NOAA Technical Memorandum OAR GSD-36

FORECAST ASSESSMENT FOR THE NEW YORK 2008 CONVECTIVE WEATHER PROJECT

S. Madine
M.P. Kay
S.A. Lack
G.J. Layne
J. Luppens Mahoney

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
August 2009

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR GSD-36

FORECAST ASSESSMENT FOR THE NEW YORK 2008 CONVECTIVE WEATHER PROJECT

Sean Madine
Michael P. Kay
Steven A. Lack
Geary J. Layne
Jennifer Luppens Mahoney

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
August 2009

UNITED STATES DEPARTMENT OF COMMERCE

Gary Locke
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Dr. Jane Lubchenco
Under Secretary for Oceans and Atmosphere/Administrator

Office of Oceanic and Atmospheric Research

Dr. Richard Spinrad
Assistant Administrator

Contents

Forecast Assessment for the New York 2008 Convective Weather Project

Quality Assessment Product Development Team
16 December 2008

Executive Summary

This study assesses the performance of the Enhanced Convective Forecast (ECF) and its associated model product, the Weather Research and Forecasting model (WRF) composite reflectivity, both of which were used as experimental forecasts during the New York 2008 Convective Weather Project. The evaluation compares the performance of the ECF and WRF directly against that of the Collaborative Convective Forecast Product (CCFP), which is the current operational baseline. Additionally, the analysis evaluates the quality of the experimental products when used to supplement CCFP as outlined in the project's concept of use.

Accounting for resolution and domain differences between the products, the study adopts two approaches for the direct comparison of the ECF and WRF to the operational baseline. The first, a sector-based method, scores the products on a "grid" of air traffic sectors, allowing for displacement of the forecast on a user-relevant scale. The second, a resolution-based approach, measures performance on a number of regularly spaced grids, ranging from 48 km down to the native resolution of the observation field, 4 km.

The analysis of the supplemental relationship of ECF and WRF with respect to CCFP relies on a decomposition approach that identifies agreement and disagreement between the forecasts. Within each subdomain, the supplemental forecasts are characterized with various measures (e.g., structure of convection) to determine the value added to CCFP.

Results from the direct comparison of the products indicate the following.

- By all measures in the sector-based approach, ECF performed poorly relative to the CCFP baseline, clearly due to the small size and limited number of individual ECF polygons within any given forecast. Overall, the WRF and CCFP showed similar skill (as measured by the Heidke Skill Score), but the forecasts have different strengths and weaknesses. The WRF forecast often misplaced convection relative to its occurrence, but effectively predicted the correct number of sectors significantly covered by hazardous weather. CCFP, while overforecasting, shows value by locating most of the sectors with significant coverage.
- As in the case of the 4- and 6-h leads, the ECF failed to show any appreciable skill in forecasting convective coverage of air traffic sectors at 8-h and 10-h leads. The WRF forecast, however, appears to retain some skill, which degrades by a

factor of two at these longer lead times. Used cautiously, the longer lead times of the WRF forecast may provide utility to planners.

- In addition to confirming the overall indications found in the sector-based approach, the resolution-based analysis suggests that the ECF and WRF products don't perform well at high resolution. Even at the most coarse resolution (48 km) studied, median performance values are very low.

Results from the analysis of the supplemental relationship indicate the following.

- When the forecasts agree on the presence of convection: The WRF simulated reflectivity appears to add more value than ECF by better forecasting the amount of hazardous convection in and around a CCFP polygon. Additionally, the WRF seems to better indicate the structure of convection within a CCFP polygon, as measured by the "center of mass" of the convection, and the distributions of the sizes and shapes of convective objects.
- When ECF and WRF forecast convection outside of CCFP polygons: Overall, neither supplemental forecast seems to effectively identify significant convection outside of a CCFP polygon. Additionally, in most cases, these regions appear to contain isolated convection that may not meet minimum CCFP criteria to warrant a polygon. More research would need to be done to completely quantify this observation.
- When only CCFP forecasts convection: Rarely – in less than 1% of the area of the domain -- does a CCFP polygon exist without some associated ECF or WRF forecast of convection. These cases almost always contain only "clipped" CCFP polygons that are found along the edge of the verification domain, likely a result of the difference in the product domains and granularity.
- When CCFP and the supplementary forecast agree on a forecast of 'no convection': The forecast combinations appear to accurately predict 'no convection.' For forecasts issued in the study period, the combined regions of 'no convection' have median convective coverage of less than 1%. Rarely did both CCFP and its supplement miss significant convection; together, in this subdomain, they appear very trustworthy for use in air traffic planning.

1. Introduction

Motivated to address the high level of weather-related delays experienced in the New York Region, the FAA System Operations Programs sponsored a study to improve the weather forecasting capabilities available to the air traffic planning community (Phaneuf 2008). The effort, called the New York 2008 Convective Weather Project, was conducted from June through August of 2008.

This study assesses the performance of the Enhanced Convective Forecast (ECF) and its associated model product, the WRF Composite Reflectivity, both of which were used as experimental forecasts during the New York study. The evaluation compares the performance of the ECF and WRF directly against that of the Collaborative Convective Forecast Product (CCFP), which is the current operational baseline. Additionally, the analysis evaluates the quality of the experimental products when used to supplement CCFP as outlined in the project's concept of use.

Accounting for resolution and domain differences between the products, the study adopts two approaches for the direct comparison of the ECF and WRF to the operational baseline. The first, a sector-based method, scores the products on a "grid" of air traffic sectors, allowing for displacement of the forecast on a user-relevant scale. The second, a resolution-based approach, measures performance on a number of regularly spaced grids, ranging from 48 km down to the native resolution of the observation field, 4 km.

The analysis of the supplemental relationship of ECF and WRF with respect to CCFP relies on a decomposition approach that identifies agreement and disagreement between the forecasts. Within each subdomain, the supplemental forecasts are characterized with various measures (e.g., structure of convection) to determine the value added to CCFP.

2. Data

This section describes the forecasts and observations used in this study. Temporal aspects of the forecasts and observations for verification purposes are discussed in section 3.2. Data thresholds are discussed in section 3.3.

Collaborative Convective Forecast Product (CCFP). This forecast is the official convective forecast product to be used for strategic planning purposes by the FAA's Air Traffic Control System Command Center (ATCSCC). The CCFP is a human-generated forecast. The goal of the forecast is to capture areas of significant hazard over the continental United States (CONUS) that may result in disruptions to traffic within the National Airspace System (NAS). Thus, the CCFP does not attempt to capture any and all convection that may occur. The forecast takes the form of a set of polygons, each of which has a set of attributes describing the expected amount of convection within the polygon, the expected maximum echo top heights, as well as forecaster confidence that the polygon will meet 'CCFP minimum criteria'. In this study, only the coverage attribute will be used in the verification. More information about the CCFP can be found in the CCFP product description document (NWS, date unknown).

Enhanced Convective Forecast (ECF). The ECF is an experimental forecast produced by forecasters at ENSCO, Inc. The stated goal of the ECF is to provide forecasts of any and all convection within the New York TRACON area (FAA, 2007). The forecast is characterized by polygons that take the form of nested ranges of equivalent radar reflectivity factor values representing low-, mid-, and high-intensity regions of convection. The ECF product also provides a graphical depiction of storm motion vectors along with maximum echo top heights. These two additional attributes were not available in a verifiable format and will not be discussed further. Like the CCFP, the ECF is a human-generated forecast product.

Weather Research and Forecast (WRF) Model Simulated Composite Radar Reflectivity. In support of the ECF forecasts, a high-resolution (3-km grid spacing) version of the WRF model was run by ENSCO, Inc. over the NY TRACON region to provide a consistent forecast from which the ECF could be derived. For the analyses within this document, only the simulated composite radar reflectivity output field was verified.

National Convective Weather Detection Field (NCWD). Because each of the forecasts in this evaluation are based upon the spatial location of a convective hazard, the choice was made to use the NCWD as the observation that the forecasts were compared against rather than an echo tops product. The NCWD is an operational product that uses vertically integrated liquid (VIL) data, along with cloud-to-ground lightning data, to provide a high spatial resolution (4-km) hazard detection field covering CONUS that is relevant to the needs of the aviation community (Megenhardt et al. 2004).

3. Methodology

This section describes the configuration of the verification exercise. The fundamental attributes considered include the spatial domain, how the data are matched temporally, and the choice of data thresholds used to define forecast and observed events. The verification techniques including the sector-based and resolution analysis are then introduced. The domain decomposition is then detailed, which provides information about the ECF and WRF forecasts relative to CCFP. Finally, the statistics used to quantify the forecast quality and behavior are identified.

3.1. Spatial domain

The forecast domain for the NY TRACON study, described in the concept of use (ConUse) document (FAA 2008), is shown in Fig. 1. This domain includes the NY TRACON itself as well as surrounding areas, with a bias towards the west to account for the importance of the airspace between New York and Chicago. A climatological analysis performed on the set of ECF forecasts available for the study dates indicated that ECF forecasts were not being produced for the entire original domain found in the ConUse document. ECF forecasts were found to occur within a subset of the WRF model domain. Therefore, the study domain has been confined to the intersection of where ECF forecasts occurred at least once with the underlying WRF model domain.

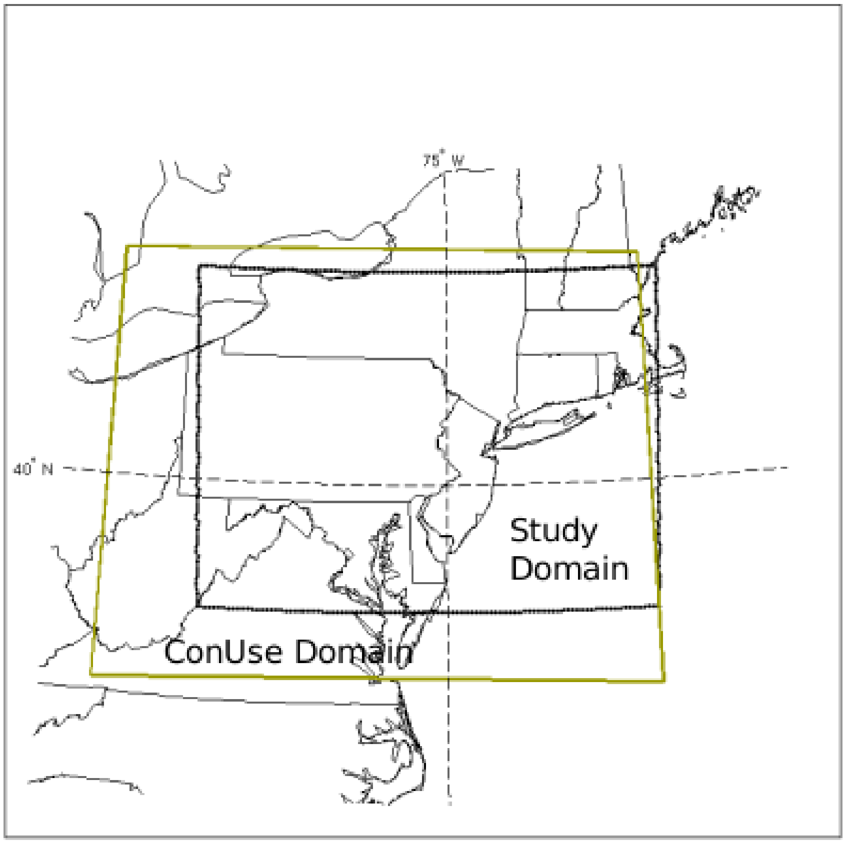

Figure 1. Study domain is shown by the black polygon. Original domain from the ConUse document is also shown in gold.

3.2. Temporal matching

The following terms are used to describe the temporal components of the forecasts and observations. A forecast is said to have an *issue time* that describes the nominal time that a forecast is produced, or issued. For the WRF model, this time refers to the model's run time, and does not include the amount of time it takes to produce the simulation and make results available (i.e. latency.) Each forecast issuance includes a variety of *valid times* in the future. The difference in time between a product's issue time and valid time is called the *lead time*.

The ECF product is issued every other hour beginning at 11 UTC and ending at 21 UTC, with each issuance providing hourly lead times from 2 to 11 hours. The CCFP forecasts are available for the same issue times as the ECF, but only contain lead times of two, four, and six hours. As a primary input to the ECF, the WRF model must run in advance in order to be available as input to the ECF creation process. The WRF model was run at 6, 9, 12, 15, and 18 UTC daily with hourly lead times available out to 15 hours.

In order to align the WRF and ECF temporally for a direct comparison at specific lead times, the study utilizes a mapping based on the issuance relationship between the two products (Table 1). The mapping adjusts the WRF lead times from the actual value (valid time minus issuance time) to an effective value, typically two hours later than actual. This alignment, designed to account for the latency of the WRF information, provides an operationally meaningful comparison by only comparing data available to the human forecaster at the ECF issuance time.

Table 1. WRF model issue times (UTC) used for each ECF issuance.

ECF	WRF
11	9
13	9
15	12
17	12
19	15
21	18

While including all issuance and lead times, this study primarily focuses on forecasts with issue times of 11, 13, and 15 UTC, and lead times beyond two hours, due to their importance in strategic air traffic management. The climatology of convective coverage of the domain (Fig. 2) supports this emphasis; initiation and subsequent storm coverage occur after approximately 18Z, coinciding with a period of high demand in the NY TRACON airspace.

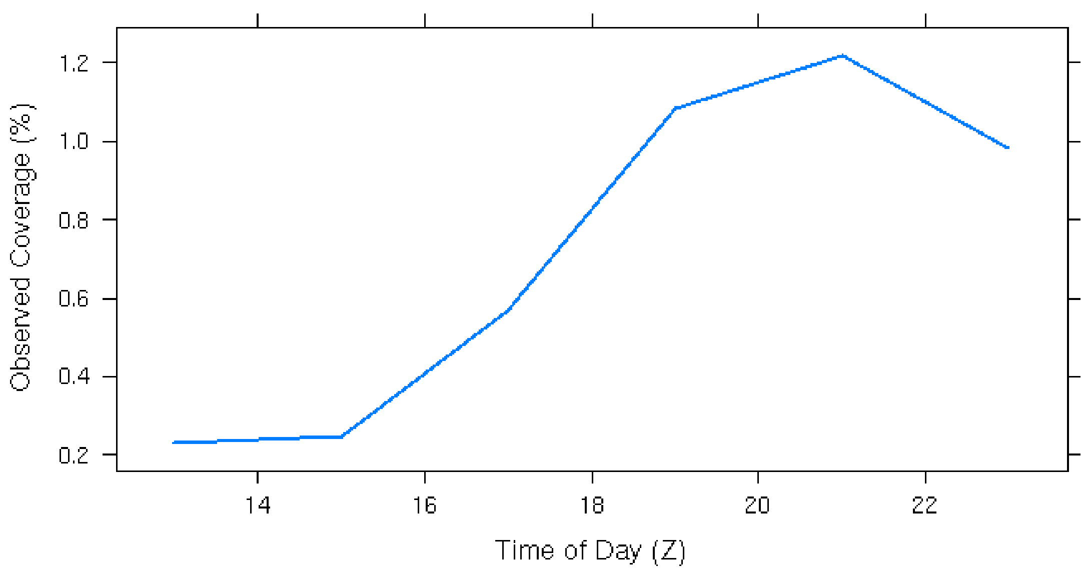

Figure 2. Mean climatological coverage amounts of VIP 3 or greater intensity convection within the study domain as a function of time of day.

