



Lead Time and Displacement Error for Thunderstorm Forecasts in Terminal and Jetway Domains

Prepared by

Quality Assessment Product Development Team
NOAA/ESRL/GSD/Forecast Impact and Quality Assessment Section (FIQAS)

Authors:

Matt Wandishin², Geary Layne², Brian Etherton¹, Melissa Petty³, Jennifer Mahoney¹

1 July 2013

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:

Missy Petty
NOAA/ESRL/GSD, 325 Broadway, Boulder, CO 80305
melissa.a.petty@noaa.gov

EXECUTIVE SUMMARY

This report follows upon a previous quality assessment of the National Digital Forecast Database (NDFD) Thunderstorm Probability field. As before, the evaluation focuses on the onset and cessation of significant thunderstorm activity around 29 major U.S. airports and the ability of forecasts to accurately place thunderstorms in space and time. Extensions of the first assessment include addition of other forecast products (i.e., the Localized Aviation MOS Program (LAMP), the WRF Rapid Refresh (RAP) model, and the High Resolution Rapid Refresh (HRRR)), as well as updates to the spatial displacement calculations and the consideration of en-route thunderstorm forecasts. In addition, a second version of the NDFD thunderstorm forecast was examined, using a lower probability threshold. The evaluation was performed on forecasts for the 2012 convective season (June-September). Primary findings include:

- Useful information in the NDFD forecasts is lost when considering on the “Likely” and above categories. Forecasts using the “Trace” and above categories performed consistently as well as or better than the “Likely” and above forecasts.
- The NDFD Trace and above forecast performance is comparable to the state-of-the-art forecast products included in this report.
- LAMP outperforms the other forecast products, especially for earlier lead times.
- Making use of the higher temporal resolution of the LAMP, RAP, and HRRR forecasts (NDFD forecasts are valid only every 3 h) improved the performance of only the HRRR.
- All forecast products examined herein fall short of the Mid-term Operating Capability requirements established by the Traffic Flow Management Weather Working Group (TRWG) for each of the four requirements categories (i.e., Probability of Detection, False Alarm Ratio, Temporal Displacement, and Spatial Displacement).

Table of Contents

1	Introduction	1
2	Requirements	1
3	Data: Characteristics and Constraints	2
3.1	National Digital Forecast Data (NDFD) Forecasts	3
3.2	WRF Rapid Refresh (RAP)	3
3.3	Localized Aviation MOS Product (LAMP)	4
3.4	High Resolution Rapid Refresh (HRRR)	4
3.5	Corridor Integrated Weather System (CIWS)	5
3.6	National Lightning Data Network (NLDN)	5
4	Approach	7
5	Methods	8
5.1	Defining the Thunderstorm	8
5.2	Methods for Terminal Assessment	8
5.2.1	Use of Coverage to Identify Instantaneous Events	8
5.2.2	Merging Events	9
5.2.3	Matching of the Forecast and Observed Merged Events	10
5.2.4	Displacement Calculations	10
5.3	Methods for Jetway Assessment	12
5.3.1	Use of Echo Top Filter	12
5.3.2	Application of Flow Constraint Index (FCI)	12
5.3.3	Use of FCI for Identification of Instantaneous Events	13
5.3.4	Event Definition, Merging, and Matching	13
5.4	Scoring Events	13
6	Results	14
6.1	Terminal	14
6.2	Jetway	21
6.3	Non-NDFD-centric	22
7	Conclusions and Discussion	25
	Acknowledgments	28

REFERENCES.....28

List of Figures

Figure 3.1: Example of observation and forecast fields. Lighter shades of red for NDFD and LAMP fields indicates lower forecast probabilities.	6
Figure 5.1: Illustration of an instantaneous event. Terminal represented by 75 nmi radius, coverage represented by red objects.	9
Figure 5.2: Schematic illustrating the matching of forecast and observed events.	10
Figure 5.3: Schematic illustrating the center of mass determination for forecast (red) and observed (blue) objects. For the observations and all forecasts except the HRRR, the center of mass is computed separately for each object (left panel). For the HRRR, individual objects are first grouped they are within 20 nm of each other; a center of mass is then computed for each group (right panel).	11
Figure 5.4: Illustration of the temporal displacement between the onset of an event and the forecast onset of an event. $\Delta T_{\text{onset}} = F_{\text{onset}} - O_{\text{onset}}$; $\Delta T_{\text{cessation}} = F_{\text{cessation}} - O_{\text{cessation}}$	12
Figure 5.5: Example of the corridors (blue polygons) defined by applying 20-nm buffers around jet route segments. The red areas designate thunderstorms within corridors.	13
Figure 6.1: Number of terminal forecast (solid) events as a function of forecast lead time for the RAP (green), LAMP (blue), HRRR (brown) models and the NDFD-L (black) and NDFD-T (gray) forecast products. The number of observed events is shown by the dashed line.	15
Figure 6.2: Probability of detection (POD) for terminal domains as a function of forecast lead time.	16
Figure 6.3: As in Fig. 6.2, but for false alarm ratio (FAR).	17
Figure 6.4: As in Fig. 6.2, but for correspondence ratio (CR).	18
Figure 6.5: As in Fig. 6.4, but for event cessation.	19
Figure 6.6: Magnitude of the average temporal displacement (min) for event onsets; the event must be present in both the forecast and the observations for a displacement to be calculated.	20
Figure 6.7: As in Fig. 6.6, but for average spatial displacement (nmi).	20
Figure 6.8: As in Fig. 6.4, but for the jetway domain.	21
Figure 6.9: As in Fig. 6.8, but for FAR.	22
Figure 6.10: Probability of detection (POD) for terminal domains as a function of forecast lead time for the non-NDFD-centric (solid) and NDFD-centric (dashed lines) approach.	23
Figure 6.11: As in Fig. 6.10, but for FAR.	24
Figure 6.12: As in Fig. 6.10, but for CR.	24

List of Tables

Table 2.1: Terminal Requirements Established by the Traffic Flow Management Weather Working Group (2011).....	2
Table 3.1: Summary of the products evaluated.....	3
Table 3.2: Available forecast lead times (maroon) at 3-h valid-time increments (green boxes along the top) for the operational NDFD product.	4
Table 4.1: Differences between Phase 1 and Phase 2 of this project.....	7
Table 5.1: NDFD forecast coverage thresholds.	9
Table 5.2: Metrics and Measures at Onset and Cessation of an Event.....	14
Table 7.1: Summary statistics for all terminal regions for the NDFD-T forecast product compared to the MOC requirements.	26
Table 7.2: POD over time for four human-generated forecasts: winter (JFM) icing AIRMETs, winter (JFM) turbulence AIRMETs, the Collaborative Convective Forecast Product (CCFP), spring (AMJ) convective SIGMETs; and two model forecast: 0.25"/day precipitation forecasts from the Global Forecast System (GFS), and visibility forecasts from the Rapid Update Cycle (RUC). The bottom row shows the expected POD for the NDFD-T 6-h terminal forecast in the year 2022, given the same linear rate of improvement as the given forecast products. POD values were taken from the Real-Time Verification System (http://rtvs.noaa.gov).	27

1 Introduction

The Federal Aviation Administration (FAA) and National Weather Service (NWS) established a joint Traffic Flow Management Weather Requirements Working Group (TRWG) 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) was assigned to baseline current forecast performance of the NWS National Digital Forecast Database (NDFD) convective forecasts.

An initial assessment of the NDFD forecasts was completed and results are summarized in Lack et al. (2012). Although initial results indicated a significant departure from the TRWG stated requirements, questions arose regarding similarities/differences between the quality of the NDFD forecasts and other forecasts used today for traffic flow planning, such as the WRF Rapid Refresh (RAP), the High Resolution Rapid Refresh (HRRR), and the Localized Aviation MOS Product (LAMP). To address these questions, the FIQAS Team was tasked by the TRWG to perform a second evaluation and to establish a baseline of forecast performance relative to the TRWG requirements. The results from this second intercomparison of forecast performance are presented in this document.

Objectives for this study are to address the following:

- Does NDFD perform as well as other convective forecasts used for ATM planning?
- How well do other forecasts perform relative to the TRWG requirements?

This second study covers the period from 1 June – 30 September 2012. Measures of forecast accuracy and skill scores are presented in terms of forecast lead time to onset and cessation of convective events. Displacement of forecasts relative to the location of the observations is also computed, both within the terminal airspace (i.e., 75-nm radius around Core 30 airports, excluding Hawaii) over the CONUS and for specific jetway domains in the northeast.

2 Requirements

The TRWG established NextGen performance metrics for a variety of aviation impact variables that are expected to meet both near-term and mid-term operating requirements. Table 2.1 presents a draft version of requirements for terminal-area metrics for convective forecast products. The statistics of interest when an event occurs are probability of detection (POD), false alarm ratio (FAR), correspondence ratio (CR), timing error with respect to time of onset and cessation, and the spatial displacement error at the time of onset and cessation. Metrics are reported specifically for the 2, 4, 6, and 8-h lead times to onset and cessation of significant convective events in the terminal area and jetway domains, although hourly leads were also considered during the assessment.

Table 2.1: Terminal Requirements Established by the Traffic Flow Management Weather Working Group (2011).

