tends to be broadly scattered and diurnally forced. Additional information included in Figure 6.12 is the count (number) of

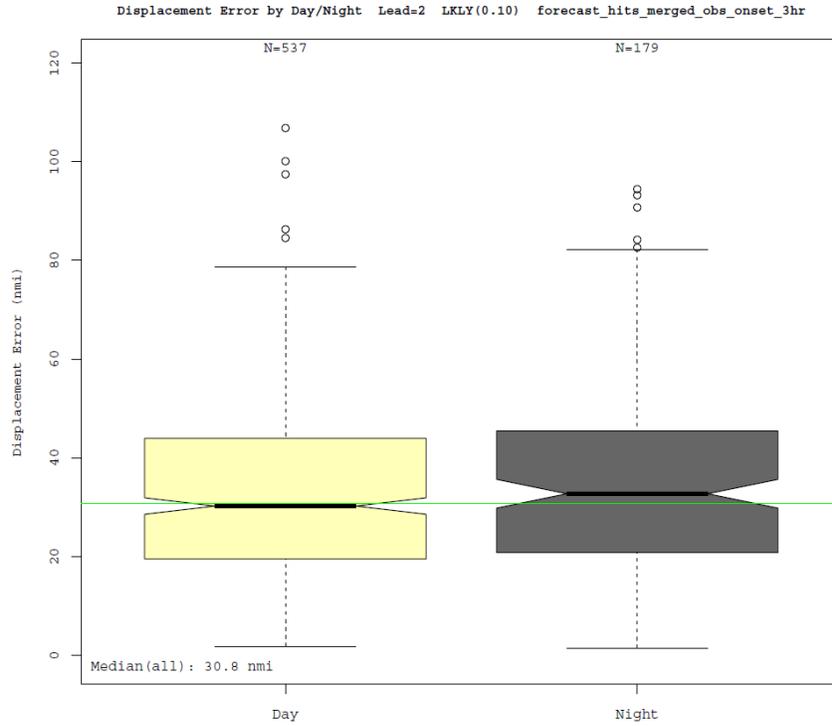
events by station and region. These counts are useful when comparing skill in different regions of the country; for example, the Western U.S. has the fewest number of thunderstorm events and Las Vegas (LAS, zero) has the fewest events by individual terminal.



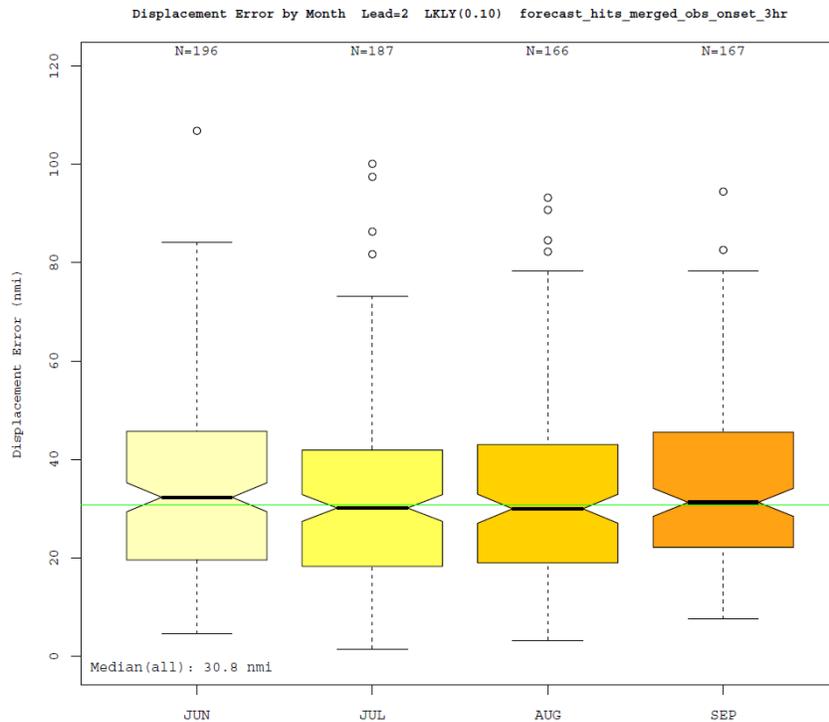
**Figure 6.12:** Box plots of displacement error by terminal airspace (top) and by regional breakdown (bottom) for convective onset at a lead-time of 2 hours for the  $\pm 3$ -hour temporal precision constraint.