3.3. Data thresholds

This verification exercise assigned thresholds to focus on hazardous convection: any convection at greater than or equal to a VIP 3 intensity in the NCWD field, the ECF yellow and red fields (levels 2 and 3), and WRF simulated composite reflectivity at 40 dBZ or higher (Table 2).

CCFP, a forecast of spatial coverage, is treated in two ways in the analysis. In the resolution-based approach (section 4.2), the forecast is converted to a dichotomous field, with any spatial coverage amount representing convection ('Sparse' or higher coverage category). In the sector portion of the study, however, the CCFP forecast at a given point is interpreted as the low end of the range of the spatial coverage valid at that point. For instance, Sparse-category polygons, which are forecasts of 25-49% coverage are assigned values of 25%. This is used to define the 'porous CCFP' which is used for determining events within sectors (section 4.1). Confidence categories of CCFP (low vs. high) are not incorporated into the percent coverage, and no polygons (including Sparse coverage, low confidence) are removed from this study.

Table 2. Verification thresholds used for the different products in this study. The verification thresholds represent significant convective weather.

Meteorological Product	Verification Threshold
National Convective Weather Diagnostic (NCWD)	VIP level 3 or greater
Enhanced Convective Forecast (ECF)	Yellow (2) and/or Red (3) levels (40dBZ or greater)
WRF Simulated Composite Reflectivity (WRF)	40 dBZ or greater

3.4. Verification approaches

The three specific approaches that were used to provide objective measures of forecast performance are described below. For all analyses, the forecasts and observations were mapped onto the 4-km NCWD grid, eliminating the need for any interpolation of the observation field. The ECF and CCFP forecast polygons were mapped directly to this grid while the simulated reflectivity field from the WRF model was interpolated to the grid using bilinear interpolation.

3.4.1. Sector-based verification

Strategic planners require convective weather information, specifically the amount and overall pattern of the hazard, to help them determine the air traffic capacity of large-scale regions (e.g., TRACON) of the airspace. High-resolution pixel-to-pixel verification approaches seemingly fail to measure the utility of forecasts, primarily because of the strict requirement that the forecast locate convection exactly to get any positive credit. To account for the shortcomings of the pixel-to-pixel method, the sector-based approach adopts a map of air traffic sectors as common grid on which the forecasts and observations are compared. The sectors naturally represent the spatial granularity of air traffic planning, and they tend to capture traffic flow with the shape and orientation. Kay et al. (2007) used sector-based verification to study convective forecasts from the summer of 2007 and found utility in the approach for aviation-related purposes.

With this approach, the forecasts are transformed from regularly spaced grids to a grid of irregularly shaped sectors, with each assigned a percent coverage value. Based on a climatological study of sector coverage for convection in the month of June 2007 (not shown), this analysis defines a sector as having significant coverage if the areal coverage of either the forecast or observed convection exceeds five percent. This threshold was used in Kay et al. (2007) as well.

3.4.2. Resolution analysis

The study utilizes the pixel-to-pixel analysis to retain a traditional view of the performance of the forecasts, but with a simple upscaling (from fine to coarse resolution) technique superimposed Relative to the TRACON domain, the forecast areas within the ECF and WRF predictions are very small, with typical diameters on the order of tens of kilometers. At fine spatial resolutions, forecasts and observations do not directly overlap but may be 'close' to one another. In recent years, a class of verification techniques called fuzzy verification techniques (Ebert, 2008) was developed to deal with the issues of verifying high-resolution gridded forecasts. With these techniques, neighborhoods of points are used to characterize forecast and observed regions throughout the domain in contrast to traditional techniques which consider results at each forecast grid point. In this study we use a simple upscaling (from fine to coarse resolution) technique to produce verification statistics for a variety of resolutions. The most liberal technique for upscaling is applied – if any one of the points within the neighborhood is an event (i.e., forecast of the hazard or observation of the hazard), then the event is said to have occurred for that neighborhood. The resolutions that are analyzed are 4 km, 16 km, 32 km, and 48 km. The coarsest scale, 48 km, was chosen to coincide with the approximately 50 km resolution pixels used by NASA SPoRT for their 'quick look' verification. (Caveat: The ‘quick look’ approach uses a window of plus or minus one pixel for comparison, effectively evaluating at a 150km resolution, according to the community literature.) The spatial scales chosen for study can be put into context by considering the approximate relative sizes of the study domains and forecasts (Fig. 3). In Fig. 3, the average areal extent of CCFP forecasts within the study domain is shown. CCFP coverage is typically a large fraction of the study domain. In contrast to the CCFP, the ECF forecasts are typically very small. The ECF sizes are discussed further in section 3.4.3. It is also evident that even at the relatively course resolution of 48 km the study domain is still covered by a large number of grid boxes.

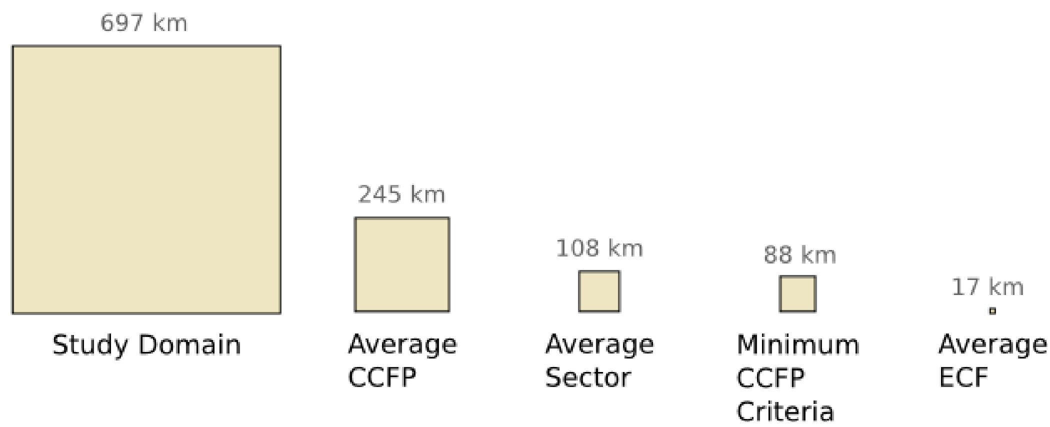

Figure 3. Comparison of average sizes of forecasts and domains used in this study.

3.4.3. ECF and WRF as supplements to CCFP

To evaluate the deterministic products (ECF and WRF) as supplemental to CCFP, the joint probability distribution (JPD), which shows forecast agreement and disagreement, is used. The JPD, in this case, does not include the NCWD observation. The NCWD observation is only used in forecast verification after forecast agreement and disagreement are determined.

Forecast agreement is indicated when any ECF or WRF (supplemental) forecast falls within a CCFP polygon or where there is no ECF or WRF forecast and no CCFP polygon. Similarly, forecast disagreement occurs when CCFP is issued without collocation of ECF or WRF, or when ECF or WRF is issued without collocation of CCFP. The subdomains of disagreement and agreement are set up using rules outlined in the following section. Once these subdomains are set up and NCWD observations are added, a variety of verification metrics can be used in each decomposition region to assess the supplementary nature of the ECF and WRF simulated reflectivity products to the CCFP.

Domain decomposition

There are four subdomains of interest in the decomposition. The YY subdomain is where there ECF or WRF are present within a CCFP polygon. The CCFP polygon basically becomes the bound of the YY subdomain if any ECF or WRF lies within the CCFP polygon. The YN subdomain occurs when there is ECF and/or WRF but without any CCFP polygon issuance. The NY subdomain consists of situations when there is a CCFP polygon and no supplemental (ECF or WRF) forecast. The NN subdomain is simply no forecast issuance from CCFP or the supplemental products. The NCWD observation is added after the forecast domain decomposition. Note that for this exercise we are thresholding the forecast and observation to the definition of convection from the CCFP baseline. That is any convection at greater than or equal to a VIP 3 intensity in the NCWD field, the ECF yellow and red field (2 and 3), and WRF simulated reflectivity at 40 dBZ or higher. The supplemental forecasts (ECF and WRF) and NCWD observations in this study are treated as objects. Objects are defined as regions of continuous pixel coverage after a threshold at the VIP 3 or equivalent level has been applied to the image. The rules for each decomposition subdomain follow.

YY subdomain

The YY subdomain is given top priority in the decomposition. This means that all NCWD observations and supplementary forecasts are considered for this subdomain first before testing for inclusion into the other subdomain types (i.e., YN, NY, NN). First, all CCFP in the TRACON verification domain are dilated by a factor of 32 km. This factor was used because it also serves as the dilation parameter for ECF and is roughly the median diameter of an ECF polygon when no intensity threshold is applied (Fig. 4). Once all CCFP polygons are dilated, any and all ECF or WRF objects that are fully contained within the dilated CCFP are included in the YY subdomain. Additionally, ECF or WRF objects are considered within the YY subdomain if at least 50% or 30 pixels of the object are within the CCFP dilation area. If an ECF or WRF object is a boundary case and meets the 50% or 30 pixel criteria it is included in the YY subdomain and is dilated using the 32-km buffer. The YY subdomain is tested once again for boundary objects (ECF or WRF). If there are new objects that meet the 50% or 30-pixel criteria the entire object is included; however, no buffer is applied and the YY subdomain reaches its fullest extent. Once the YY subdomain is determined, observations are overlaid and included in the YY subdomain if there is intersection. Additionally, intersection of observations with the YY subdomain that meet the 50% or 30-pixel criteria are included as YY observations with no additional buffer applied.

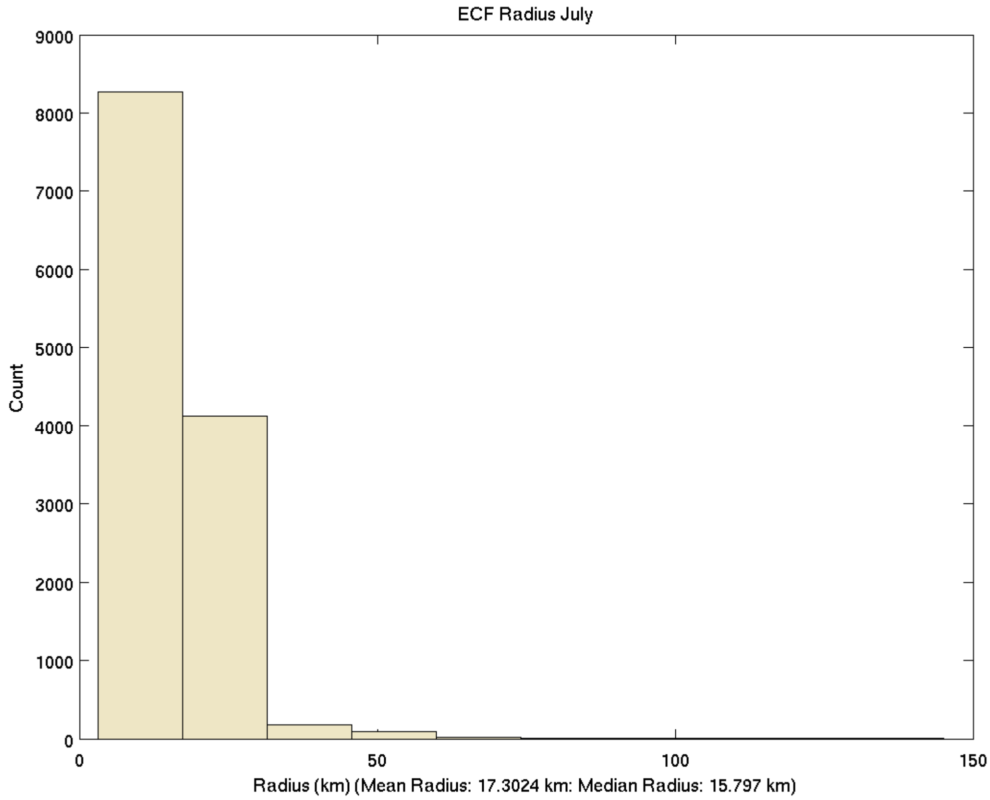

Figure 4. The distribution of ECF radiuses (km) for the month of July 2007; similar results can be seen for June 2007.

YN subdomain

The YN subdomain (ECF or WRF without a CCFP polygon) is given the next priority. Once the YY subdomain is established, the remaining ECF and WRF objects are considered. The remaining objects are dilated using the 32-km buffer and where there are intersections between buffered objects, an additional dilation is performed coupled with an erosion of the same size to smooth out the YN regions. This effectively couples neighboring ECF or WRF buffered objects to give some hint of added structure outside of the YY subdomain. Observations are coupled with the YN subdomain similarly to the YY methodology.

NY and NN subdomains

The NY subdomain (CCFP without an ECF or WRF object) is given the last priority as the NN subdomain is simply the residual of the entire TRACON verification subdomain that is neither YY,YN, or NY. The CCFP polygons that are dilated with the 32-km buffer without a supplementary forecast are simply left as the NY subdomain. These subdomains are normally quite small as will be seen in the resulting JPDs later on. Observations are coupled with the NY subdomain similarly to the YY and YN methodology. An example of the domain decomposition is shown in Fig. 5 (ECF) and Fig. 6 (WRF).