		Near Term	MOC	Near Term	MOC	Near Term	MOC	Near Term	MOC	Near Term	MOC	
			± 5 min	± 5 min	≥ 90%	≥ 90%	≤ 20%	≤ 15%	± 5 min	± 5 min	≤ 3 nm	≤ 3 nm
Thunderstorms for Core Airports with: Probability ≥ 50% Area Diameter ≤ 150 nm	Time of Onset	1 h	± 5 min	± 5 min	≥ 90%	≥ 90%	≤ 20%	≤ 15%	± 5 min	± 5 min	≤ 3 nm	≤ 3 nm
		0 - ≤ 2 h	± 15 min	± 10 min	≥ 80%	≥ 85%	≥ 20%	≥ 15%	± 15 min	± 10 min	≤ 3 nm	≤ 3 nm
		> 2 - ≤ 4 h	± 30 min	± 20 min	≥ 75%	≥ 80%	≥ 25%	≥ 20%	± 30 min	± 20 min	≤ 3 nm	≤ 3 nm
		> 4 - ≤ 6 h	± 45 min	± 30 min	≥ 70%	≥ 75%	≥ 35%	≥ 25%	± 45 min	± 30 min	≤ 3 nm	≤ 3 nm
		> 6 - ≤ 8 h	± 60 min	± 45 min	≥ 65%	≥ 75%	≥ 40%	≥ 30%	± 60 min	± 45 min	≤ 3 nm	≤ 3 nm
		> 8 - ≤ 18 h		± 60 min		≥ 70%		≥ 35%		± 60 min		≤ 3 nm
		> 18 - ≤ 36 h		± 75 min		≥ 65%		≥ 40%		± 75 min		≤ 3 nm
		> 36 - ≤ 48 h		± 90 min		≥ 60%		≥ 45%		± 90 min		≤ 3 nm
		1 h	± 5 min	± 5 min	≥ 90%	≥ 90%	≤ 20%	≤ 15%	± 5 min	± 5 min	≤ 3 nm	≤ 3 nm
Time of Cessation		0 - ≤ 2 h	± 15 min	± 10 min	≥ 80%	≥ 85%	≥ 20%	≥ 15%	± 15 min	± 10 min	≤ 3 nm	≤ 3 nm
		> 2 - ≤ 4 h	± 30 min	± 20 min	≥ 75%	≥ 80%	≥ 25%	≥ 20%	± 30 min	± 20 min	≤ 3 nm	≤ 3 nm
		> 4 - ≤ 6 h	± 45 min	± 30 min	≥ 70%	≥ 75%	≥ 35%	≥ 25%	± 45 min	± 30 min	≤ 3 nm	≤ 3 nm
		> 6 - ≤ 8 h	± 60 min	± 45 min	≥ 65%	≥ 75%	≥ 40%	≥ 30%	± 60 min	± 45 min	≤ 3 nm	≤ 3 nm
		> 8 - ≤ 18 h		± 60 min		≥ 70%		≥ 35%		± 60 min		≤ 3 nm
		> 18 - ≤ 36 h		± 75 min		≥ 65%		≥ 40%		± 75 min		≤ 3 nm
		> 36 - ≤ 48 h		± 90 min		≥ 60%		≥ 45%		± 90 min		≤ 3 nm

3 Data: Characteristics and Constraints

Table 3.1 lists the products used in the study, the fields considered for each product, and the thresholds considered for each field. A summary of the products and the observations are provided in Sections 3.1 – 3.6.

Table 3.1: Summary of the products evaluated.

Product	Field and Threshold for Terminal Domain	Field and Threshold for En-route Domain
National Digital Forecast Database (NDFD)	Assessed at two thresholds: 1) Thunderstorm \geq LIKELY, or SVR 2) Thunderstorm \geq TRACE, or SVR	Same as terminal
WRF Rapid Refresh (RAP)	Variable used: Convective precip (ACPCP) \geq 1mm	Same as terminal
High Resolution WRF Rapid Refresh (HRRR)	Variable: Composite reflectivity (REFL) \geq 35 dBZ and lift index(LI) \leq 1	Terminal with echo top filter (ET > 30 kft)
LAMP	Variable: T-storm probability	Same as terminal
Observations: Derived deterministic Thunderstorm using CIWS and NLDN	VIL combined with lightning in a climatologically-based statistical thunderstorm diagnostic (see below)	Terminal + echo top filter (ET > 30 kft)

3.1 National Digital Forecast Data (NDFD) Forecasts

The NDFD Forecast Thunderstorm Probability field disseminated on a 5-km output grid (Glahn et al. 2003) is used in this study. Two thresholds are applied to this field for assessment: trace-and-above ($\geq 15\%$ probability) and likely-and-above ($\geq 55\%$ probability).

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, see Table 3.2), which restricts the number of forecasts that can be incorporated into the analysis. 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, a much coarser issuance frequency than the other products included in the assessment.

3.2 WRF Rapid Refresh (RAP)

The RAP model (Benjamin et al. 2006) is a regional, mesoscale model produced hourly with 13-km grid spacing. The model resolution is too large to allow for explicit representation of thunderstorms. Rather, the model employs a parameterized convective scheme to simulate the

effect of thunderstorms on the environment (e.g., the redistribution of heat and moisture in the column). RAP forecasts are issued every hour with hourly leads out to 18 h.

3.3 Localized Aviation MOS Product (LAMP)

LAMP is a forecast system that produces post-processed statistical output from the Global Forecast System (GFS) model (Ghirardelli, 2005). The LAMP Thunderstorm Probability field uses recent surface observations combined with the Global Forecast System (GFS) model and a climatological background field to produce forecast probabilities for the likelihood of a thunderstorm in a 2-h window. The definition of a thunderstorm is closely tied to the occurrence of lightning. The LAMP Thunderstorm Probability field is available on the same 5-km grid as the National Weather Service’s (NWS) National Digital Forecast Database (NDFD), with hourly updates, and forecast lead times from 1 to 25 h.

Table 3.2: Available forecast lead times (maroon) at 3-h valid-time increments (green boxes along the top) for the operational NDFD product.

		Valid Time (UTC)																							
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Issuance Time (UTC)	0				3			6																	
	1				2			5			8														
	2				1			4			7														
	3							3			6														
	4							2			5			8											
	5							1			4			7											
	6										3			6											
	7										2			5			8								
	8										1			4			7								
	9													3			6								
	10													2			5			8					
	11													1			4			7					
	12																3			6					
	13																2			5			8		
	14																1			4			7		
	15																			3			6		
	16	8																		2			5		
	17	7																		1			4		
	18	6																					3		
	19	5			8																		2		
	20	4			7																		1		
	21	3			6																				
	22	2			5			8																	
	23	1			4			7																	

3.4 High Resolution Rapid Refresh (HRRR)

The HRRR model (Weygandt et al. 2010) is available hourly with 15-min lead-time increments and provides a host of output grids, including the composite reflectivity, lifted index (a measure of convective instability), and echo-top fields used in this evaluation. The boundary conditions for the HRRR are provided by the WRF Rapid Refresh (RAP) model. At 3-km grid spacing, the HRRR model is at the edge of what is called “convection-permitting” resolution. The model is unable to resolve all of the processes within a convective cloud, but grid-scale convection is well-behaved (e.g., one

type of “bad” behavior is that at larger grid spacing, without the use of convective parameterization, models often produce single-grid-box-scale anomalously high precipitation totals).

3.5 Corridor Integrated Weather System (CIWS)

The observations used in this study are the Corridor Integrated Weather System (CIWS) Vertically Integrated Liquid (VIL) and echo top fields (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 parameters. The CIWS analysis VIL and echo top fields are issued every 2.5 min at a spatial resolution of 1 km. In this study, the CIWS echo top is only used to stratify relevant convection within the jetway domains. The CIWS VIL field is used along with the National Lightning Data Network (NLDN) data to infer areas of thunderstorms.

3.6 National Lightning Data Network (NLDN)

The National Lightning Data Network (NLDN) combines over 100 ground-based sensors across the United States. The sensors identify cloud-to-ground lightning strikes based on an electromagnetic signature. Triangulation between sensors is used to determine the location of the strike. The NLDN is capable of capturing 90-95% of all cloud-to-ground lightning strikes, but cloud-to-ground strikes make up only 20-30% of all lightning (Mackerras et al. 2012). An example of the forecasts and observations using the fields and thresholds listed in Table 3.1 is shown in Figure 3.1.