#### 6.4 Displacement results by time of day or season

Figure 6.13 shows event displacement errors broken down by time of day. Nighttime skill, in terms of displacement, is slightly worse than daytime skill, due to inherent difficulties in forecasting nighttime convective cessation. This is especially true for long-lived MCSs that persist late into the evening hours. Although displacement errors are slightly larger during nighttime, the sample size is also much lower than it is for daytime, and the differences are not statistically significant (note overlap in the notched box plots). Figure 6.14 provides a breakdown of displacement errors by month during the 2010 and 2011 convective seasons. It is interesting to note that the highest errors occur in June, when much of the convection still has spring-related forcing through the northern parts of the country.



**Figure 6.13:** Box plots of displacement error for daytime (yellow) and nighttime (dark gray) for convective onset at a lead time of 2 hours for the  $\pm 3$ -hour temporal precision constraint.



**Figure 6.14:** Box plots of displacement error by month for convective onset with a lead time of 2 hours for the  $\pm 3$ -hour temporal precision constraint.

## 6.5 Skill summary with MOC requirements

Skill measured for the NDFD thunderstorm probability field is summarized in Table 3. The top table is the skill of the NDFD product in terms of the product itself; that is, merged observations are used to simulate events at the coarseness of the raw forecasts, and the  $\pm 3$  hour temporal constraint is used to mirror the product's temporal resolution. In addition, a 10% coverage threshold on the forecast and observed regions of convection is used to identify onset and cessation times, and to restrict events to those that have potential impact on air traffic.

In terms of the categorical statistics (POD and FAR) calculated within the event space, the NDFD thunderstorm product misses the mark by a considerable degree for the MOC timeframe and even for the near-term timeframe (not shown). In addition, the timing error for onset and cessation for all lead-times are well over one hour, and the displacement errors are greater than 30 nmi. These summary measures once again illustrate that even with the coarse NDFD product the forecasting of cessation is a more difficult problem than forecasting onset in a terminal region.

**Table 3:** Summary statistics for all terminal regions for the NDFD forecast product for the 2010 and 2011 convective seasons compared to the MOC requirements at the *likely-and-above* threshold, a 10% coverage threshold of convective onset or cessation, and a  $\pm 3$ -h temporal precision constraint (top) and an idealized  $\pm 1$ -h temporal precision constraint (bottom).

**Skill with 10% coverage threshold and  $\pm 3$ hr temporal precision; Timing and Location Error ( $\pm 3$ hr)**

			POD		FARatio		Timing (min)		Location (nmi)	
			NDFD	MOC	NDFD	MOC	NDFD	MOC	NDFD	MOC
Thunderstorms for Core Airports with: Probability $\geq$ LKLY Area Diameter $\leq$ 150 nmi	Time of Onset	2-h	0.38	$\geq .85$	0.64	$\leq .15$	91.45	$\pm 10$ min	33.83	$\leq 3$ nmi
		4-h	0.32	$\geq .80$	0.69	$\leq .20$	89.64	$\pm 20$ min	35.33	$\leq 3$ nmi
		6-h	0.29	$\geq .75$	0.70	$\leq .25$	89.69	$\pm 30$ min	36.77	$\leq 3$ nmi
		8-h	0.26	$\geq .75$	0.70	$\leq .30$	87.41	$\pm 45$ min	37.21	$\leq 3$ nmi
			POD		FARatio		Timing (min)		Location (nmi)	
			NDFD	MOC	NDFD	MOC	NDFD	MOC	NDFD	MOC
	Time of Cessation	2-h	0.34	$\geq .85$	0.68	$\leq .15$	93.40	$\pm 10$ min	42.64	$\leq 3$ nmi
		4-h	0.30	$\geq .80$	0.71	$\leq .20$	97.06	$\pm 20$ min	42.05	$\leq 3$ nmi
		6-h	0.26	$\geq .75$	0.73	$\leq .25$	91.29	$\pm 30$ min	42.43	$\leq 3$ nmi
		8-h	0.24	$\geq .75$	0.72	$\leq .30$	83.70	$\pm 45$ min	43.40	$\leq 3$ nmi