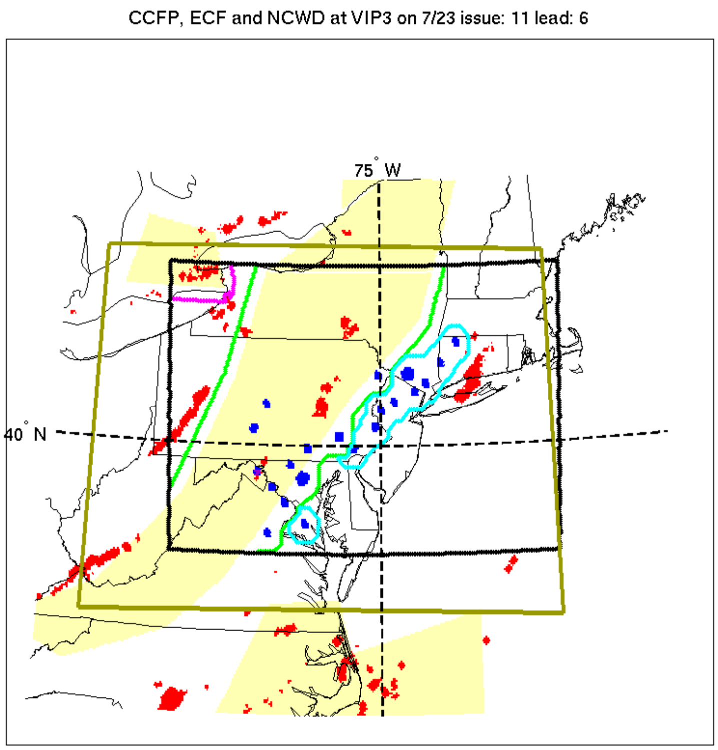

Fig. 5. CCFP/ECF domain decomposition on 23 July 2007 issued at 11 UTC valid at 17 UTC. Blue objects represent ECF yellow/red regions only. Red objects represent NCWD observations at VIP level 3 or greater. The yellow regions indicate CCFP (sparse coverage/ low confidence). Green outlines the YY domain. Cyan outlines the YN domain. Magenta outlines the NY domain. The remaining region is the NN domain. The black outline represents the verification domain.

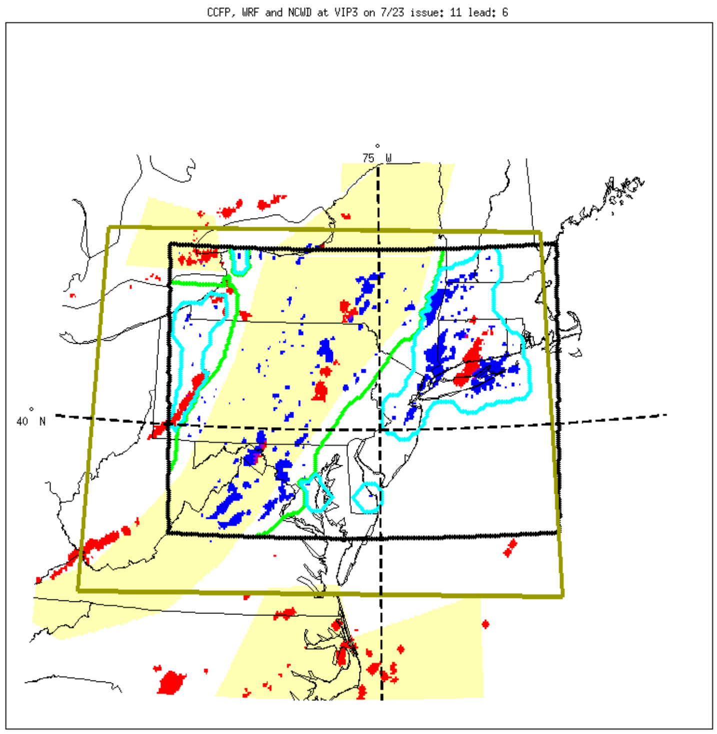

Fig. 6. CCFP/WRF domain decomposition on 23 July 2007 issued at 11 UTC valid at 17 UTC. Blue objects represent WRF at 40 dBZ or higher. Red objects represent NCWD observations at VIP level 3 or greater. The yellow regions indicate CCFP (sparse coverage/ low confidence). Green outlines the YY domain. Cyan outlines the YN domain. The remaining region is the NN domain. The black outline represents the verification domain.

3.5. Dichotomous verification statistics

In this study a limited subset of statistics were used to describe the important aspects of the forecast quality. The reader is encouraged to consult Wilks (2006) for more information about the scores presented below as well as other scores that may be used for dichotomous events. The four possible event scenarios that comprise the statistical measures in a dichotomous setting are represented by a 2x2 contingency table (Table 3).

Table 3. 2x2 contingency table illustrating the various states that are allowed.

		Observed Yes	Observed No
Forecast	Yes	a	b
	No	c	d

Some of the information relevant to air traffic management that can derived includes the amount of convection captured by the forecasts, the amount of over- or underforecasting, and overall forecast skill. These aspects can be measured by the probability of detection (POD), bias (BIAS), and Heidke Skill Score (HSS), respectively. The statistics are defined and summarized in Table 4.

Table 4. Dichotomous summary statistics used in this study.

Score	Long Name	Definition	Interpretation
POD	Probability of Detection	$POD=\frac{a}{a+c}$	The amount of convection occurring within the forecast areas. Values range from zero to one.
BIAS	Bias	$BIAS=\frac{a+b}{a+c}$	The amount of over- or underforecasting. A bias of 1 means the forecast occurs as much as the observation but says nothing about location. Values range from zero to infinity.
HSS	Heidke Skill Score	$HSS=\frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}$	Skill score where perfect forecasts achieve a score of one, forecasts no better than random score zero, and forecasts worse than random score less than zero.

3.6. Domain decomposition statistics

A number of verification metrics are utilized for evaluating the supplemental nature of both the ECF and the WRF products. The metrics are primarily used to evaluate the value added in terms of structure over or in addition to the CCFP. Different verification metrics are used within the different domains and will be described further.

3.6.1. YY subdomain

In the YY subdomain, we want to test if structure is added within a CCFP polygon. Structure can be defined and evaluated in two ways, the amount of supplementary forecast compared to the observation in the subdomain and additionally some measure of placement of the mass within the subdomain. The first of these is simply measured by the bias (bias=Forecast Area/Observation Area; as in Table 4) of both the CCFP to the observation and the supplementary (ECF or WRF) forecast to the observation. In this case, CCFP's forecast area is defined as 25% for sparse coverage, 50% for medium coverage and 75% for solid coverage ignoring confidence (i.e., the low values assigned to each CCFP coverage ranges). Second, placement is measured by a center of mass differencing equation. Inside the YY subdomain, the center of mass is calculated by finding the weighted average of the centroids of the objects of interest. The CCFP center of mass normally comprises of simply finding the centroid of the one CCFP polygon used to make up the YY subdomain and weighting by the percent coverage value if more than one CCFP polygon is present with differing coverage values. Second, the supplementary forecast center of mass is found in the YY subdomain by weighting the centroids by the forecast object sizes. Finally, the observation center of mass within the YY subdomain is calculated using the same size weighting technique as above. Differences are then calculated between CCFP and the supplementary product, CCFP and the observation, and the supplementary product and the observation. In this case, one hopes that the CCFP and supplementary product are significantly different and that the supplementary product is in fact closer to the observation than the CCFP center of mass. An example plot of the centers of mass for ECF, NCWD, and CCFP is shown in Fig. 7. The combination of bias within the YY subdomain and the center of mass differences provides a good baseline for determining if structure is added with the supplementary forecast product of ECF and WRF simulated reflectivity. Additionally, information on the size distribution of objects within the YY subdomain can be inferred from the mean and median supplementary forecast and observed object sizes.

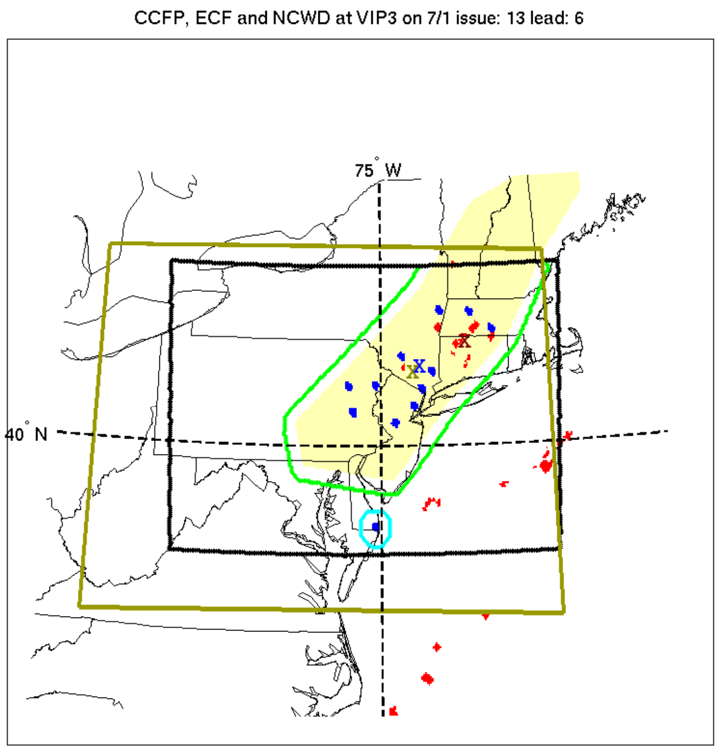

Fig. 7. Center of mass locations marked with a gold X for CCFP, blue X for ECF, and maroon X for NCWD. Note in this case the center of mass for ECF and CCFP are nearly the same.

3.6.2. YN subdomain

In the YN subdomain, the usefulness of identifying convection by the supplementary forecast product outside of a CCFP region is studied. There tend to be multiple YN regions inside the verification domain. Thus, hits and false alarms are key behaviors that are examined in this domain. For example, if there is a YN region with any non-zero amount of convection, the region is then considered a 'hit'. Furthermore, if the region is a hit, a bias measure is associated with it to yield understanding on the quality of the hit. Note that a hit requires that the observation fell within the 32-km buffer zone. Conversely, if a YN region contains no observation it is considered a false alarm. The size of the YN regions that score a hit is compared to the size of YN regions that score a false alarm. The mean and median biases can then be calculated for the combination of YN regions that hit or false alarm or biases may be calculated on just those regions that would be deemed a hit. Additionally, like the YY subdomain, information of the size distribution of objects with the YN subdomain can be inferred from the mean and median supplementary forecast and observed object sizes. The combination of these hits and false alarms with the distribution of object sizes can yield results on the quality of the forecast outside of CCFP (YY) regions.

3.6.3. NY subdomain

Within the NY subdomain, the characterization of the observation field and the area weighted ratio between hits and false alarms are of primary interest. However, after looking at the JPD summary for the NY subdomain it is determined that the NY subdomain primarily consists of CCFP polygons only partially overlapping the TRACON verification domain. Therefore, it may not be meaningful to assess the quality of the CCFP forecast where no ECF or WRF forecast is present. These findings will be briefly summarized in section 4.3.4.

3.6.4. NN subdomain

The only verification metrics of interest within the NN subdomain are those that characterize the size, shape, and coverage area of the observation field. Forecasts are considered perfect in the NN subdomain if there is no observation present. When the characterization of the observation field in the NN subdomain is used in conjunction with statistics from the YN and YY subdomains, additional information may be gleaned.

4. Results

Results are summarized below for each of the verification approaches. Each section represents an analysis aimed at providing a different view of forecast performance. The sections aim to complement one another by providing a more complete view of behavior and quality than could be achieved by any one technique individually.

4.1. Sector-based verification

This section presents results using the sector-based grid. Section 4.1.1 focuses on results for strategic lead times while Section 4.1.2 analyzes the performance of ECF and WRF for longer lead times than are available for the CCFP.

4.1.1. Comparison of CCFP, ECF, and WRF forecasts for strategic times

In this section, forecast performance is summarized for the forecasts using the sector-based verification method described in section 3.4.1. For this part of the analysis, we focus on the strategic forecast times – forecasts issued at 11, 13, and 15 UTC having lead times of four hours or greater. Both dichotomous and porous interpretations of CCFP are considered.

Overall, the Heidke skill score (HSS) depicts the CCFP and WRF simulated reflectivity product significantly outperforming the ECF (Fig. 8). Though none of the forecasts perform particularly well in an absolute sense, median scores for CCFP and WRF are near 0.14, whereas the forecast score for ECF shows virtually no skill.

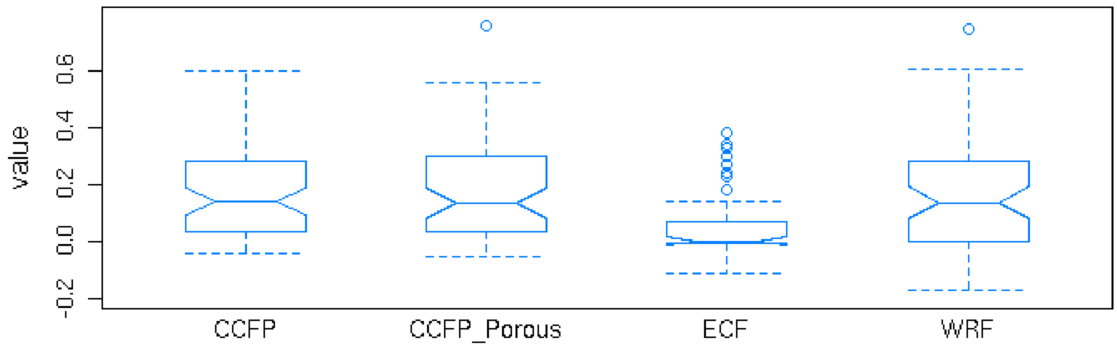

Figure 8. Boxplot of the distribution of Heidke Skill Score (HSS) values for all individual forecasts within the study period for the CCFP, CCFP Porous, ECF, and WRF products verified using the sector grid for issuance times 11, 13, and 15 UTC with lead times 4 h, and 6 h. Notches indicate approximate 95% confidence interval about the median.

Indicative of significant underforecasting, the ECF forecast has a much lower bias, 0.15, than any of the other forecasts studied (Fig. 9). Forecasting small amounts of hazardous convection, the ECF product rarely covers more than 5% of any sector with weather, resulting in a very low measure of skill. Overall, the CCFP forecast has a bias of about 3.0, nearly twice that of the WRF. The WRF reflectivity has a bias of 1.2, indicating that the model does a very good job of predicting the number of sectors with significant convective coverage.