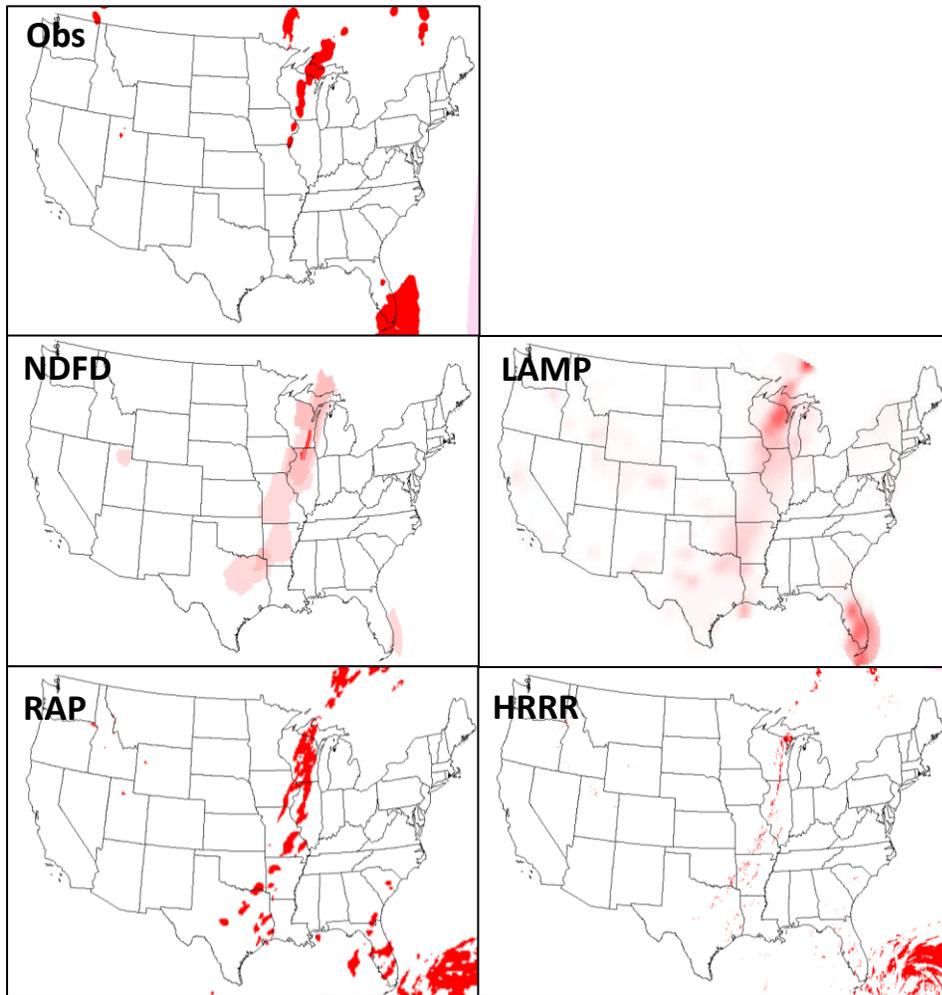


Figure 3.1: Example of observation and forecast fields. Lighter shades of red for NDFD and LAMP fields indicates lower forecast probabilities.

4 Approach

Although the overall approach used in this study is similar to Lack et al. (2012), significant changes to the methodologies are introduced. These changes are summarized in Table 4.1.

Table 4.1: Differences between Phase 1 and Phase 2 of this project.

Phase 1	Phase 2
NDFD thunderstorm coverage derived deterministically (threshold)	Thunderstorm coverage for probabilistic forecasts (NDFD, LAMP) based upon probability values.
CIWS echo top with 30 kft threshold was used as observation	CIWS VIL in conjunction with NLDN lightning used as observation
Single Center of mass for spatial displacement calculation, where center of mass is computed using all forecast (observation) objects in the terminal region.	<p>For all products except HRRR, the center of mass is computed individually for each forecast and observation object in the terminal region. Overall spatial displacement between the forecast and observation is computed using distances between the centers of mass for the individual forecast and observation objects.</p> <p>For HRRR, observation objects of close proximity are first grouped, and the center of mass is computed for each group of objects. Overall spatial displacement is computed using distances between the centers of mass of the HRRR groups and the individual observation objects.</p>
+/- 3h window for temporal matching of onset, cessation of events	A +/- 3 h time window will be applied, after which the Gale-Shapley procedure for optimizing pairings between forecast and observations will be used.

Statistics for the Core 30 airports are computed. Each terminal region is determined by a 75 nm radius around the airport. Note that in the application of the technique, only the field intersecting the terminal region is used. The extent of the field outside of the terminal region is eliminated from the event characterization.

Product comparisons will be performed in two ways:

- An ‘NDFD-centric’ way that constrains the issues, leads, and temporal characteristics of the other forecast products to that of NDFD. The NDFD has the coarsest temporal resolution

(3 h vs 1 h for the other forecasts) and has only a subset of issues and leads that the other products have at a given valid time. To ensure fairness to NDFD, in this type of comparison only the issues and leads corresponding to those of NDFD will be included, and a 3-h time window is used for event definition.

- A ‘non-NDFD-centric’ way where NDFD is excluded from the comparison for fairness. All issues and leads for the other products are included, and a 1-h time window is used for event definition.

Data outages are tracked for reference, but are excluded from the process of determining an event. Further concept development is necessary to determine the correct approach for incorporating outages that treats all products fairly, as more simplistic approaches of eliminating event information for a product due to an outage could penalize some of the products. A missing forecast is treated as though it were a forecast of no event.

5 Methods

The verification approaches vary according to the domain: terminal and jetway. Each will be summarized in this section.

5.1 *Defining the Thunderstorm*

For this assessment a thunderstorm is defined as moist convection with lightning. For the terminal domain, convection of any height is considered; for the en-route domain, only convection with echo tops greater than or equal to 30 kft is considered. The observation fields are created through a machine learning process in which VIL intensities and NLDN lightning were used to identify the existence of lightning strikes found in the Global Lightning Detection Network, GLD360, over a convective season. The resulting VIL and NDLN characteristics required to produce a thunderstorm observation were chosen to ensure that at least 95% of all GLD360 lightning strikes were captured and vary geographically. Note that GLD360 data is not operational and so was not used in the assessment itself, but only as an independent dataset for developing the VIL- and NLDN-based observation field.

5.2 *Methods for Terminal Assessment*

5.2.1 **Use of Coverage to Identify Instantaneous Events**

Percent coverage over each terminal region is computed for each product issuance based upon the field threshold defined in Table 3.1. An example is shown in Figure 5.1. Probabilistic forecasts are treated probabilistically in that the coverage computed at each pixel is weighted by its probability value. Table 5.1 provides the NDFD categories and corresponding probability value ranges. The NDFD-T field includes all four categories with 15% probability; NDFD-L includes ‘Likely’ and ‘Occasional’ Categories with 55% probability. All probability values are used for the LAMP product. Values for HRRR and RAP are either 0 or 1. An instantaneous event is determined when the computed coverage reaches or exceeds 10%.

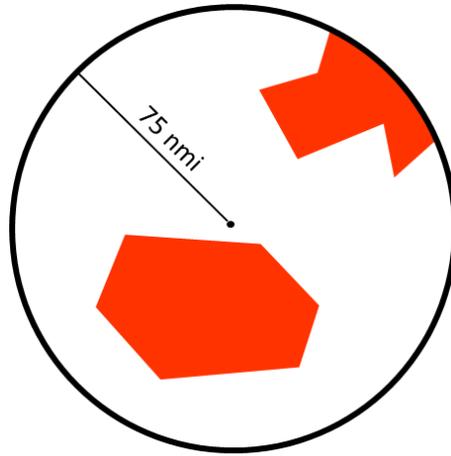


Figure 5.1: Illustration of an instantaneous event. Terminal represented by 75 nmi radius, coverage represented by red objects.

Table 5.1: NDFD forecast coverage thresholds and corresponding probability value ranges.

		Category	Probability
Trace and above		slight chance	15 - 24.99 %
		chance	25 - 54.99 %
	Likely and above	likely	55% - 74.99%
		occasional	75% - 100%

5.2.2 Merging Events

Instantaneous events are merged as part of a larger event if the temporal gap between them is less than or equal to a given temporal threshold. For the assessment, results are processed using two temporal thresholds: 1 h (non-NDFD-centric) and a 3 h (NDFD-centric). For the forecasts, events will be constructed for a fixed lead using consecutive issuances.

Note that the spatial and temporal characteristics of an event onset (e.g., onset time and location) are determined from the instantaneous event marking the onset of the merged event (i.e., the first in the set of instantaneous events comprising the merged event). Similarly, the spatial and temporal characteristics of an event cessation are taken from the last of the set of instantaneous events comprising the merged event. Spatial information from intermediate forecasts will not contribute to the overall event information.

5.2.3 Matching of the Forecast and Observed Merged Events

Forecast and observation events are matched temporally (spatial criteria are not included in the matching). Onset and cessation are treated separately, in that a forecast and observation can be matched as a hit for onset but not necessarily cessation, and vice versa. To be a candidate for matching, forecast and observation onset/cessation events must occur within 3 h of each other. Of the candidate matches, a final matching is determined using the Gale-Shapley (1962) procedure, where shorter temporal distances between forecast and observation are preferred. Note that a forecast event may not be matched to its closest temporal observation if that observation event has another, more preferred (temporally closer) forecast object. It would be matched with the next-closest available candidate, should one exist. Matches are computed per forecast lead.

Note that a forecast event will be matched with only one observation event, and vice versa. Some forecast or observation events may remain single, with no match, either by not meeting the 3 h criteria, or by failing to be a sufficient candidate during the matching algorithm. A match is considered a hit, a forecast with no observation match is considered a false alarm, and an observation with no forecast match is considered a miss. Examples of the matching outcomes are provided in Figure 5.2.



Figure 5.2: Schematic illustrating the matching of forecast and observed events.

5.2.4 Displacement Calculations

Displacement errors between events will only be computed for forecast/observation matches (hits), and therefore will be considered separately for event onset and cessation. As mentioned earlier, the spatial and temporal information corresponding to the instantaneous forecast (observations) determining the onset will be used in displacement computations for onset, and analogously for cessation.

5.2.4.1 Computation of Center of Mass for Thunderstorm Objects

The displacement calculation is computed by measuring the distance between the forecast and observation centers of mass. For all products, only the field intersecting the terminal region is included in the center of mass computation. The thunderstorm objects within the terminal region are determined using the fields and thresholds as described in Table 3.1.

5.2.4.1.1 Multi-object approach

For all products (RAP, LAMP, NDFD, and the CIWS/NLDN-derived thunderstorm observations) except the HRRR, the center of mass is computed for each individual object in the terminal. The center of mass for NDFD and LAMP is derived by weighting each pixel by its probability value.