**Skill with 10% coverage threshold and  $\pm 1$ hr temporal precision; Timing and Location error (any association)**

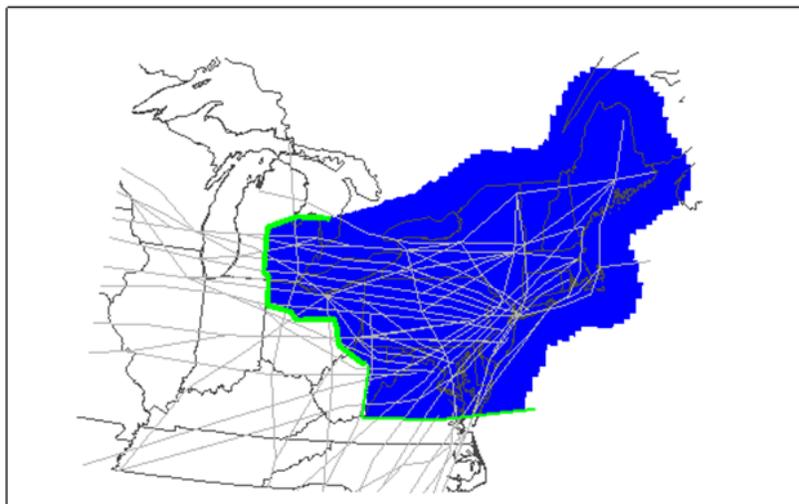
			POD		FARatio		Timing (min)		Location (nmi)	
			NDFD	MOC	NDFD	MOC	NDFD	MOC	NDFD	MOC
Thunderstorms for Core Airports with: Probability $\geq$ LKLY Area Diameter $\leq$ 150 nmi	Time of Onset	2-h	0.14	$\geq .85$	0.87	$\leq .15$	232.93	$\pm 10$ min	37.78	$\leq 3$ nmi
		4-h	0.12	$\geq .80$	0.88	$\leq .20$	262.91	$\pm 20$ min	38.90	$\leq 3$ nmi
		6-h	0.11	$\geq .75$	0.89	$\leq .25$	267.97	$\pm 30$ min	40.16	$\leq 3$ nmi
		8-h	0.10	$\geq .75$	0.88	$\leq .30$	272.84	$\pm 45$ min	40.15	$\leq 3$ nmi
			POD		FARatio		Timing (min)		Location (nmi)	
			NDFD	MOC	NDFD	MOC	NDFD	MOC	NDFD	MOC
	Time of Cessation	2-h	0.13	$\geq .85$	0.88	$\leq .15$	264.71	$\pm 10$ min	42.41	$\leq 3$ nmi
		4-h	0.11	$\geq .80$	0.90	$\leq .20$	288.35	$\pm 20$ min	42.67	$\leq 3$ nmi
		6-h	0.11	$\geq .75$	0.89	$\leq .25$	306.56	$\pm 30$ min	43.04	$\leq 3$ nmi
		8-h	0.12	$\geq .75$	0.87	$\leq .30$	308.83	$\pm 45$ min	44.47	$\leq 3$ nmi

The bottom portion of Table 3 presents skill of the NDFD forecast product. It is treated with a very small temporal precision constraint of  $\pm 1$  hour, as if the forecast were capable of providing higher-resolution information, such as is provided by simulated reflectivity from a state-of-the-art, mesoscale forecast model like the operational Rapid Refresh Model (RAP, Benjamin et al. 2006). When utilizing the  $\pm 1$  hour temporal precision constraint, the categorical skill (POD and FAR) are considerably lower, to the point that the forecast would be considered rather difficult to use. The temporal errors shown in this case are from using the criteria of *any associated event*. This means that a hit is recorded if a forecast event simply overlaps an observed event at any time, or is within 3 hours of the observed event. The mean timing errors are well over 4 hours in these cases, which makes the forecast seemingly of little value to planning for the onset and cessation at critical lead-time horizons within the terminal airspace.