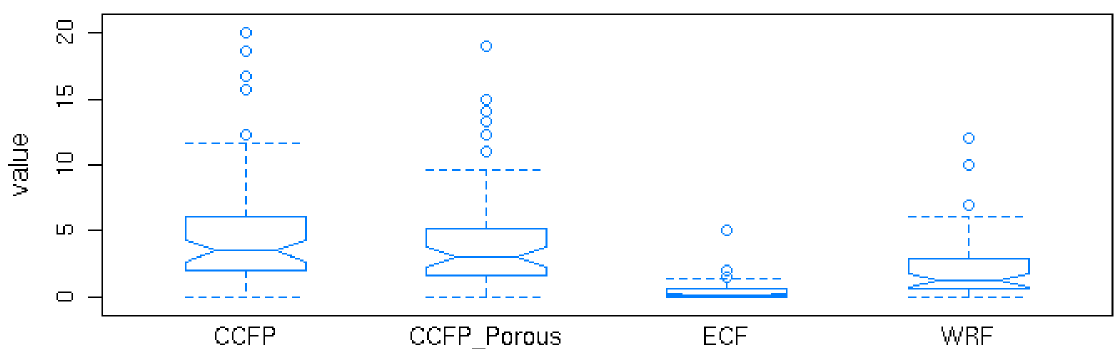

Figure 9. As in Fig. 8, but for BIAS.

Fig. 10 shows the probability of detection (POD) for each of the products in the study. Covering large areas of the domain with polygons, CCFP has a very good POD of ~0.7. The WRF, while approximately forecasting the correct number of covered sectors, has a POD of ~0.25, indicating that the forecast doesn't effectively locate convection at the sector scale. ECF, because of its very low bias, has a POD of virtually zero for a sector threshold of 5%.

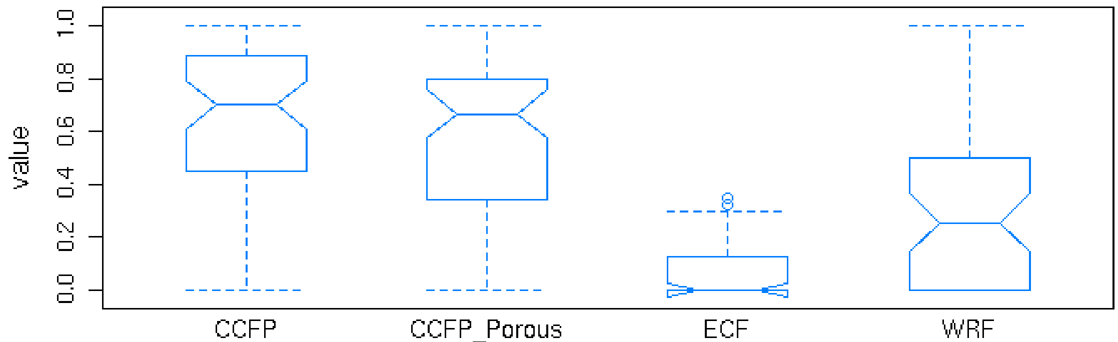

Figure 10. As in Fig. 8, but for POD.

The preceding discussion reveals a significant difference in performance amongst the various forecasts. By all measures in the sector-based approach, ECF performed poorly, clearly due to the small size and limited number of individual ECF polygons within any given forecast. Overall, the WRF and CCFP showed similar skill as measured by the HSS, but the forecasts appear to have complimentary strengths and weaknesses. The WRF forecast often misplaced convection relative to its occurrence, but effectively predicted the correct number of sectors significantly covered by hazardous weather. CCFP, while overforecasting, shows value by locating sectors with significant coverage of hazardous convective weather.

Sensitivity to sector coverage threshold

In the preceding results, the sector coverages were thresholded into events/non-events by specifying a 5% coverage threshold. Clearly the verification results will differ if an alternate threshold were chosen to define events. However, despite the changing scores, it is not necessary for the forecast rankings to change as the sector coverage threshold is varied. To examine the sensitivity of verification measures, scores from the sector-based approach were computed for sector coverage thresholds between 1% and 5% (Fig. 11). Most importantly, the relative ranking of the forecasts does not change. Results for the HSS and bias measures, which are not shown here, indicate the same relative ranking. Therefore, while the choice of a different threshold may be considered, the results indicate that the behavior of the ECF, WRF, and CCFP forecasts relative to one another is not affected for the range of sector coverage thresholds that we believe are appropriate to consider.

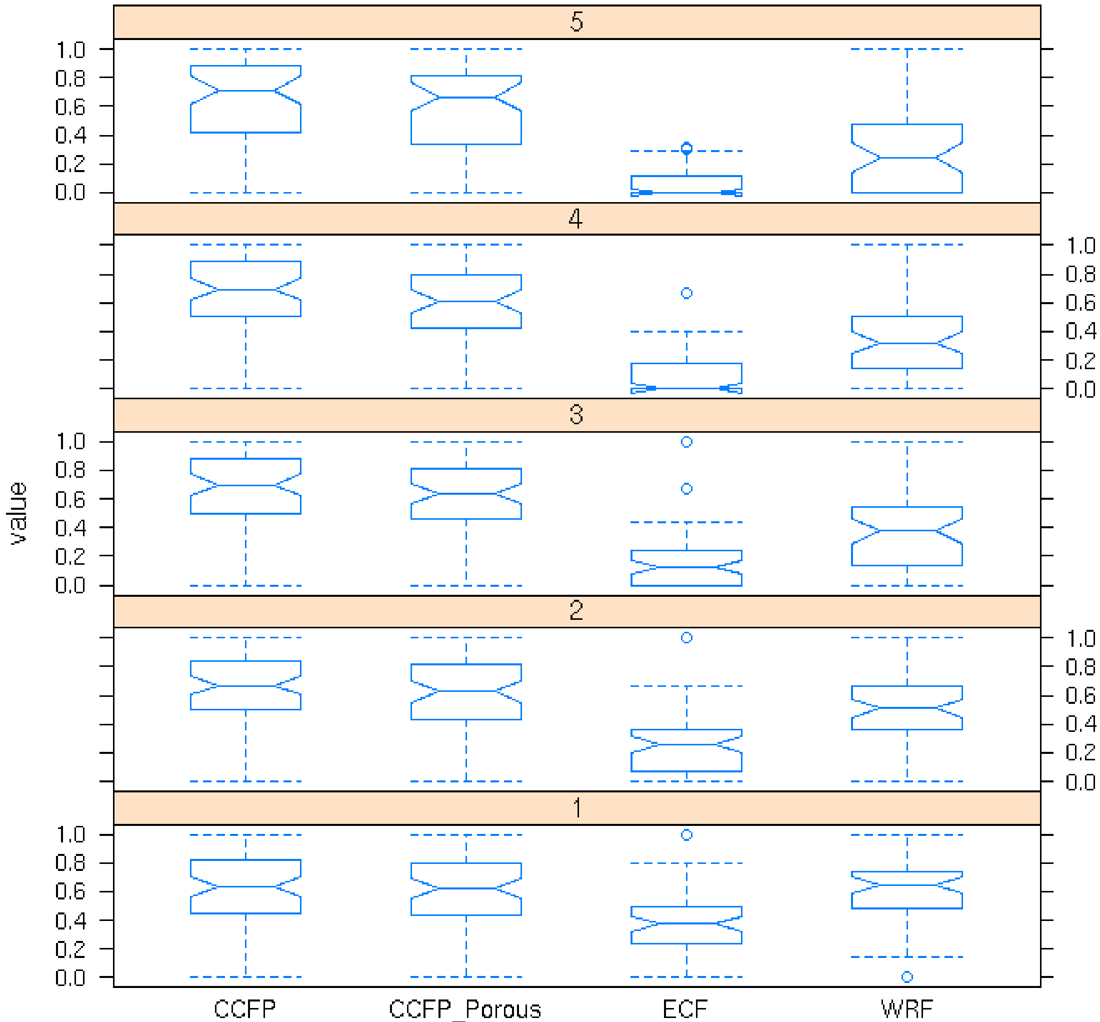

Figure 11. Boxplots of POD as in Fig. 8 for sector coverage thresholds of 1%, 2%, 3%, 4%, and 5%. Title bar above each panel indicates the particular threshold used.

4.1.2. Comparison of ECF and WRF for 8-h and 10-h lead times

Since the CCFP is not available for lead times beyond six hours, the ECF and WRF forecasts were evaluated to determine if they provide useful forecasts at 8- and 10-h lead times for the strategic issuances (Fig. 12). As in the case of the 4- and 6-h leads, the ECF failed to show any appreciable skill. The performance of the WRF forecast, which was quite modest at 4 h and 6 h, degraded by a factor of two for these longer lead times. Used cautiously, the WRF forecast may provide utility to planners.

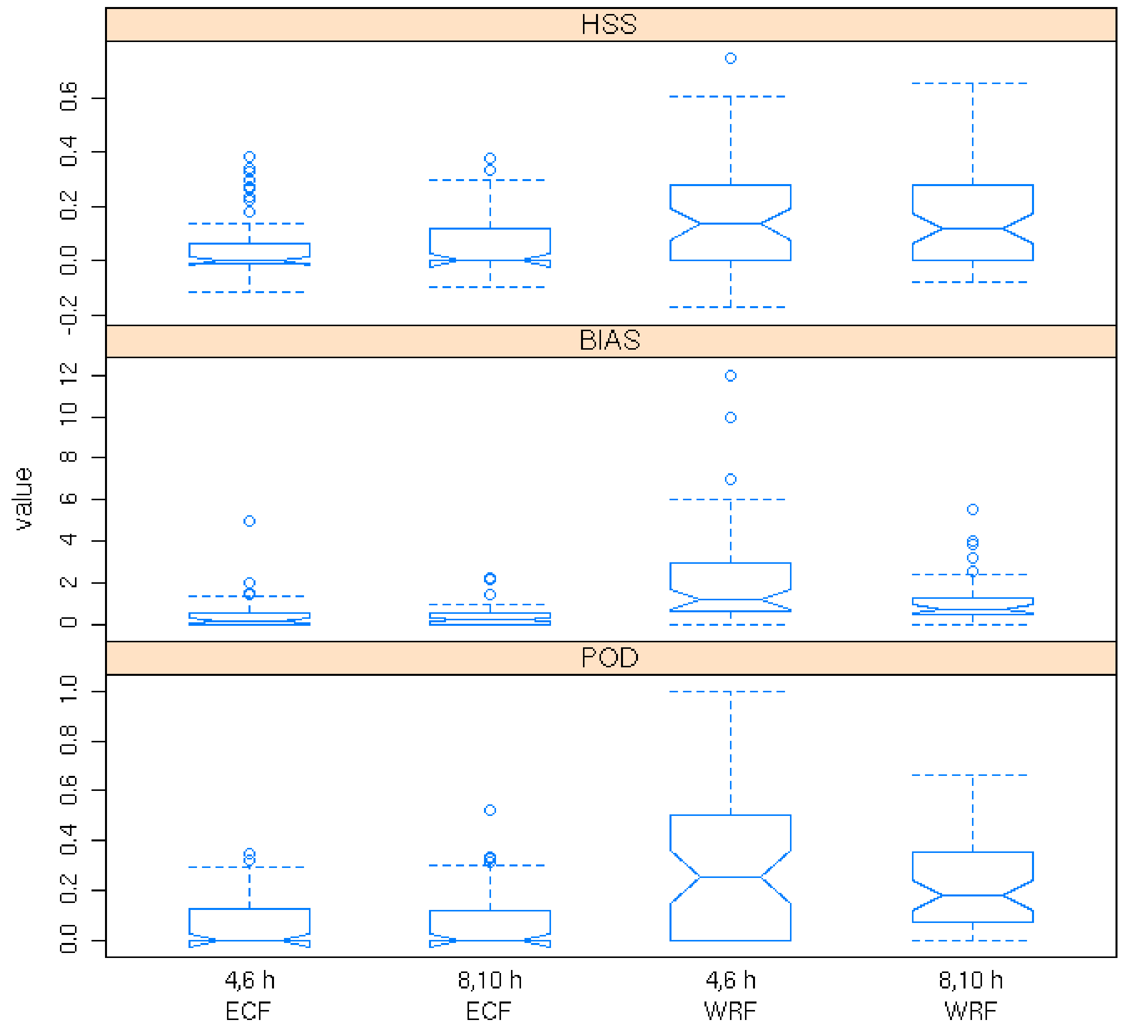

Figure 12. Boxplots of HSS, BIAS, and POD for the 8-h and 10-h lead times for the ECF and WRF forecasts.

4.2. Resolution-based verification

Extending the comparison of ECF and WRF, performance of each forecast is summarized using the resolution analysis described in section 3.4.2. High-resolution forecasts often capture weather effectively, but fail to locate it precisely. Employing a liberal verification scheme, the resolution-based approach upscales the forecasts and observations to produce verification measures at many spatial resolutions, attempting to credit high-resolution forecasts for "near misses."