Figure 5.3 shows an illustration of the matched forecast (red) and observation (blue) objects where the distance between the centers of mass is measured and the displacement calculated.

5.2.4.1.2 Grouping approach for high resolution products

For HRRR, which has smaller-scale thunderstorm objects, objects are grouped according to proximity to determine larger-scale objects. Individual objects within a Euclidean distance of 20 nm of each other will determine a group. As illustrated in Figure 5.3 (right panel), once the groups are identified, the center of mass is computed for each group using the individual thunderstorm objects composing the group.

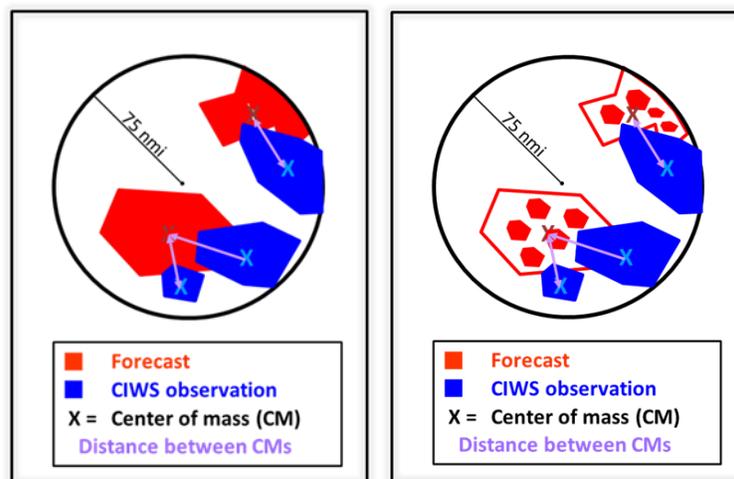


Figure 5.3: Schematic illustrating the center of mass determination for forecast (red) and observed (blue) objects. For the observations and all forecasts except the HRRR, the center of mass is computed separately for each object (left panel). For the HRRR, individual objects are first grouped they are within 20 nm of each other; a center of mass is then computed for each group (right panel).

5.2.4.2 Spatial Displacement

To compute spatial displacement for a hit, spatial information of the forecast at onset (cessation) is compared to the spatial information of the observations at onset (cessation), regardless of the temporal offset between the two. Spatial displacement is computed using the centers of mass of the forecast and observation objects (or in the case of HRRR, groups of objects) corresponding to the onset or cessation event (Figure 5.3). For each forecast object (or group) associated with the event, the minimum distance of its center of mass to that of an observation object will be identified. Similarly, for each observation object, the minimum distance to a forecast object (group) will be determined. The overall displacement is computed by taking the average of the minimum distances between forecast and observation objects. To aggregate results of displacement, the average is used.

5.2.4.3 Temporal Displacement

Temporal displacement will merely be the difference in valid times between the forecast and observation (Figure 5.4). Aggregation of temporal displacement uses the average of the time differences.

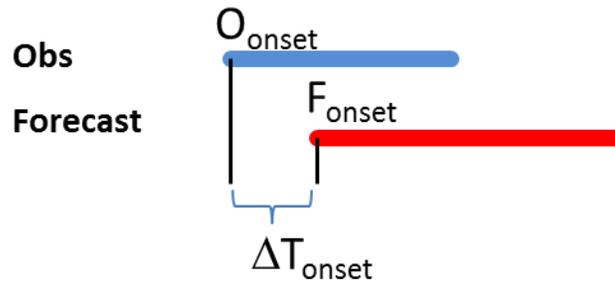


Figure 5.4: Illustration of the temporal displacement between the onset of an event and the forecast onset of an event. $\Delta T_{\text{onset}} = F_{\text{onset}} - O_{\text{onset}}$; $\Delta T_{\text{cessation}} = F_{\text{cessation}} - O_{\text{cessation}}$.

5.3 Methods for Jetway Assessment

The methods for the jetway domain focuses on a subset of jetways within the region in the NE corridor bounded by Flow Constrained Areas (FCA) A05 and A08 that either cross or are contained within these A05/A08 boundaries. The specific list used in the study was provided by the NWS and is provided in Appendix A. For the study, jetways are categorized as either North-South (N-S) or East-West (E-W), and statistics will be computed for jetway groups N-S, E-W, and All.

5.3.1 Use of Echo Top Filter

An echo top threshold is applied to CIWS and the HRRR to filter convective activity below 30 kft, which typically does not impact aircraft en-route to a destination. This filter is applied as a first step, prior to any translation of the product to identify instantaneous events.

5.3.2 Application of Flow Constraint Index (FCI)

The FCI technique (Layne and Lack 2010), a measure of flow constraint within a given corridor, will be applied using a geometry defined by jetway segments. Specifically, the jetways are buffered by 20 nm on each side and partitioned by segments 80 nm in length, so that the FCI is computed for each 40x80 nm segment along the jetway (Figure 5.5). Note that the buffered jetways do not completely cover the NE region. For all products, only the portion of the field intersecting the buffered jetways is included in the computations for FCI and center of mass.



Figure 5.5: Example of the corridors (blue polygons) defined by applying 20-nm buffers around jet route segments. The red areas designate thunderstorms within corridors.

5.3.3 Use of FCI for Identification of Instantaneous Events

For deterministic forecasts, the field threshold as defined in Table 3.1 is used to compute the FCI. For probabilistic forecasts, the FCI is computed probabilistically (e.g. 80% probability area coverage over exactly half of the jetway cross-section produces an FCI value of $0.8 * 0.5 = 0.4$). Given a jetway grouping (N-S, E-W, or All), for each jetway in the grouping, the max FCI is determined by taking the maximum over the segments composing that jetway, yielding an overall FCI score for each jetway. An FCI threshold of 0.5 is used to determine if a jetway is constrained. If the number of constrained jetways for the jetway grouping exceeds 10%, an instantaneous event is identified.

5.3.4 Event Definition, Merging, and Matching

Similar to the coverage approach for the terminal, the FCI is used to identify instantaneous events for each issuance of a forecast or observation. Once the instantaneous events are identified, the remainder of the verification approach follows the terminal technique, including: the object grouping technique for HRRR to support the center of mass computation, the center of mass computation itself, the merging of instantaneous events, event characterization, matching of events, computation of spatial displacement, and computation of statistics. Note that the center of mass computation for the jetway uses the objects determined by the forecast or observation field intersected with the jetways in the grouping (N-S, E-W, or All).

5.4 Scoring Events

For the onset and cessation of the event, the following scores are computed (Table 5.2). The scores are computed for onset and for cessation separately.

Table 5.2: Metrics and Measures at Onset and Cessation of an Event

Statistic	Formula	Description
POD	$H/(H+M)$, H = hit, M=miss	<u>Probability of Detection</u> : Proportion of observed events that were correctly detected
FAR	$FA / (FA + H)$, FA = false alarm	<u>False Alarm Ratio</u> : Proportion of forecast events that actually did not occur
CR	$F \cap O / F \cup O$, F = forecast, O = observation	<u>Correspondence Ratio</u> : A measure of agreement (ratio of intersection to union) of forecasts and observations
Temporal Displacement		Time difference between onset of forecast and corresponding observed events; time difference between cessation of forecast and corresponding observed events.
Spatial Displacement		Location difference between forecast and corresponding observed event at onset; location difference between forecast and corresponding observed event at cessation.

6 Results

6.1 Terminal

Before looking at the verification scores for the terminals it is instructive to compare the number of events produced by each of the forecast products with the number of observed events (Figure 6.1) to provide further insight into the performance results discussed in subsequent sections. RAP (green) slightly overforecasts the number of events relative to the observations (dashed) at all lead times. LAMP (blue) moves from a slight overforecast to a slight underforecast as lead time decreases. NDFD-T (gray) has a somewhat stronger underforecast for all lead times, while HRRR (brown) and NDFD-L (black) have only about one-third of the observed number of events. (Note that in all subsequent plots the forecast products are represented by the same colors as in Figure 6.1.) Observe that the number of forecast events in the HRRR increases substantially over the first three hours, in contrast to the slow decline in forecast events from the other forecast products. This is likely a result of model spin-up: during this time period, the HRRR forecasts are initialized using the RAP model's data assimilation. Consequently, there is an adjustment period as the 3-km HRRR processes its 13-km initial fields. On 11 April, 2013, the HRRR began using its own 3-km assimilation scheme, with the expectation that this model spin-up should be greatly reduced.

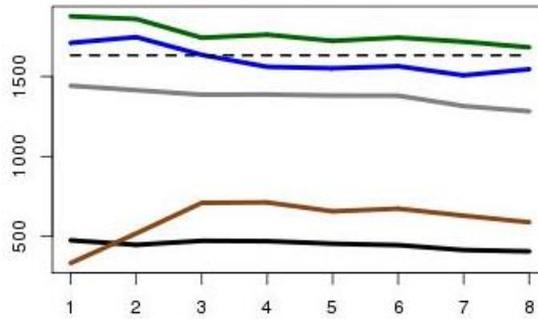


Figure 6.1: Number of terminal forecast (solid) events as a function of forecast lead time for the RAP (green), LAMP (blue), HRRR (brown) models and the NDFD-L (black) and NDFD-T (gray) forecast products. The number of observed events is shown by the dashed line.