## 7 En-route considerations

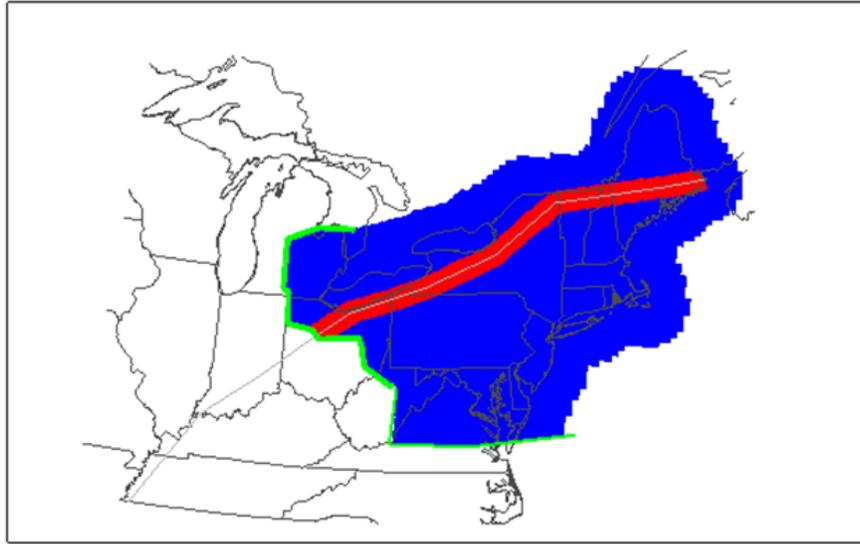
In addition to the terminal airspace requirements summarized by the lead-time to the occurrence or cessation of hazardous weather, an en-route event-based framework was requested for this report. The en-route consideration involves the identification of jetways found in the AFP-bounded regions in the NE U.S. AFPs FCAA05 and FCAA08 effectively frame the bounded region of interest for this portion of the study (see Figure 7.1). When identifying events, it was deemed important to first ascertain the impact on individual jetways by constructing a 20-nmi boundary around jetways within the domain. Figure 7.2 shows an example of a buffer surrounding one sample jetway.

Domain in blue, FCAA05 and FCAA08 in green with Jetways in gray



**Figure 7.1:** FCAA05 (western boundary) and FCAA08 (southern boundary) in green with jetways overlaid in gray. This region defines the en-route region of interest (shown in blue).