As in the sector-based approach, the WRF and ECF forecasts are compared for the strategic issuances and lead times. Trends in the results for longer lead times (8 and 10 hours) mirror those found in the previous section for the sector-based verification. In

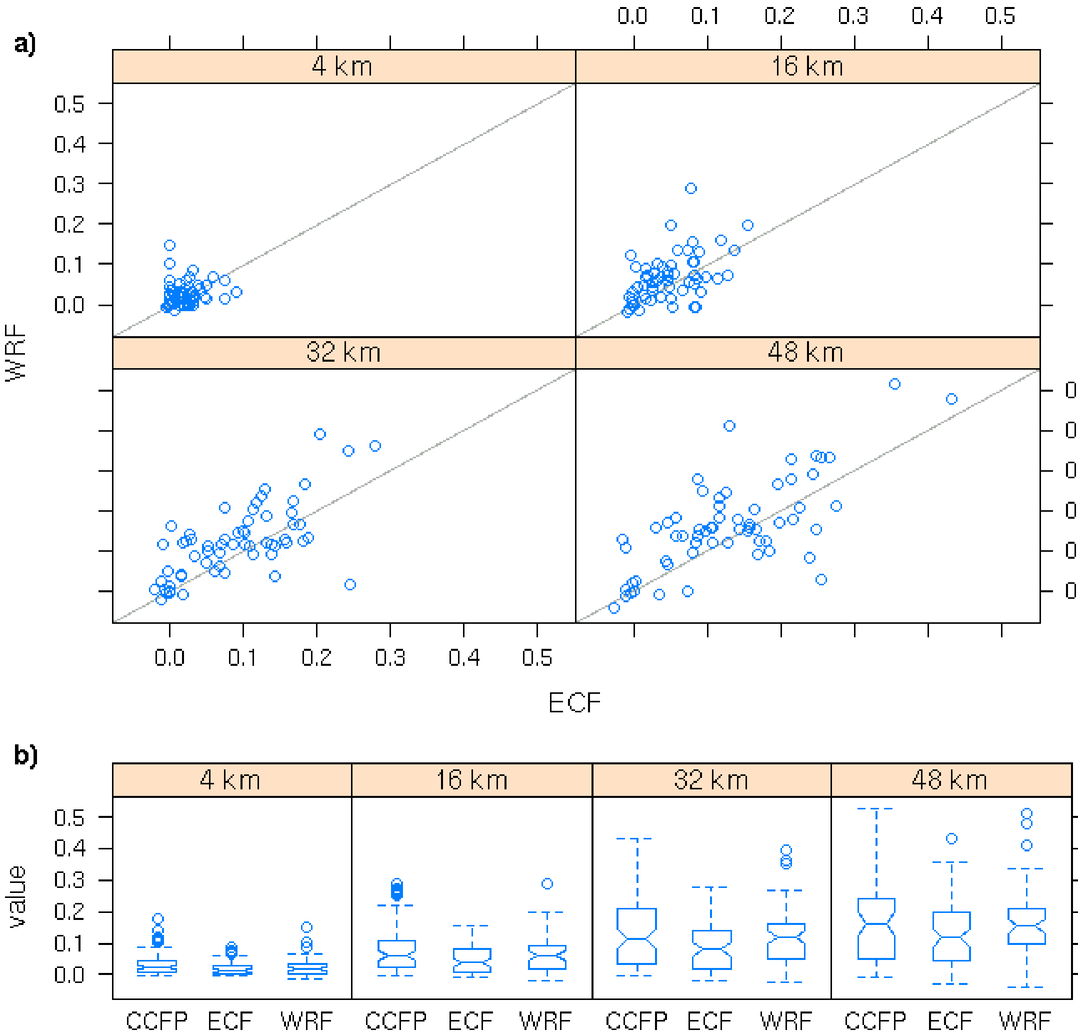

Figure 13. a) Scatterplot matrix of HSS for the WRF and ECF forecasts for verification resolutions of 4 km, 16 km, 32 km, and 48 km. Thin grey line in each panel denotes y=x. b) Boxplots of the distribution of HSS scores from a) along with values of CCFP for comparison for each resolution verified.

terms of overall skill (HSS), the ECF forecasts do not improve upon the WRF forecasts most of the time (Fig. 13 a.). The WRF forecasts, and to some extent the ECF as well, show comparable skill to CCFP but with reduced variability in performance (Fig. 13 b.). At the coarsest resolution (48 km), median performance values are still quite low, with all forecasts having median HSS values below 0.2. As noted previously in section 4.1.1, significant differences in POD and bias underlie the aggregate skill scores for each of the forecast products.

For probability of detection, the WRF model significantly outperforms the ECF forecasts (Fig. 14 a.). Even at the coarse resolutions shown, the ECF maintains much lower POD values than the other forecasts. In terms of bias, the ECF median values show little change as the fields are upscaled to 48-km resolution. In contrast, the WRF biases increased from near 1.25 to approximately 2.25 (Fig. 15 b.). The ECF has a higher bias than the WRF in only a small number of forecasts (Fig 15 a.). The WRF biases increase because of the more scattered nature of the composite reflectivity forecasts compared to the ECF. The ECF forecasts often show biases below one indicating underforecasting even for the coarse resolutions.

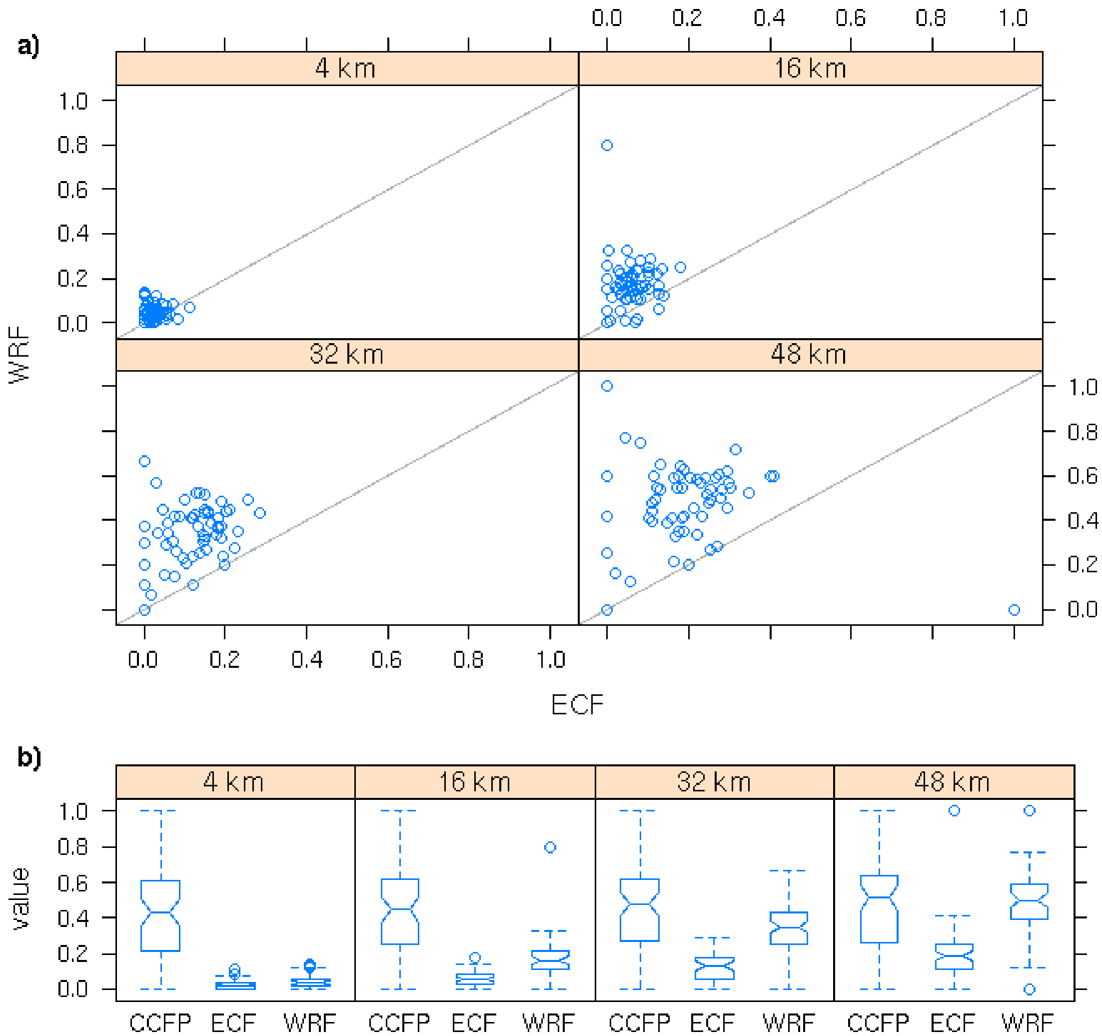

Figure 14. As in Fig. 13, but for POD.

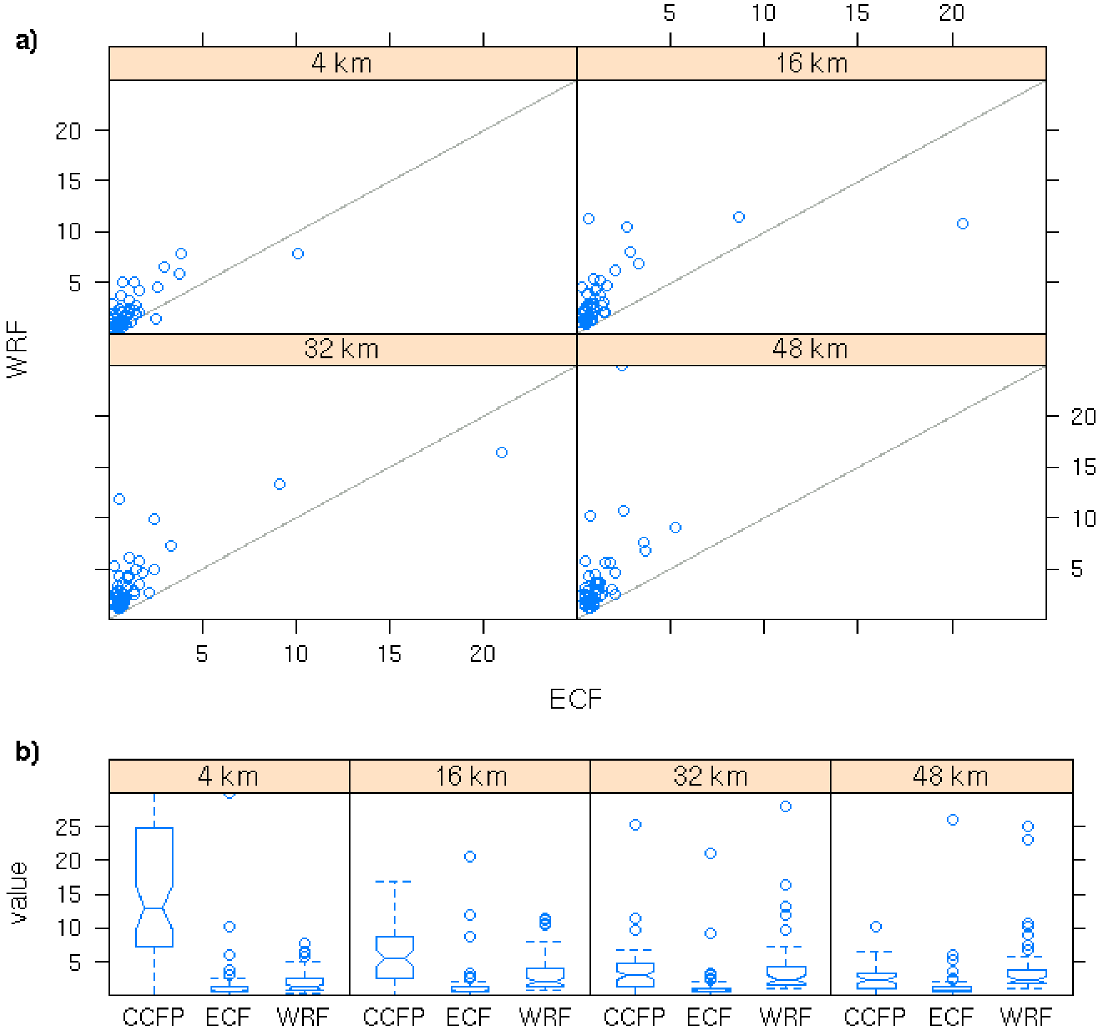

Figure 15. As in Fig. 13, but for BIAS.

Overall, the results of the resolution analysis mirror those from section 4.1.1, with the WRF model outperforming ECF. Therefore, by the sector-based and resolution-based methods, it is not clear that the ECF was able to add any appreciable forecast information over the model.

4.3. Domain decomposition analysis

The results will be broken down into overall JPD results to look at how the forecasts compare without observations in the different subdomains of agreement and disagreement. Additionally there will be a further examination of the primary agreement subdomain (YY) and the primary disagreement subdomain (YN).

4.3.1. JPD summary

The overall JPDs for both the CCFP-ECF supplement and the CCFP-WRF supplement are shown with respect to total number of forecasts and average percent coverage of the TRACON verification domain in Tables 5 (a) and 5 (b). The sample size for the CCFP-WRF supplement is slightly smaller due to some of the WRF data being unavailable during the study period.

Table 5. The joint probability distribution (JPD) for total number of forecasts with each subdomain decomposition region and the mean fractional area coverage for (a) the CCFP-ECF supplement and (b) the CCFP-WRF supplement.

(a)

CCFP-ECF Decomp. Region	Number of Forecast (%)	Mean Fractional Area Coverage (%)
YY	639 (70.14)	0.2285
YN	757 (83.10)	0.0332
NY	166 (18.22)	0.0072
NN	911 (100)	0.73

(b)

CCFP-WRF Decomp. Region	Number of Forecast (%)	Mean Fractional Area Coverage (%)
YY	620 (73.11)	0.2358
YN	838 (98.82)	0.1133
NY	93 (10.97)	0.0033
NN	848 (100)	0.6475

In the primary agreement subdomain (YY), it is interesting that both the mean fractional area coverage and the percentage of forecasts within a YY subdomain are approximately the same for both supplementary forecasts. This is not surprising as the CCFP dominates the area defined by the YY boundary. However, as the numbers do not vary much between the supplementary forecasts some interesting results may be revealed. Discrepancies exist between the YN, NY, and NN subdomains in terms of both the percentage of forecasts having these decomposition subdomains and the mean fractional area coverage. These differences can be explained by the WRF simulated reflectivity product forecasting slightly more significant convection that is more widespread than the ECF product over the entire TRACON verification domain.

4.3.2. YY subdomain

In the YY subdomain (forecast agreement of convection), convection occurs in greater than 90% of the forecast issuances; when the supplementary products (ECF or WRF) agree with the CCFP forecast, significant convection can be expected. As supplements to CCFP, ECF and WRF differ in utility. The WRF simulated reflectivity appears to add more value than ECF by better forecasting the amount of significant convection (NCWD VIP level 3 or greater) in and around a CCFP polygon. WRF also seems to better forecast the center of mass of the convection. However, this may not be a significant difference. To a first order approximation, significant convection is often located near the center of CCFP polygons. The ECF does not discriminate between observed objects of different size and shape, where the WRF appears to present more structural information. Overall, the WRF simulated composite reflectivity product alone adds more supplementary value than the ECF in terms of structure in the YY subdomain.

The YY subdomain for both supplementary products has about the same mean fractional area coverage (~23%) and a similar percentage of number of forecasts with the YY subdomain present. In the YY region, significant convection occurs greater than 90% of the time. This means that where there is forecast agreement between either CCFP and ECF or CCFP and WRF, there is at least some convection in this subdomain greater than or equal to VIP level 3. The remaining 10% of the time, the YY region tends to be smaller and made up of clipped CCFP polygons along the edge of the subdomain where verification may be less meaningful. One conclusion from this overview is that the end-user may have increased confidence of the occurrence of significant convection when ECF or WRF exists within a CCFP polygon. Further examination of the supplemental forecasts may indicate which (ECF or WRF) adds more value within this CCFP polygon.