As noted in Section 5.45.4, the probability of detection (POD) measures the ability of a forecast to capture an observed event. LAMP forecasts exhibit the best performance in capturing observed events (Figure 6.2): the POD is highest for the 2-h lead forecast, with almost 60% of all observed events correctly forecast, then declines gradually to nearly 40% by the 8-h lead time. The RAP model captures 10-15% fewer events at the early leads, but experiences a smaller drop in performance with longer leads, such that its POD matches that of the LAMP by the 7-h lead. NDFD forecasts using the “Trace” threshold follow the same pattern as the RAP model, but with a 10% reduction in POD. Meanwhile, the NDFD-L and HRRR forecasts capture fewer than 20% of all observed events. As shown in Figure 6.1, however, these forecasts suffer from a strong low bias, precluding the possibility of a high POD. For example, even if every NDFD-L forecast event was matched with an observed event, the resulting POD would be only around 0.3. It is for this reason that the NDFD-T forecasts were included. Similarly, if through post-processing the number of forecast events in the HRRR could be increased without reducing the model’s level of forecast performance, its POD should be in the range of the other forecasts examined in this study.

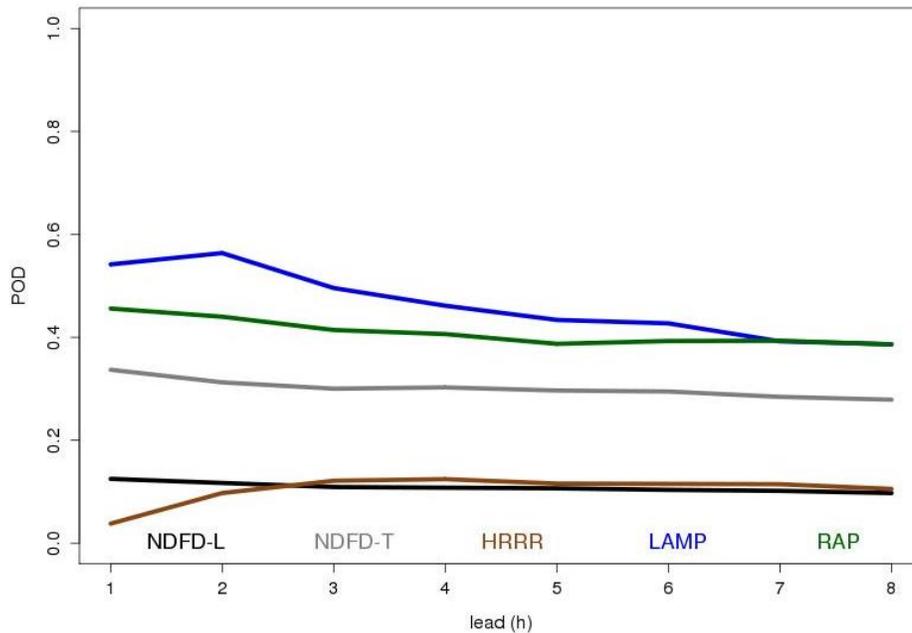


Figure 6.2: Probability of detection (POD) of event onset for terminal domains as a function of forecast lead time, for NDFD-Light threshold (black), NDFD-Trace threshold (gray), HRRR (brown), LAMP (blue), and RAP (green).

The false alarm ratio (FAR; Figure 6.3) measures the likelihood of a forecast to incorrectly identify the occurrence of an event, thus the score is inverted with 0 being a perfect FAR. Once again, the LAMP forecasts perform best at the earliest times, but also possess a larger decline with longer forecast leads, such that its FAR is similar to the other forecasts by the 7-h lead time. The RAP and the two NDFD forecasts all perform similarly with about 60% of all forecast event onsets not matching with an observed event onset. For RAP, a small percentage of these false alarms are a result of the model forecasting too many events (see Figure 6.1), but most of the error, and all of the error for the other forecasts, is a result of failing to place events in the right place at the right time. Forecast events that occur outside of the 3-h time window relative to observed events will be counted as false alarms. However, false alarms could also be the result of spatial errors: for example, an area of observed thunderstorms could lie slightly outside of a terminal domain so that a shift in the area covered by the forecast thunderstorms could place the latter inside the terminal domain. Consequently, the forecast meets the coverage threshold while the observations do not, resulting in a false alarm. The HRRR has a somewhat higher false alarm rate than the other forecasts.

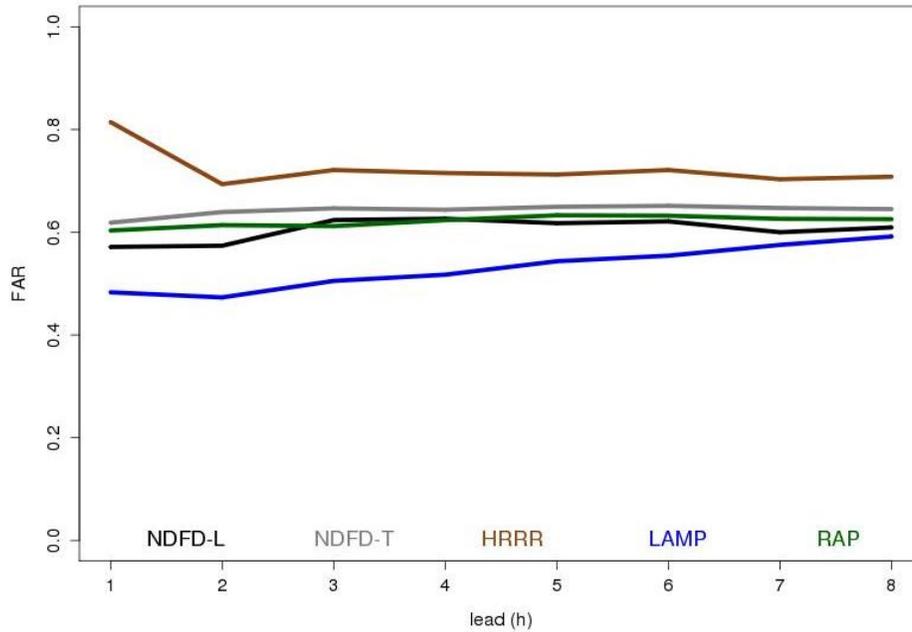


Figure 6.3: As in Fig. 6.2, but for false alarm ratio (FAR).

The information in the POD and FAR can be combined in the Correspondence Ratio (CR), a measure of association between forecasts and observations which credits forecasts for each hit while penalizing them equally for each miss and false alarm. In this case, the CR (Figure 6.4) looks very similar to the POD, with LAMP performing best and HRRR and NDFD-L receiving substantially lower scores. The discussion above, on the likelihood that post-processing would improve the performance of the HRRR forecasts, applies equally here. NDFD forecasts would likely benefit from post-processing as well; one expects that treating the probabilities as coverage amounts introduces biases into the forecasts that could be improved through calibration.

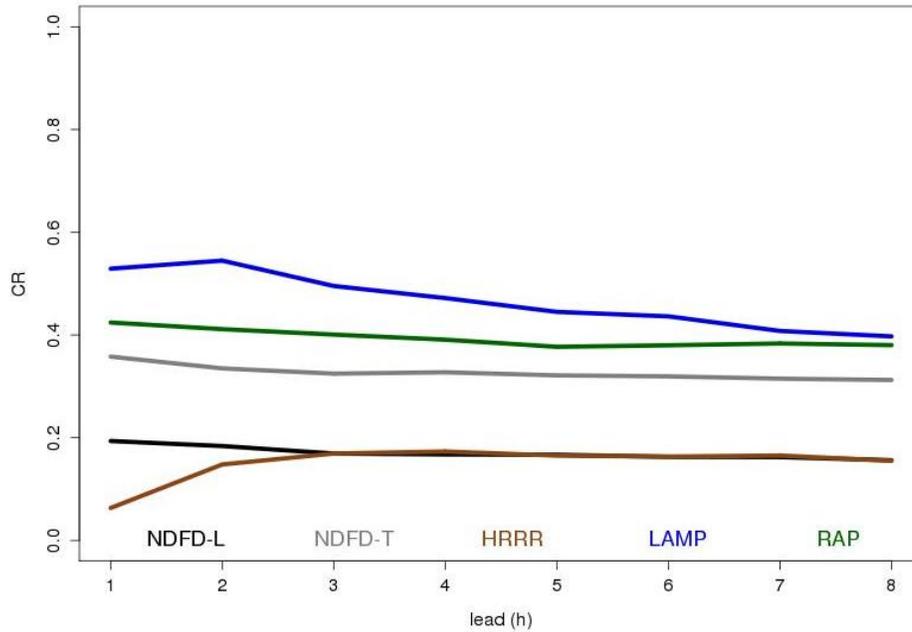


Figure 6.4: As in Fig. 6.2, but for correspondence ratio (CR).

When the focus is shifted from event onset to event cessation, some changes are seen (Figure 6.5). LAMP forecasts perform similarly at short leads as for event onset, but the drop in performance with longer leads is larger for cessation. RAP forecasts perform very similarly for both onset and cessation, but the NDFD-T forecast performance improves for cessation, such that the two forecasts have nearly identical CR. Interestingly, whereas the NDFD-T forecasts improve for cessation, NDFD-L forecast performance is mostly unchanged. HRRR forecasts perform better at the 1-h lead and then have a longer spin-up period (3-h vs. 2-h) compared with event onset, resulting in a significantly improved score for the middle and longer leads.