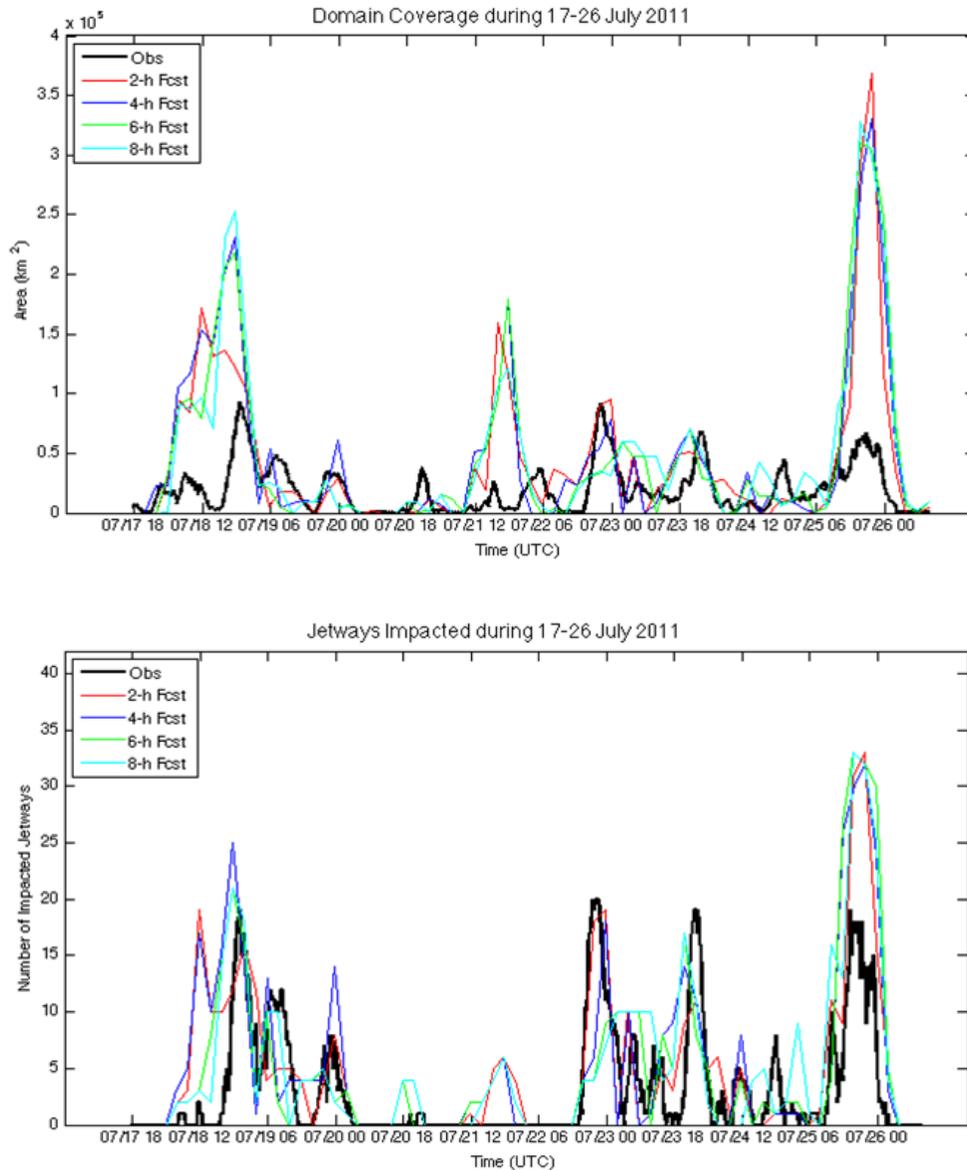
J29 with 20-nmi buffer in red applied



**Figure 7.2:** Jetway J29 with a 20-nmi buffer shown in red in the domain of interest (blue).

Determining impact to the NE region utilizing this en-route framework requires a summary of the total impact on numerous jetways that reside within the domain. There are various methodologies that can be employed to determine impact along a jetway. The simplest is to define a coverage threshold for convection along the jetway itself. This is problematic in that a jetway is likely to be considered impacted if it is severed anywhere along its path. In the coverage framework we have adopted, a jetway may be minimally covered by convection and be completely severed, likewise, it may experience significant coverage and still be traversable.