The bias of ECF and WRF inside a YY subdomain is shown in Figs. 16 and 17, respectively. Although WRF simulated reflectivity generally overforecasts for the entire TRACON domain, the median bias within a CCFP polygon is close to one, whereas the median bias for ECF is around 0.5. In this case, where the CCFP and the supplementary products are in agreement, WRF typically adds more structure in terms of amount of convection present. In comparison to the CCFP, WRF reasonably captures the amount of convection while CCFP has a bias much greater than one on average, indicating overforecasting.

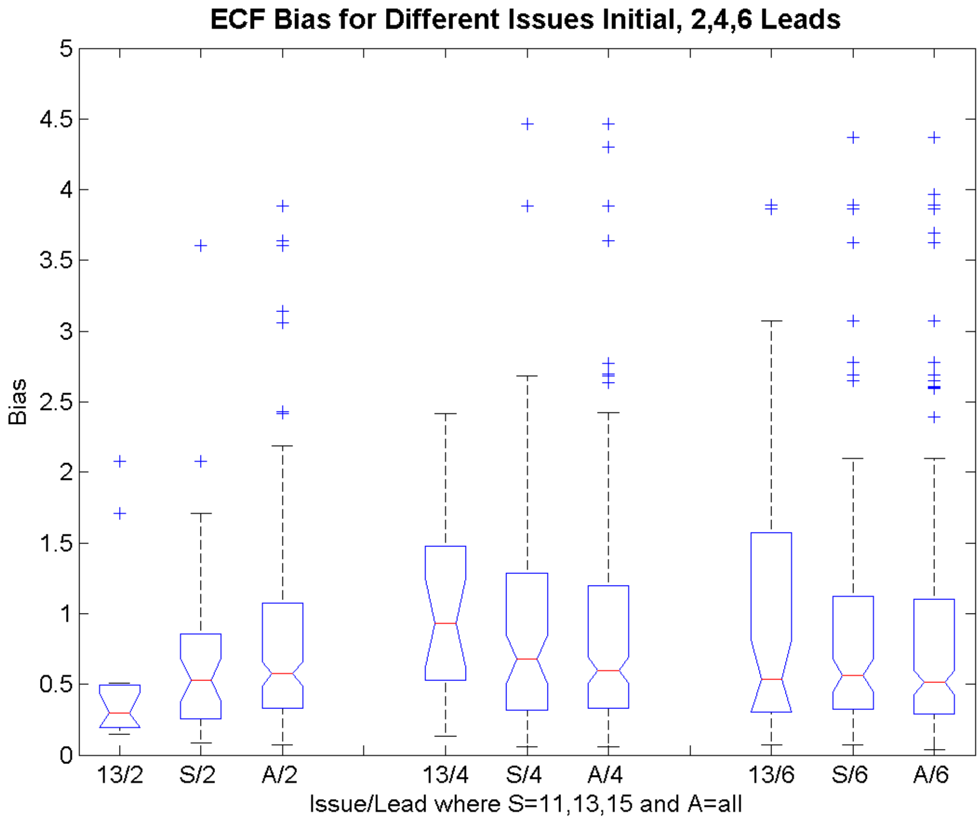

Figure 16. Boxplots of the bias of ECF within the YY region for the 13 UTC issuance, the strategic issuances (11, 13, 15 UTC), and all issuances available for the 2-, 4-, 6-h leads. The notch represents the 95% confidence interval, the upper and lower bounds of the box represent the quartile ranges, the whiskers represent 1.5 standard deviations and the blue +'s represent outliers.

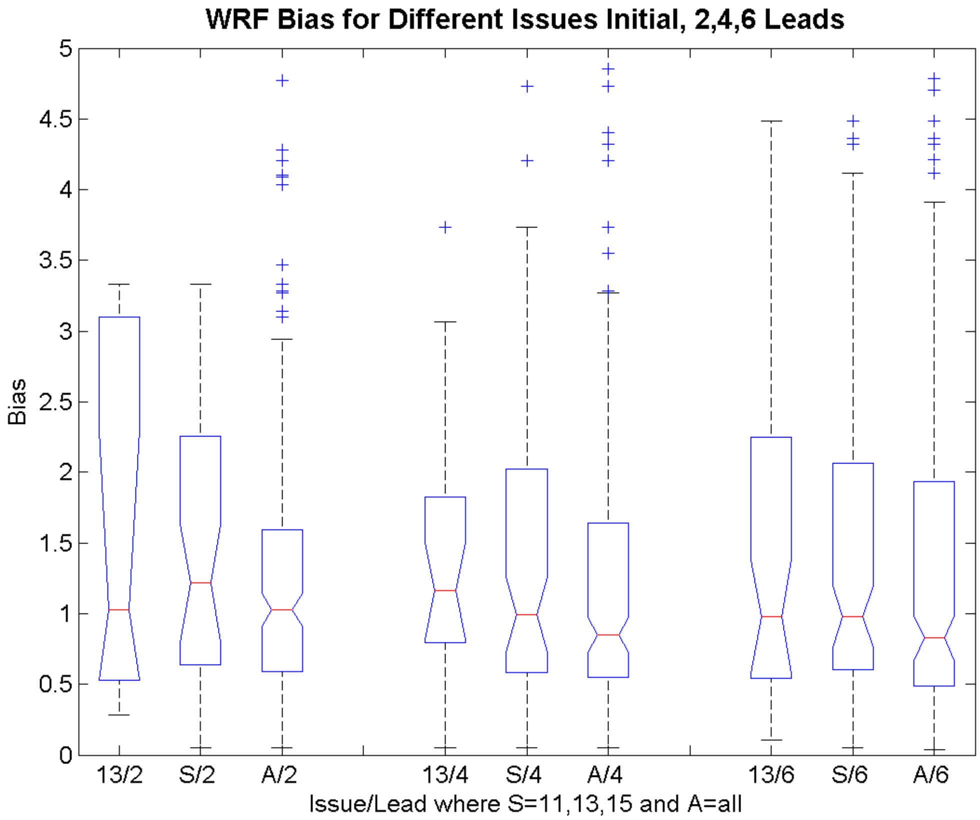

Figure 17. Boxplots of the bias of WRF within the YY region for the 13 UTC issuance, the strategic issuances (11, 13, 15 UTC), and all issuances available for the 2-, 4-, 6-h leads.

The center of mass displacements between CCFP and the observation, the supplementary product (ECF or WRF) and the observation, and CCFP and the supplementary product are used as a first order approximation for placement accuracy inside a YY region. Clearly, if there is no difference between the center of mass between the CCFP and the supplementary product combined with no differences between the observation field and the forecast products, no first order structural information is gained in terms of convective placement. The center of mass displacements from CCFP to the supplementary forecasts are shown in Figs. 18 (CCFP-ECF) and 19 (CCFP-WRF). As this is a significant difference (~70 km in both ECF and WRF), it can be inferred that ECF and WRF are not often collocated with the center of mass of CCFP polygons.

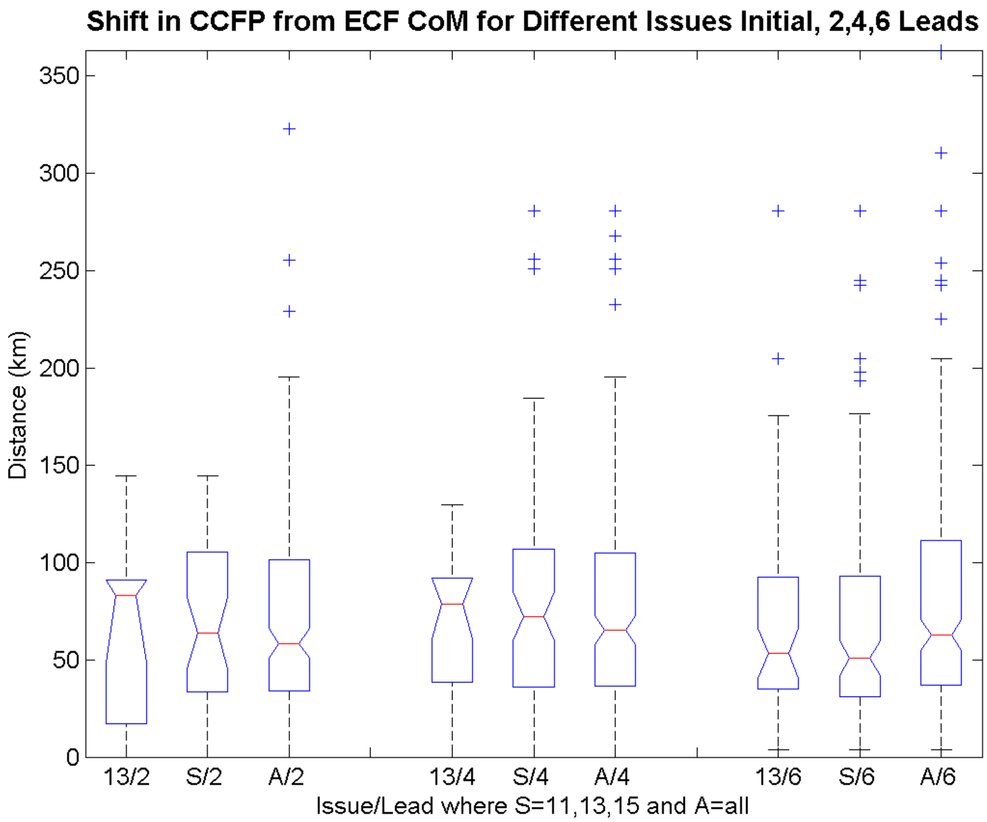

Figure 18. Boxplots of the shift in center of mass between the CCFP and ECF in km. Note the positive difference in terms of the median for all issues and leads.

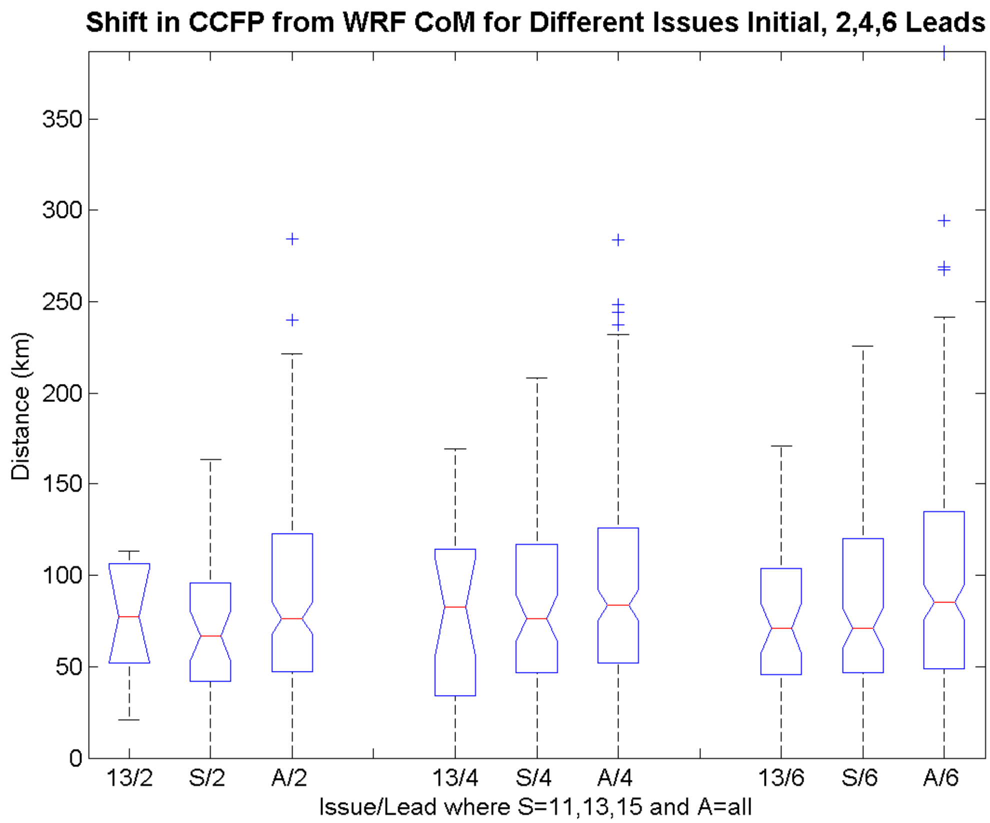

Figure 19. Boxplots of the shift in center of mass between the CCFP and WRF in km. Note the positive median difference for all issues and leads.

The baseline for comparison is the displacement of observation from the CCFP center of mass (Fig. 20). To a first order, if the forecasts do not show improvement over the baseline CCFP, the only structural information available from the supplementary product would be in terms of bias. The displacements between the supplementary forecasts (ECF and WRF) and the NCWD observations are shown in Figs. 21 and 22, respectively. These displacements show that although the median displacement is slightly lower for the WRF to the observation than the ECF to the observation and the CCFP to the observation it may not be a significant difference (using the notch as a confidence interval proxy). However, as these differences may not be significant, it may be concluded that the WRF and ECF do no worse at identifying the placement of convection inside a YY region than the CCFP center. This result actually favors CCFP and demonstrates that there is some correlation between the location of convection and the center of a typical CCFP polygon.

Additionally, structure inside a YY region can be measured and compared by identifying similar object traits between the observation (NCWD) field and the supplemental field. The results of this study were obtained by saving information on the mean and median object size for ECF, WRF, and NCWD observations (see Appendix for Figs. A1-A6). Of note, the mean size of observation objects inside a YY region closely match those of the ECF objects; however, the distribution of observed (NCWD) objects share a right-skewed distribution (mean higher than the median) with the WRF simulated reflectivity objects. ECF in this case clearly does not discriminate significant convective object size or shapes.