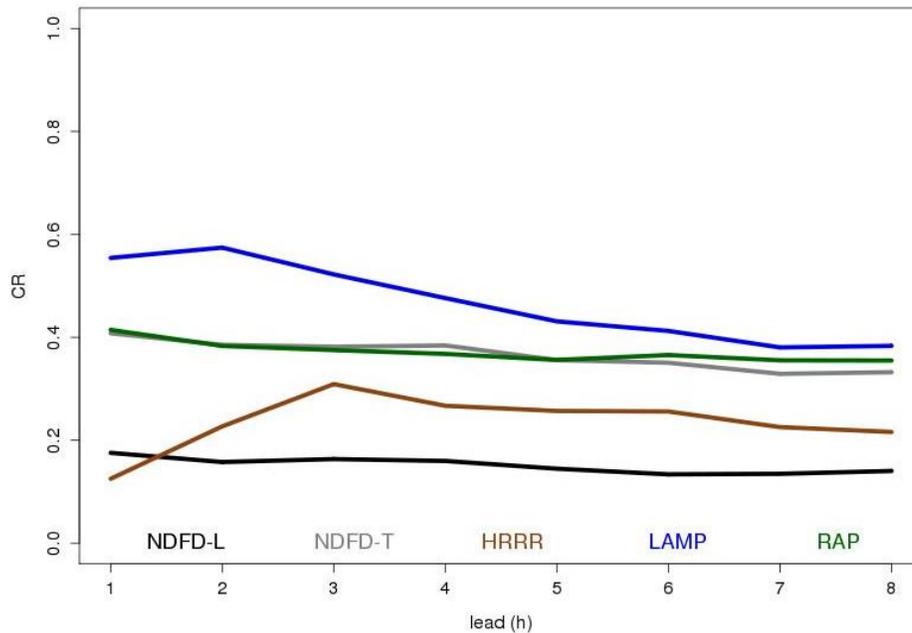


Figure 6.5: As in Fig. 6.4, but for event cessation.

Whenever a forecast event (either onset or cessation) exists within 3-h of an observed event, i.e. a hit, it is possible to compute temporal (Figure 6.6) and spatial (Figure 6.7) displacements of those forecast events, according to the methods described in Section 5.2.4. For these scores, the maximum possible value is determined by the event and region definitions: 3-h and 150 nmi for the temporal and spatial displacements, respectively. As with the previous scores, LAMP outperforms the other forecast products for shorter leads, but suffers from greater forecast degradation than the other forecasts and so possesses similar scores for longer leads. RAP again performs just a little behind LAMP, but is similar to the other forecasts, especially for temporal displacement. There is little difference between the two NDFD thresholds for either temporal or spatial displacement. The spatial displacement in the HRRR is similar to that of the NDFD forecasts, but the spin-up issue appears to affect the temporal displacements, with displacements exceeding 2-h for the 1-h lead forecasts. However, the timing errors decrease steadily with lead time, such that the HRRR errors are only slightly worse than the other forecasts by the 8-h lead.

In summary, the temporal forecast performance is only slightly better than a random sample. If forecasts and observations were randomly distributed within a 3-h window, the expected error would be 90 min. The RAP and NDFD forecasts are very close to this threshold, while LAMP beats it by around 10% at earlier lead times. It is somewhat more complicated to place the spatial errors in context. Generally, a random placement of objects within a 150-nmi diameter domain would yield an expected error of 75 nmi. However, because only the portion of the forecast object within the domain is considered, and because of the large size of many of the objects considered (Figure 3.1), the centroids of the objects will tend to lie closer to the center of the domain, thereby reducing the

expected baseline displacement. It is likely that the displacements shown in Figure 6.7 are near this baseline magnitude.

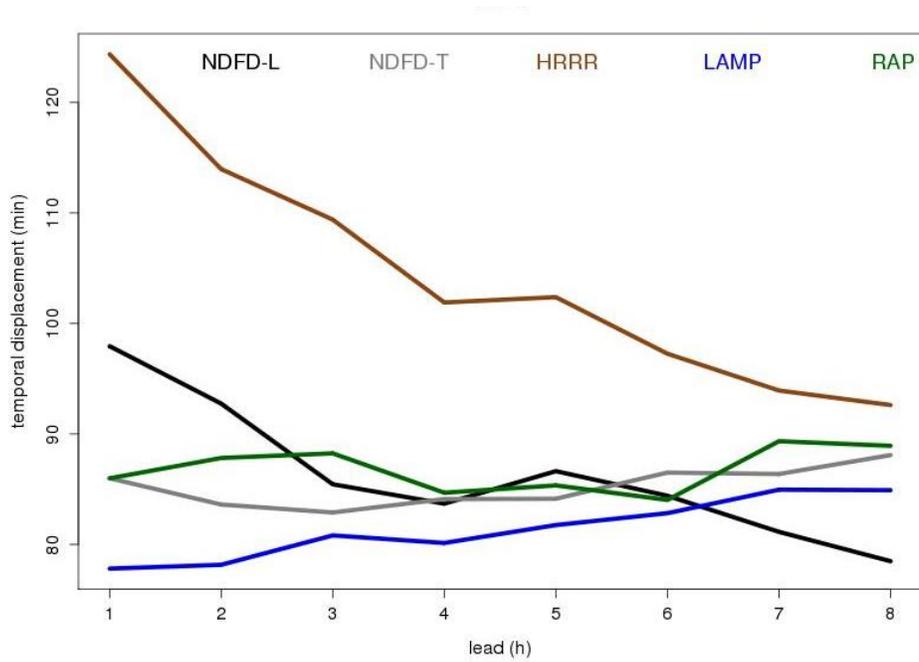


Figure 6.6: Magnitude of the average temporal displacement (min) for event onsets; the event must be present in both the forecast and the observations for a displacement to be calculated.

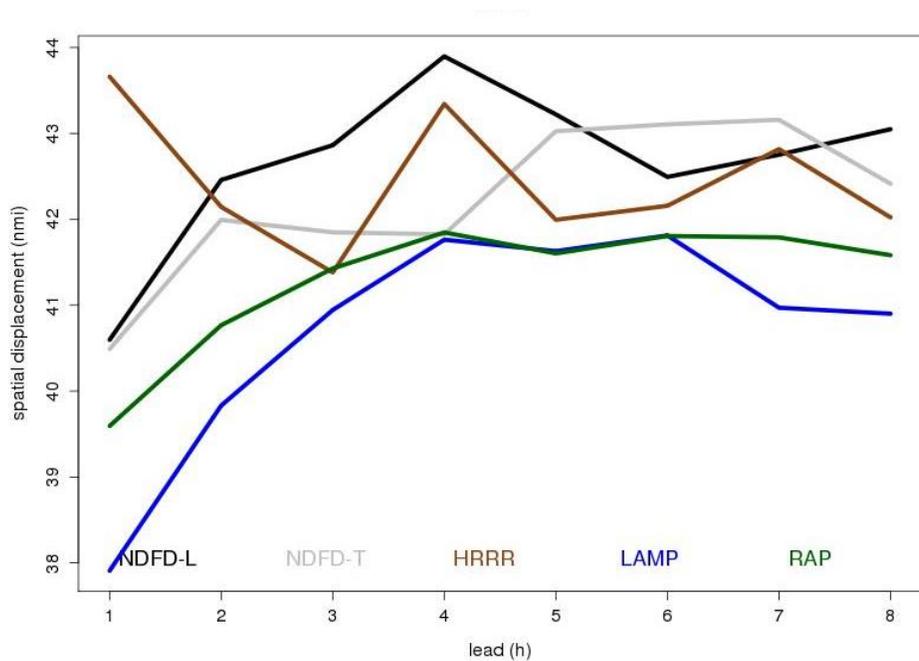


Figure 6.7: As in Fig. 6.6, but for average spatial displacement (nmi).

6.2 Jetway

Even though the jetway domain is much larger than the 75-nmi radius terminal domains, greater precision is required in the placement of the storms for the jetway domain. This is because the storms must be located within the jetway corridors in order to block traffic for that jetway. Consequently, forecast performance within the jetway domains is expected to be different than for the terminal domains, as is shown in Figure 6.8.

Once again LAMP outperforms the other products at the 1-h lead time. However, the gap between LAMP and the other products is not as large for the jetway domain as for the terminal domain and the degradation with increasing lead time is greater: the CR falls to less than 0.2 compared with just over 0.4 for the terminal domain. The performance of the high-resolution HRRR forecasts is much improved in the jetway domain over that measured for the terminal domain, such that it outperforms the RAP forecasts for most lead times. As explained above, the jetway domain requires greater precision than the terminal domain and so the improvement of the HRRR relative to the RAP forecasts is exactly what would be expected. The forecast improvement achieved by using the NDFD-T threshold compared with the NDFD-L threshold disappears for the jetway domain, with the two products performing almost identically.

The difference in CR for the jetway domain compared with the terminal domain comes mostly from differences in the FAR (Figure 6.9, cf. Figure 6.3). The HRRR FAR drops somewhat, while the RAP and LAMP FARs increase substantially, to over 0.8 (i.e., only 20% of all forecasts are hits).