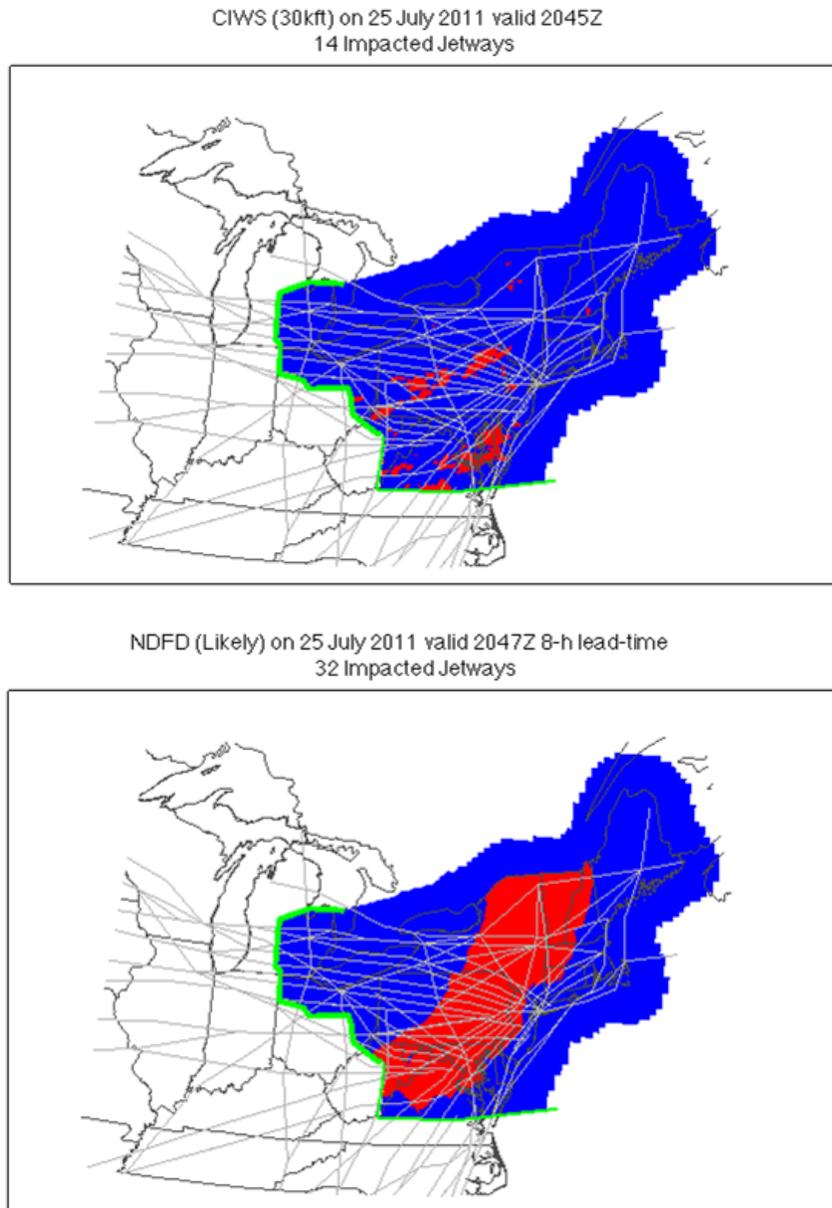
A Mincut technique to determine whether a jetway is severed is the most appropriate method to employ to gauge the impact along a specific route; however, it was computationally expensive to employ this technique for this portion of the study. In lieu of this approach, a Euclidean distance method was utilized. This approach takes into account the coverage and orientation of convection along the buffered jetway path. This approach is briefly explained in Layne and Lack 2010, along with the preferred implementation of the Mincut Bottleneck technique. Once the Euclidean distance approach is applied to each jetway in the domain of interest, a summary of en-route impact in the NE domain is computed. In this early implementation, a straightforward summary statistic is the number of jetways impacted by both observed (CIWS echo tops  $\geq 30$ kft) and forecast (NDFD likely-or-greater thunderstorm probability) convection within the AFP-bounded region of interest.



**Figure 7.3:** A plot of the convective coverage (top) and the number of jetways impacted (bottom) during 17-26 July 2011. The observations are represented by a bold black line. The 2-h forecast (red), 4-h (blue), 6-h (green), 8-h (cyan) are also shown.

Figure 7.3 shows the difference in considering just convective coverage in the domain of interest versus actually considering the number of jetways impacted within the domain. Although coverage can be similar, the number of impacted jetways can vary considerably. The domain of interest for this study includes 42 jetways that are contained within or intersect FCAA05 or FCAA08 from NE terminals. A small amount of convective coverage near the NY terminal area can impact several high-density jetways, while similar coverage amounts over large expanses of the western U.S. would result in little constraint to air traffic flow. Due to the density of terminals in the NY region, it is difficult to make appropriate

distinctions between the en-route traffic impact and the impact on arrivals and departures. Figure 7.4 illustrates the difference between observed and forecast convection in the domain of interest. When comparing high-resolution observations to coarse-resolution forecasts, the effectiveness of pinpointing specific regions of potential impact is limited. This is particularly evident in the NE U.S. where the density of jetways is high.



**Figure 7.4:** CIWS echo top observation and the 8-h lead-time NDFD forecast convection valid at 2045 UTC on 25 July 2011. 14 jetways were impacted from the observations whereas only 32 jetways were impacted by the forecast. There are 42 total jetways in this domain.