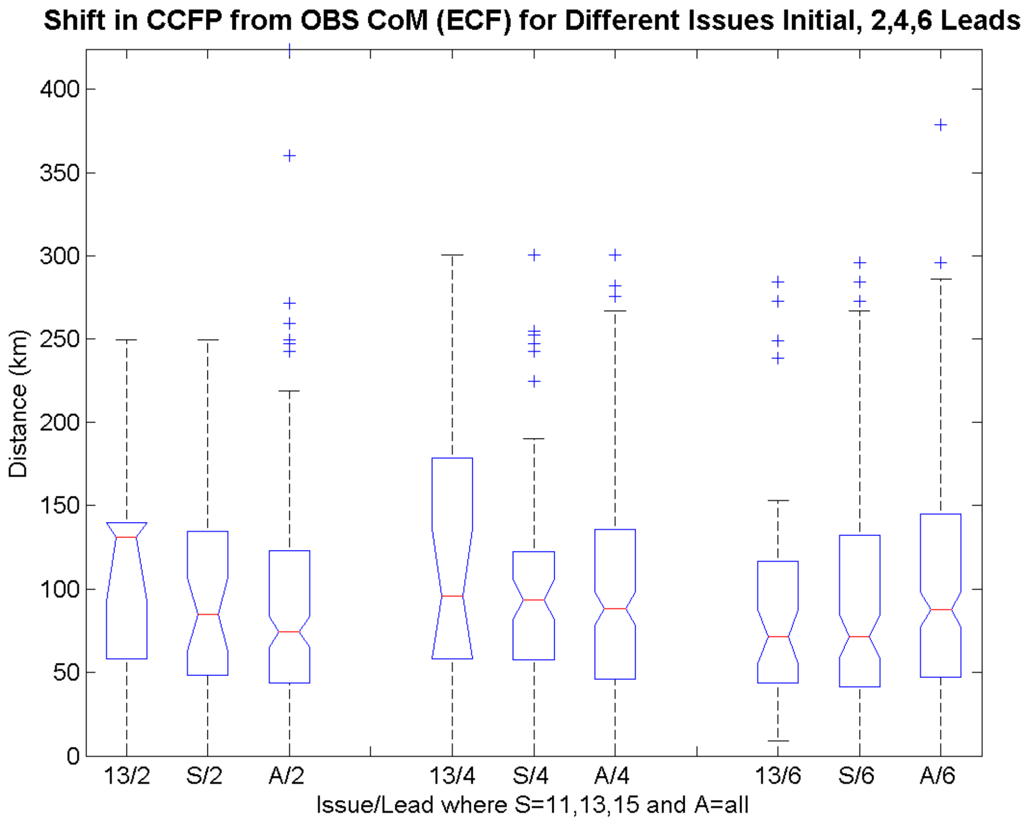

Figure 20. Boxplots of the shift in center of mass between the CCFP and NCWD observations (OBS) in km. The median difference for all issues and all lead combinations is roughly 100 km.

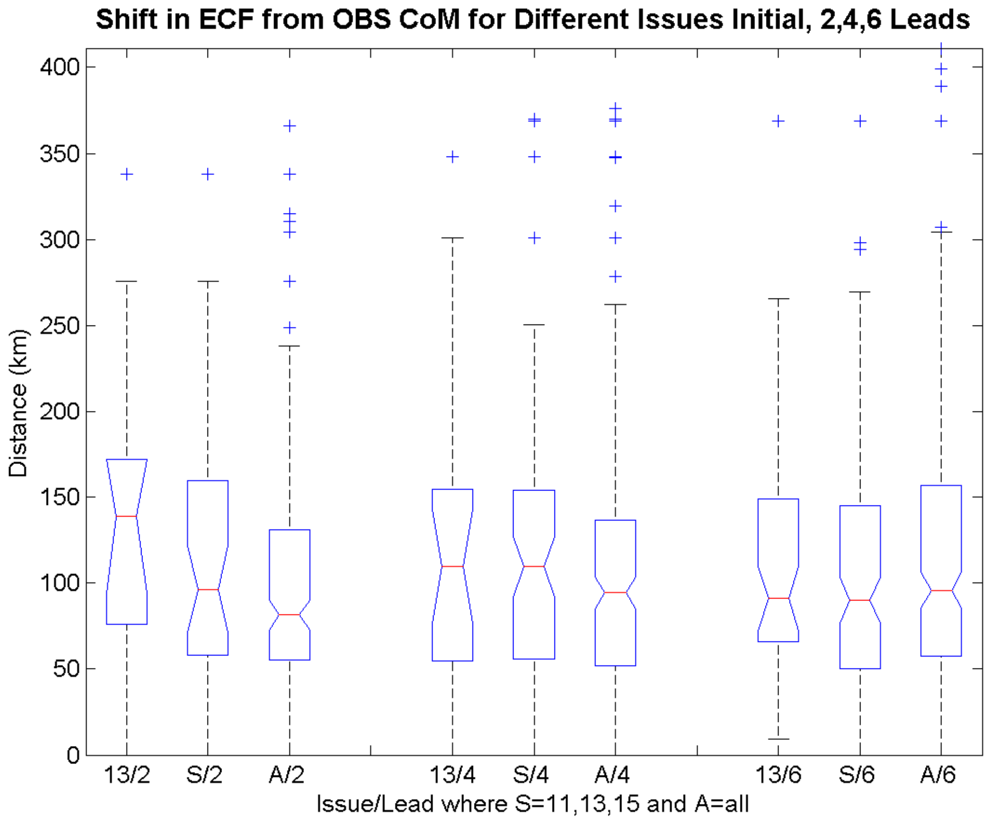

Figure 21. Boxplots of the shift in center of mass between the ECF and NCWD observations (OBS) in km. The median difference for all issues and all lead combinations is slightly above 100 km.

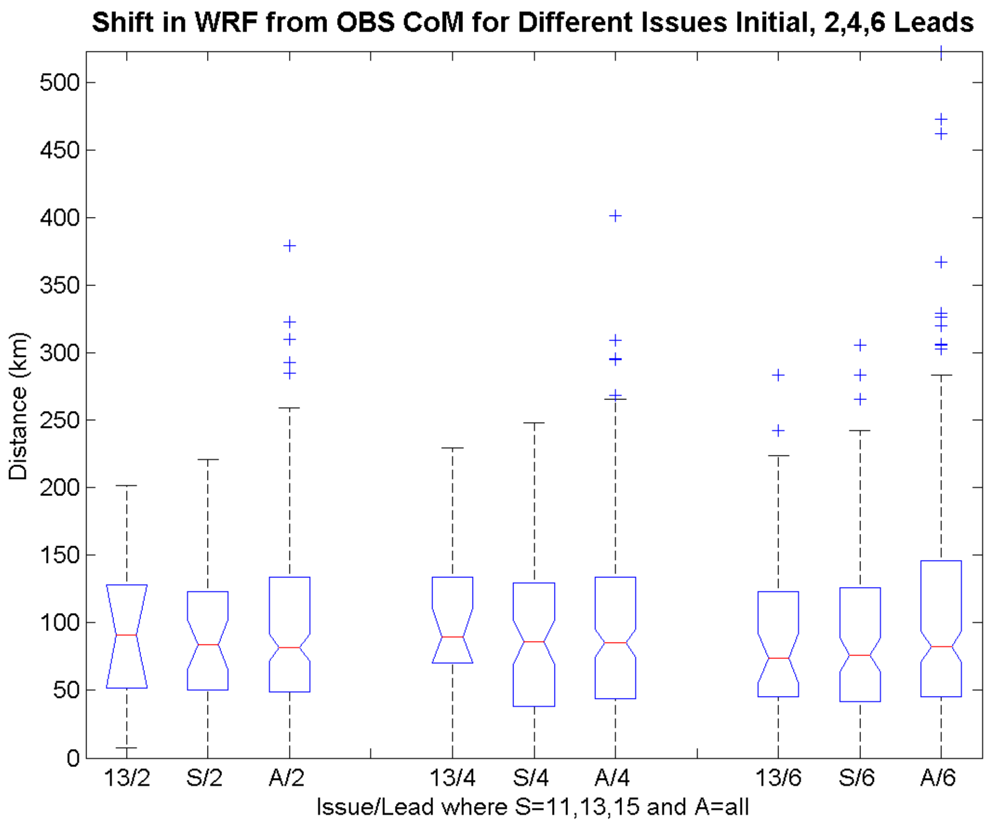

Figure 22. Boxplots of the shift in center of mass between the WRF simulated reflectivity and NCWD observations (OBS) in km. The median difference for all issues and all lead combinations is slightly below 100 km.

4.3.3. YN subdomain

Overall, neither supplemental forecast seems to consistently identify significant convection outside of a CCFP polygon; however, the WRF simulated reflectivity product with its larger areal coverage in the YN subdomain does capture slightly more convection than the ECF. Despite the larger areal coverage of the WRF forecast domain, the bias is lower for the WRF than the ECF. Additionally, in most cases the YN regions tend to be regions of isolated significant convection that may not meet minimum CCFP criteria to warrant a polygon. More research would need to be done to completely quantify this observation. Regardless, with median biases greater than three for both WRF and ECF in the YN subdomain when convection is present, this overforecasting trend would suggest underdelivery of air traffic to these regions.

The YN subdomain has a large mean fractional area coverage difference between the ECF and WRF supplementary forecasts. ECF has a mean areal subdomain coverage of ~3% while WRF has a mean areal coverage of ~11%. This difference results from WRF having more overall convective coverage within the TRACON and being slightly more widespread, especially for small areas of significant convection. Despite this difference in areal coverage, meaningful results can be given to address whether ECF or WRF adds value in places lacking CCFP polygon coverage.

The number of forecasts containing a YN region and any observed convection within the 32-km buffer zone provides a valuable statistic for the YN subdomain. For ECF, out of the 757 forecasts that have a YN region, only 323 forecasts (42.67%) contain any convection within the 32-km buffer. For WRF, out of the 838 forecasts that have a YN region, 515 (61.46%) contain any significant convection within the buffer. The next question that follows from this is when the forecasts capture any convection how well does it capture the amount of convection within the buffer. This question can be answered by looking at the bias of the ECF and WRF when convection is present in the YN region (Figs. 23 and 24). Although the WRF simulated reflectivity YN regions are on average three times larger than the ECF forecast region, the bias of the WRF appears to be lower for all issue and lead combinations although it may not be a significant improvement in some cases.

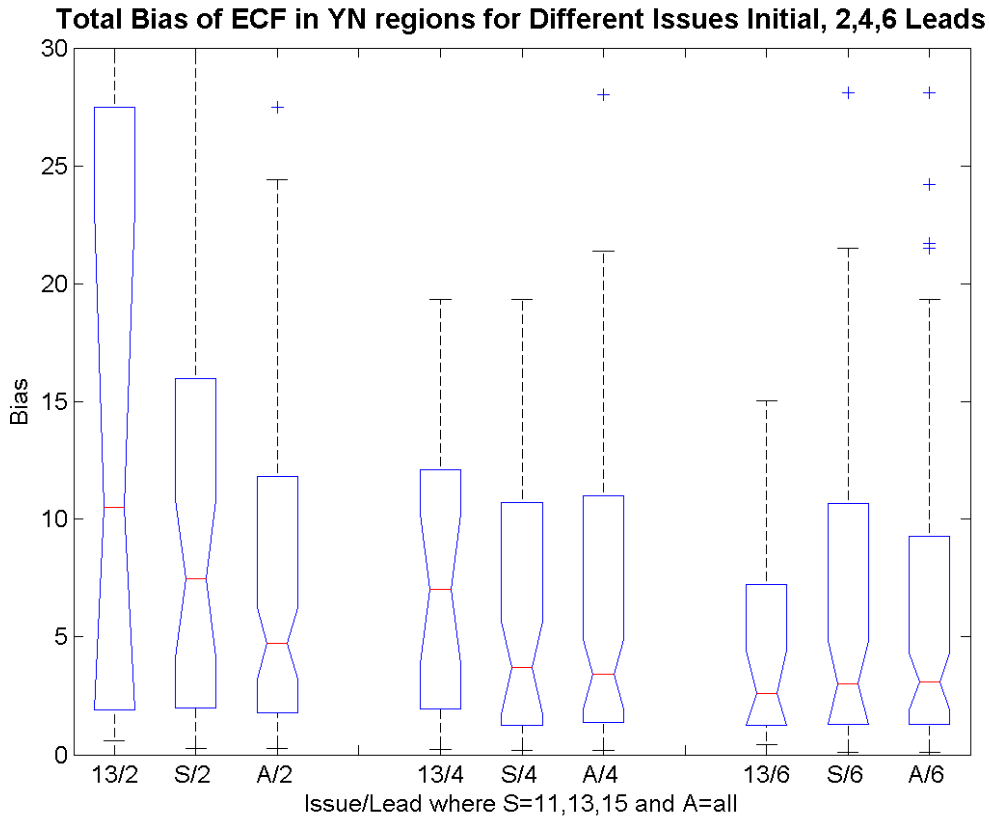

Figure 23. Boxplots of total bias of ECF forecasts when convection is present in a YN region.

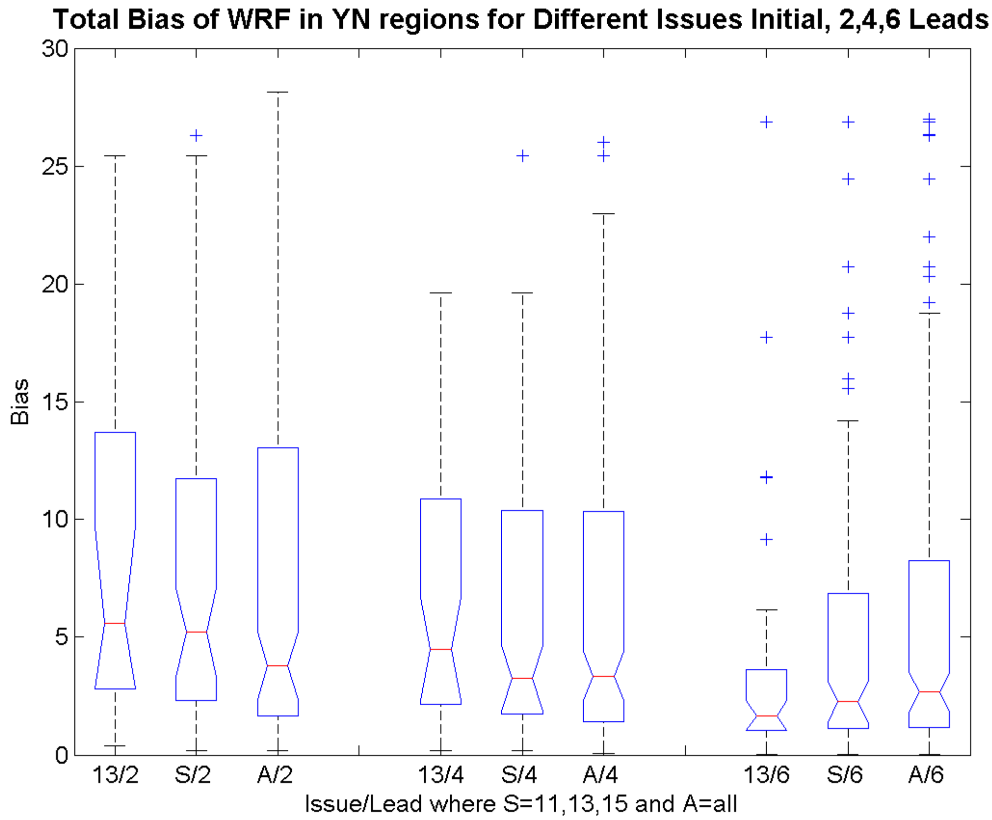

Figure 24. Boxplots of total bias of WRF forecasts when convection is present in a YN region.