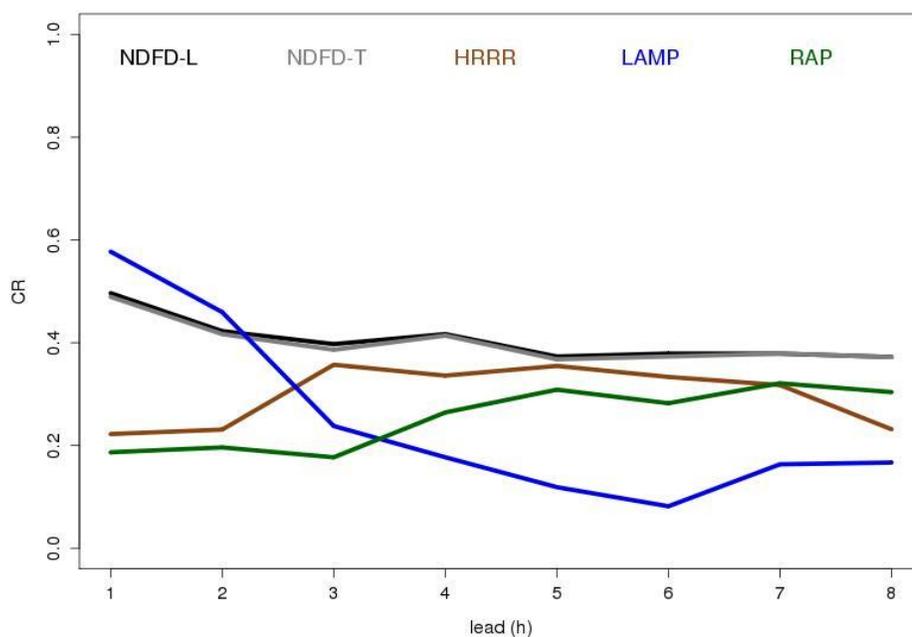


Figure 6.8: As in Fig. 6.4, but for the jetway domain.

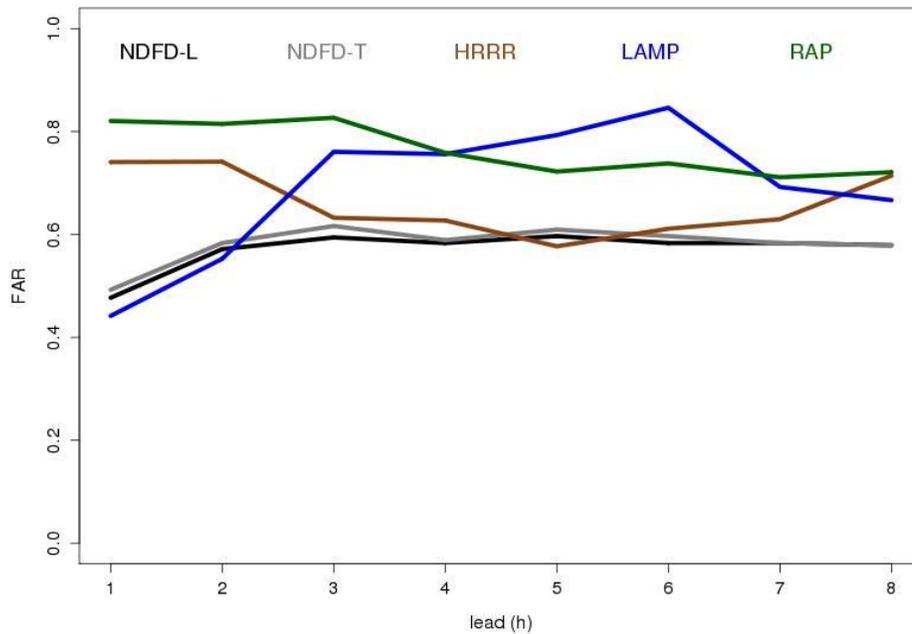


Figure 6.9: As in Fig. 6.8, but for FAR.

6.3 Non-NDFD-centric

The results presented in Section 6.1 used an NDFD-centric approach in which all products were “thinned” to match the 3-hour valid time increments of the NDFD forecasts. In this section, the effect of the thinning on the other forecasts products is examined. The allowable temporal separation for matching forecast and observed events is still 3 h, but instantaneous events must now be within 1 h in order to be merged into a single event.

Figure 6.10 shows the resulting POD scores (solid lines) along with the scores using the 3-h approach (dashed lines) for event onset—results are consistent with that seen for event cessation (not shown). Both RAP and HRRR PODs improve by 0.1 to 0.15 consistently for all leads. In contrast, the LAMP POD actually declines slightly for most lead times. Further inspection reveals that these results are not surprising. RAP and HRRR produce truly hourly forecasts. For LAMP, however, although the forecasts are produced every hour, each forecast is valid over a 2-h window and so the output resolution is somewhat misleading.

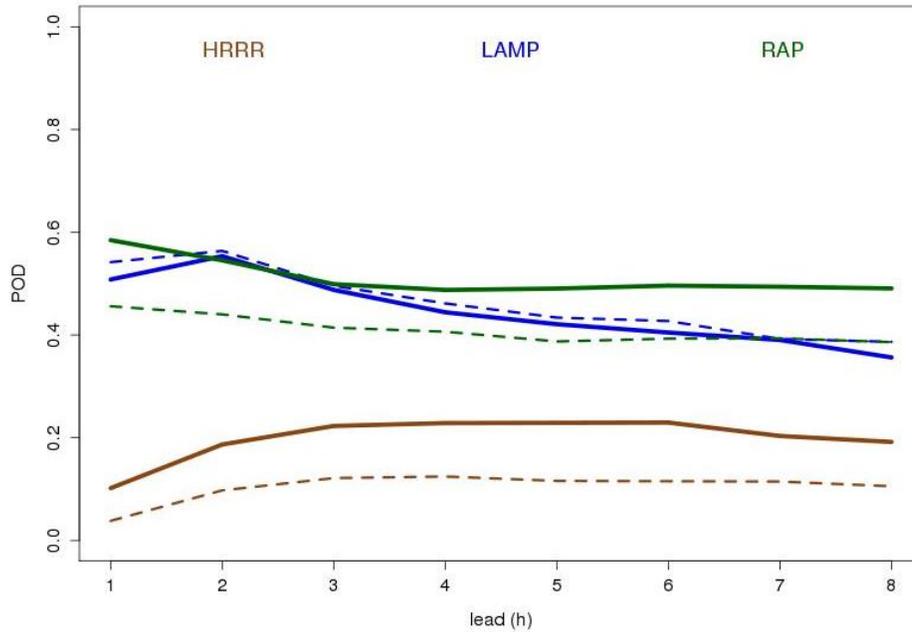


Figure 6.10: Probability of detection (POD) for terminal domains as a function of forecast lead time for the non-NDFD-centric (solid) and NDFD-centric (dashed lines) approach.

For FAR (Figure 6.11) as compared with POD, the HRRR forecasts see the same level of improvement, but LAMP experiences a larger degradation and RAP flips from improving when using 1-h merging to getting worse. It is not clear what could be responsible for this split behavior: better POD, but worse FAR. One possibility is that the 1-h window results in fewer merges and so more events; an indiscriminate increase in the number of forecasts will often produce an increase in both POD and FAR. The result, in terms of CR (Figure 6.12), is that switching from 3-h output to 1-h output leads to a near doubling of the performance of HRRR, little change for RAP, and a small decline for LAMP. Note that even with this boost in performance, HRRR still lags behind the other forecast products (and behind NDFD-T; cf. Figure 6.4).

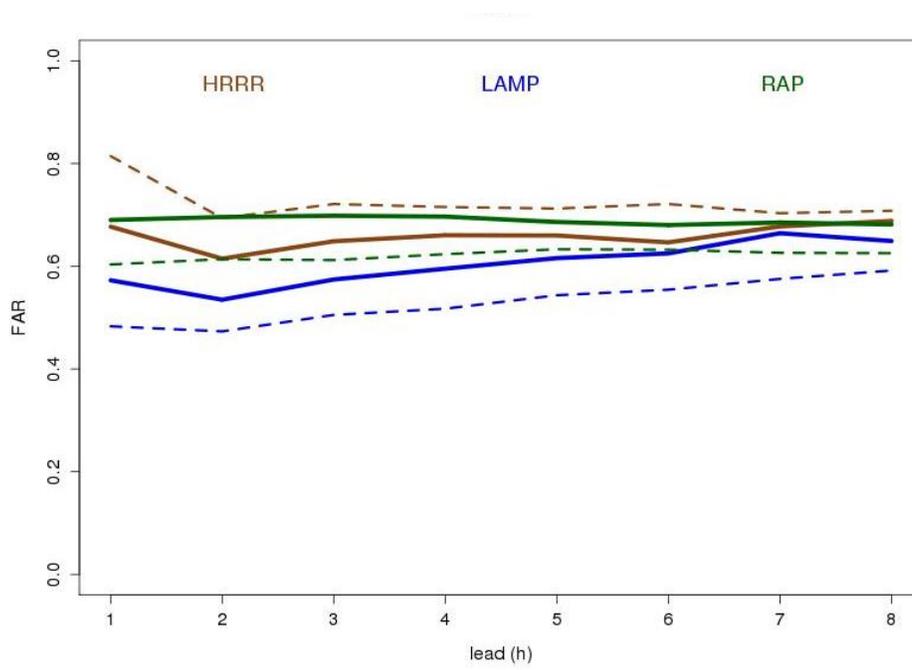


Figure 6.11: As in Fig. 6.10, but for FAR.