## 8 Conclusions

For the 2010 and 2011 convective seasons, performance measures have been computed and a sensitivity analysis has been performed leading to these primary findings:

- NDFD convective forecasts in their current form are unable to resolve the finer temporal and spatial scales that are required for forecasts of onset and cessation at the terminal/tactical scale:
  - The scheduled forecast update cycle at individual Weather Forecast Offices (WFOs) limits NDFD's effective use when considering impact along high-traffic arrival and departure routes and for en route traffic within congested corridors.
  - Event coverage over the terminal space is far greater for NDFD forecasts than it is for observed convection. It was determined that this is a characteristic of forecast feature resolution, not the NDFD grid spacing.
- NDFD forecast characteristics do not capture attributes of convection such as growth, decay, and advection:
  - There is no significant movement of forecast convective systems into, out of, or through terminal areas.
  - Forecast objects are not well correlated with observations in either space or time.
- Onset of significant convective weather is generally forecast to occur too early while cessation is typically forecast to occur too late.
- NDFD temporal and spatial accuracy is better for onset than it is for cessation.
- Skill is quite similar for the scattered-and-above and the likely-and-above convective thresholds.
- For  $\pm 3$  hour temporal precision, there is a decrease in skill with increasing lead time for all regions, but this is not as noticeable for the extended lead times (out to 8 hours) due to the persistent nature of coarse-resolution NDFD forecasts.
- Displacement errors are greater at the time of cessation than at the time of onset, and the median displacement error for each of these exceeds 30 nmi.
- Skill varies by region. Displacement errors are higher in the northeast than in the southeast according to the predominate form of convection in each region.

In summary, the NDFD thunderstorm probability field is limited by its effective temporal and spatial resolution. Although the forecast is issued hourly with 3-hrly valid times, it is limited by the infrequent updates provided by individual WFOs.

The NDFD grid does have a horizontal grid spacing of 5 km, but the actual spatial resolution of the forecast can be determined by examining the features that are forecast by each WFO. The effective spatial resolution is more coarse than the horizontal grid spacing might suggest, and is more in-line with the resolution of a CWA, or even the outer boundary of a WFO.

## 9 Future work

This work was specifically aimed at determining the performance of NDFD thunderstorm forecasts at key decision points for implementing Traffic Management Initiatives. Through this effort, FIQAS has developed an objective methodology for assessing forecast skill in terms of forecast lead time and spatial displacement. A follow-on assessment of other high-resolution NWS forecast products using this established methodology will be performed to fully document current forecast capabilities relative to requirements established by the TRWG. Products to be included in the assessment are the NDFD forecast at 2.5 km resolution, the RAP, the NAM, and the LAMP thunderstorm forecast. Work will include necessary adaptations of the lead-time and displacement assessment methodology to these forecast products, extension of the jetway methodology, analysis of the results, and a presentation of findings to the TRWG. An automated prototype application with online access to current results will be developed and deployed in the spring of 2013. Subsequent efforts will include backfilling the application with historical data from previous seasons to expand the baselining effort.

## Acknowledgments

This research was funded by the National Weather Service Aviation Services Branch in support of requirements established by the Traffic Flow Management Requirements Working Group.

## References

- Benjamin, S. G., D. Devenyi, T. Smirnova, S. Weygandt, J. M. Brown, S. Peckham, K. Brundage, T. L. Smith, G. Grell, and T. Schlatter, 2006: From the 13-km RUC to the Rapid Refresh. *12th Conf. on Aviation, Range, and Aerospace Meteorology*, Amer. Meteor. Soc. <https://ams.confex.com/ams/pdfpapers/104851.pdf>.
- Evans, J.E. and E. Ducot, 2006: Corridor Integrated Weather System. *Lincoln Laboratory Journal*, **16**, 59-80.
- Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database at the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–201.

Lack, S.A., G. L. Limpert, and N. I. Fox, 2010b: An object-oriented multi-scale verification scheme. *Weather and Forecasting*, **25**, 79-92.

Layne, G.J. and S.A. Lack, 2010: Methods for estimating air traffic capacity reductions due to convective weather for verification. *14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM)*. Atlanta, GA, Amer. Met. Soc.