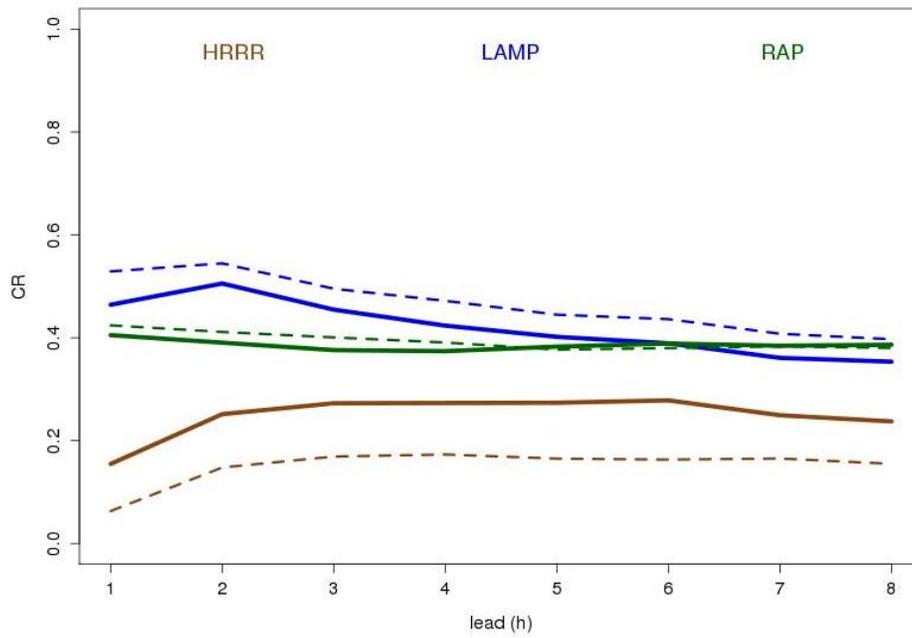


Figure 6.12: As in Fig. 6.10, but for CR.

7 Conclusions and Discussion

This assessment included two main areas of focus: 1) determine if the NDFD performs as well as other convective forecasts used for ATM planning and 2) determine how well other forecasts perform relative to the TRWG requirements.

Results indicate that there is useful information in the NDFD forecasts that is lost when considering only the “Likely” and above categories; the NDFD-T forecasts consistently performed at least as well as, and typically better than the NDFD-L forecasts. Furthermore, the NDFD-T forecast performance is comparable to the state-of-the-art forecast products examined herein.

Overall, for the terminal domain, LAMP outperforms the other forecast products, especially for the earlier lead times. The RAP model typically provides the next best forecast, followed by the NDFD-T forecasts. The HRRR shows evidence of suffering from model spin-up; forecast performance improves over the first few hours of the forecast, but remains well below the other forecast products, with the exception of NDFD-L. (This spring the HRRR updated its data assimilation package in a way that could substantially alleviate this spin-up problem.) Furthermore, several of the forecasts could likely be improved through post-processing; no post-processing or calibration was performed for this assessment.

The NDFD-T forecast performance as well as the performance of all the convective products falls short of meeting the TRWG-MOC requirements (Table 7.1). Although the requirements provide a target for the level of weather information needed for traffic flow planning, the state of the science and current forecast products are not yet at the spatial and temporal scales that allow these requirements to be adequately met. For example, the MOC requirement for the timing error for a forecast with a 2-h lead is 10 min, but the NDFD forecasts have only 3-h resolution. The only way to achieve an error less than 10 min would be if observed event onset and cessation occurred only within 10 min of the top of the hour for eight of the 24 valid hours in the day. Similarly, the 3 nmi requirement for spatial errors is roughly at the grid resolution of the forecasts. In other words, to meet the requirement the forecasts must place storms (technically the center of mass of the storms) in the very same pixel grid box as the observation. Meeting the MOC requirements would necessitate, at a minimum, forecast output at horizontal resolution below 1 km every 5-10 min. The hardware upgrade to support such an increase in spatial and temporal resolution would be substantial.

Table 7.1: Summary statistics for all terminal regions for the NDFD-T forecast product compared to the MOC requirements.

		POD		FARatio		Timing (min)		Location (nmi)	
		NDFD	MOC	NDFD	MOC	NDFD	MOC	NDFD	MOC
Onset	2 h	0.31	≥ 0.85	0.64	≤ 0.15	83.6	± 10	42.0	≤ 3
	4 h	0.30	≥ 0.80	0.64	≤ 0.20	84.1	± 20	41.8	≤ 3
	6 h	0.29	≥ 0.75	0.65	≤ 0.25	86.5	± 30	43.1	≤ 3
	8 h	0.28	≥ 0.75	0.65	≤ 0.30	88.1	± 45	42.4	≤ 3
Cessation	2 h	0.36	≥ 0.85	0.58	≤ 0.15	84.2	± 10	41.8	≤ 3
	4 h	0.36	≥ 0.80	0.58	≤ 0.20	86.3	± 20	43.1	≤ 3
	6 h	0.32	≥ 0.75	0.62	≤ 0.25	87.0	± 30	43.5	≤ 3
	8 h	0.30	≥ 0.75	0.62	≤ 0.30	88.7	± 45	43.5	≤ 3

To provide context for the degree of improvement necessary to bring the terminal event POD and FAR values up to the MOC requirements, the improvement in POD over the past decade of a variety of products serves as a set of predictors for NDFD improvement by 2022. Table 7.2 shows the POD for a suite of human- and model-generated forecasts over the last decade. A least-squares linear trend is then fit to each set of forecasts and the forecast improvement calculated as a fraction of the total possible improvement according to

$$improvement = \frac{POD_{2013} - POD_{2002}}{1 - POD_{2002}}$$

The improvement is then applied to the NDFD-T 6-h lead POD (see Table 7.1), assuming a rate of improvement equal to that achieved over the previous decade for this sample of forecasts, to give a set of predicted POD values for the year. Only with using the fastest rate of improvement (e.g., CCFP) is the predicted NDFD forecast POD more than halfway toward its goal a decade from now. Extrapolating forward in time, the rates of improvement shown in Table 7.2 would bring the NDFD 6-h forecast up to the MOC value between the years 2029 and 2044. It is worth noting that the human-generated forecasts shown in Table 7.2 are all at a substantially larger scale than the terminal forecasts examined in this evaluation, which might explain the faster rate of improvement achieved by these forecasts over the model-based forecasts included in Table 7.2.

Table 7.2: POD over time for four human-generated forecasts: winter (JFM) icing AIRMETs, winter (JFM) turbulence AIRMETs, the Collaborative Convective Forecast Product (CCFP), spring (AMJ) convective SIGMETs; and two model forecast: 0.25"/day precipitation forecasts from the Global Forecast System (GFS), and visibility forecasts from the Rapid Update Cycle (RUC). The bottom row shows the expected POD for the NDFD-T 6-h terminal forecast in the year 2022, given the same linear rate of improvement as the given forecast products. POD values were taken from the Real-Time Verification System (<http://rtvs.noaa.gov>).

Year	Icing AIRMETs	Turbulence AIRMETs	CCFP	Convective SIGMETs	GFS 0.25"/day	RUC visibility
2003	0.665	0.547	0.462	0.387		
2004	0.677	0.562	0.449	0.413		
2005	0.683	0.601	0.471	0.456	0.427	
2006	0.632	0.676	0.443	0.427	0.449	0.467
2007	0.684	0.643	0.521	0.467	0.416	0.429
2008	0.677	0.648	0.537	0.512	0.449	0.486
2009	0.719	0.648	0.575	0.532	0.436	0.443
2010	0.738	0.678	0.608	0.569		0.455
2011	0.734	0.672	0.657	0.551		0.457
2012	0.696	0.683	0.609	0.569		0.443
improvement	0.066	0.125	0.211	0.192	0.016	-0.015
NDFD Predicted (2022)	0.427	0.499	0.552	0.515	0.310	0.270

All of this suggests that the MOC requirements as presently conceived may be set higher than is realistically achievable. A re-assessment of these requirements in light of the practical (i.e., hardware) demands and general advancement of the state of the science they imply is encouraged.

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. Preprints, 12th Conf. 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.
- Gale D. and L. S. Shapley, 1962: College Admissions and the Stability of Marriage, Amer. Math. Monthly, 69, 9-14.
- 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.
- Ghirardelli, J., 2005: An overview of the redeveloped Localized Area MOS Program (LAMP) for short-range forecasting. Preprints, 21st Conference on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction, Washington, D.C
- Lack, S.A., A.F. Lough, M.P. Kay, M.A. Petty, and J.L. Mahoney, 2012: The NDFD thunderstorm field: An assessment of forecast lead time and spatial displacement. 38 pp.
- 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
- Mackerras, D., M. Darveniza, R.E. Orville, E.R. Williams, and S.J. Goodman, 2012: Global Lightning: Total, cloud and ground flash estimates. *J. Geophys. Res.*, **103**, 19791-19809.
- Weygandt, S., T. Smirnova, C. Alexander, S. Benjamin, K. Brundage, B. Jamison, and S. Sahm, 2010: The High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2009 convective season. Preprints, 14th Conference on Aviation, Range and Aerospace Meteorology (ARAM), Atlanta, GA.

Appendix A – Jet Routes Used for Assessment

Jet Routes that cross AFP 05 (FCA 05)	Jet Routes that cross AFP 08 (FCA 08)	Jet Routes that do not cross AFP 05 or 08 (but are within AFP boundaries)
J16	J121	J49
J29	J209	J63
J547	J79	J95
J94	J134	Q480
Q42	J149	J225
J110	J193	J211
J584	J42	J222
J91	J61	J518
J43	J53	J190
J85	J174	J162
J36	J48	
J60	J6	
J80	J75	
J82	J51	
J64		
J70		
J34		
J30		
J146		