4.3.4. NY subdomain

The NY subdomain is given little regard as most of the CCFP polygons associated with this region are almost always clipped polygons that are found along the edge of the verification domain. This can be seen in Table 5a. as very low areal percent coverages (<1%) on average. Due to the more widespread nature of the WRF simulated reflectivity product it is of no surprise that the NY regions are smaller than for the ECF forecasts. Additional case study work would need to be done to illustrate cases in which CCFPs within the TRACON are of significant size and perform well or poor in the absence of ECF or WRF forecasts. One particular case, not necessarily typical, shows a relatively good CCFP forecast in significant disagreement with both supplemental forecasts, (Figs. 25 and 26).

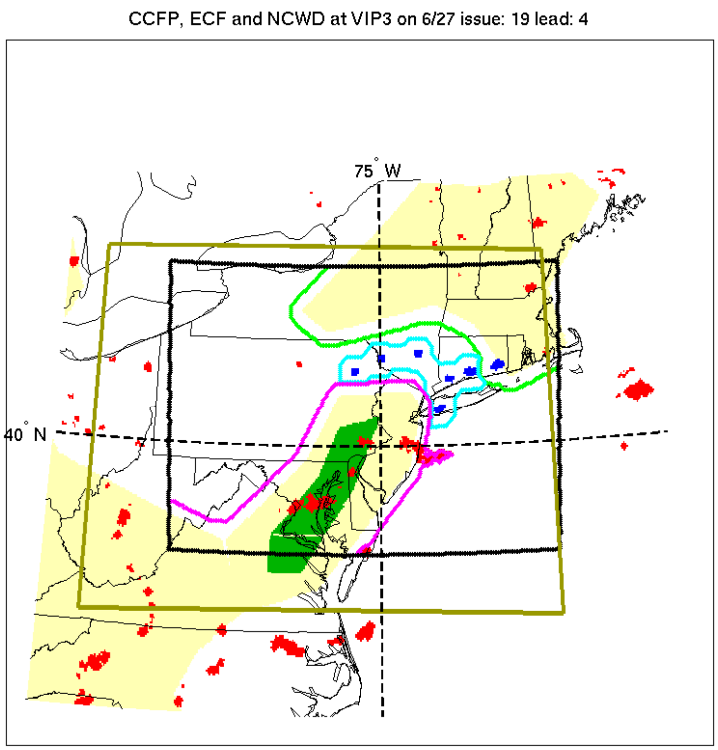

Figure 25. CCFP (yellow and green) and ECF forecasts (blue) issued at 19 UTC valid at 23 UTC on 27 July 2008 overlaid with significant NCWD (red). The green indicates a medium-coverage, high-confidence CCFP polygon within a NY region outlined by magenta.

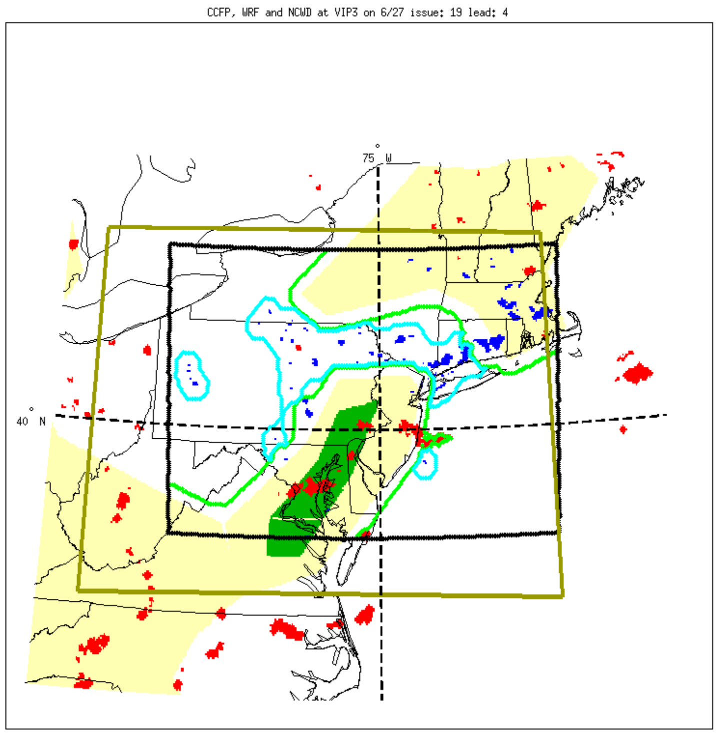

Figure 26. CCFP (yellow and green) and WRF simulated reflectivity (blue) issued at 19 UTC valid at 23 UTC on 27 July 2008 overlaid with significant NCWD (red). The green indicates a medium-coverage, high-confidence CCFP polygon within a YY region outlined by green. This YY region has some very isolated WRF reflectivity associated with it.

4.3.5. NN subdomain

The NN region highlights that when CCFP and the supplementary forecast agree with a forecast of no convection, the forecast combination tends to accurately predict no convection. The NN subdomain is characterized by the space that remains once the regions covered by forecasts and associated buffers are removed. The NN subdomain can be used as a measure of forecast misses. If a given forecast has a lot of convection present in the NN region, an impactful event may have been missed. For each forecast issued in the study period, both the CCFP/ECF and CCFP/WRF NN regions had median convective coverage of less than 1% (Figs. 27 and 28). Additionally, only 6 forecasts for the CCFP/ECF combination and 19 forecasts for the CCFP/WRF combination had significant convection covering more than 5% of the NN region.

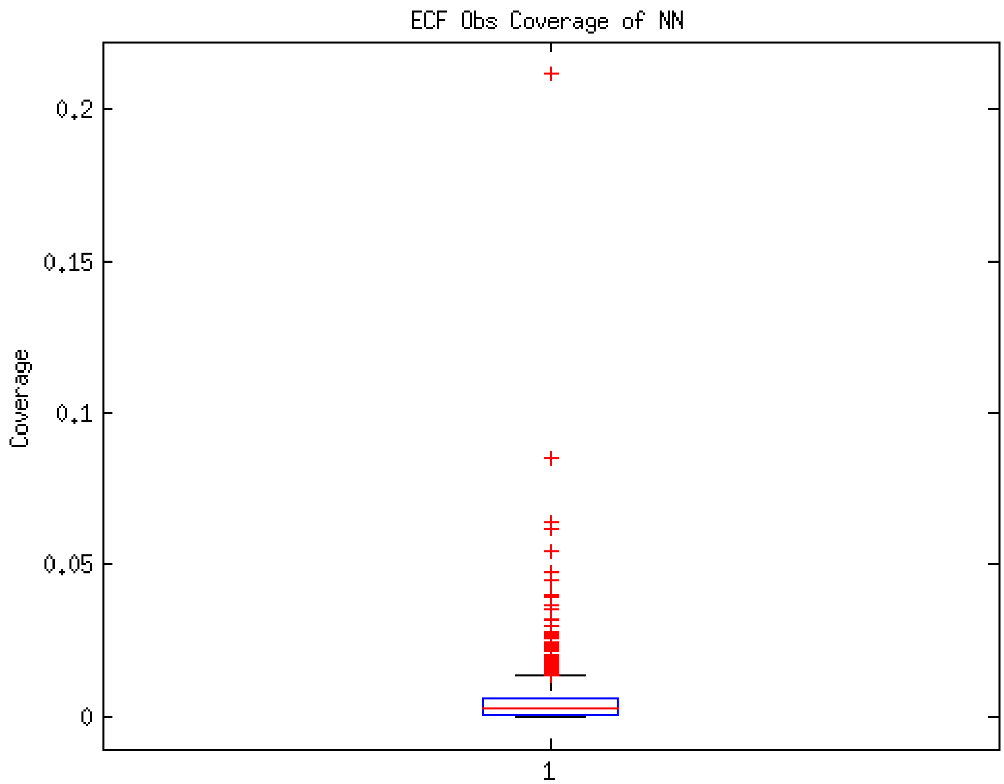

Figure 27. Boxplot of significant convective coverage in the NN subdomain for all forecasts issued for the CCFP/ECF combination.

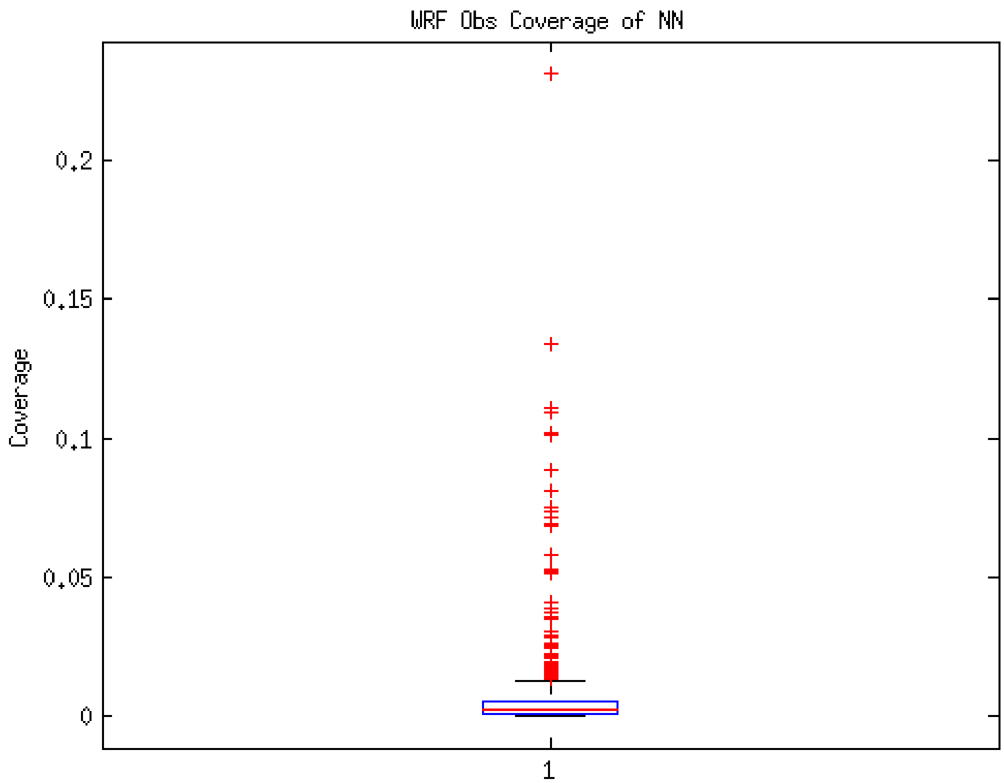

Figure 28. Boxplot of significant convective coverage in the NN subdomain for all forecasts issued for the CCFP/WRF combination.

4.3.6. Discussion of domain decomposition results

Overall findings from the domain decomposition can be summarized in the following bullets:

- When CCFP and the supplementary products are in agreement (YY region), the WRF simulated reflectivity tends to capture the amount and placement (in terms of center of mass) better than the ECF. However, CCFP may have implied structure as its center of mass measure is slightly better than ECF.
- ECF tends to not forecast object sizes consistent with observation objects. ECF tends to have small, circular forecast objects that seemingly vary little from day to day.
- The WRF forecasts smaller objects than observed, however, the distribution of object shapes closely resemble those of observed objects from NCWD. The WRF also varies these object sizes on a cases-by-case basis.
- When the supplementary forecasts are in agreement with CCFP for no convection, the no convection forecast can be trusted 99% of the time.

5. Conclusions

Accounting for resolution and domain differences between the products, two approaches are adopted for the direct comparison of the ECF and WRF to the operational baseline. The first, a sector-based method, scores the products on a “grid” of air traffic sectors, allowing for displacement of the forecast on a user-relevant scale. The second, a resolution-based approach, measures performance on a number of regularly spaced grids, ranging from 48 km down to the native resolution of the observation field, 4 km.

The analysis of the supplemental relationship of ECF and WRF with respect to CCFP relies on a decomposition approach that identifies agreement and disagreement between the forecasts. Within each subdomain, the supplemental forecasts are characterized with various measures (e.g., structure of convection) to determine the value added to CCFP.

Results from the direct comparison of the products indicate:

- By all measures in the sector-based approach, ECF performed poorly relative to the CCFP baseline, clearly due to the small size and limited number of individual ECF polygons within any given forecast. Overall, the WRF and CCFP showed similar skill (as measured by the Heidke Skill Score), but the forecasts appear to have different strengths and weaknesses. The WRF forecast often misplaced convection relative to its occurrence, but effectively predicted the correct number of sectors significantly covered by hazardous weather. CCFP, while overforecasting, shows value by locating most of the sectors with significant coverage.
- As in the case of the 4- and 6-h leads, the ECF failed to show any appreciable skill in forecasting convective coverage of air traffic sectors at 8-h and 10-h leads. The WRF forecast, however, appears to retain a small amount of skill, which degrades by a factor of two at these longer lead times. Used cautiously, the longer lead times of the WRF forecast may provide utility to planners.
- In addition to confirming the overall indications found in the sector-based approach, the resolution-based analysis suggests that the ECF and WRF products don’t perform well at high resolution. Even at the most coarse resolution (48 km) studied, median performance values are very low.

Results from the analysis of the supplemental relationship indicated:

- When the forecasts agree on the presence of convection: The WRF simulated reflectivity appears to add more value than ECF by better forecasting the amount of hazardous convection in and around a CCFP polygon. Additionally, the WRF seems to better indicate the structure of convection within a CCFP polygon, as measured by the “center of mass” of the convection, and the distributions of the sizes and shapes of convective objects.

- When ECF and WRF forecast convection outside of CCFP polygons: Overall, neither supplemental forecast seems to effectively identify significant convection outside of a CCFP polygon. Additionally, in most cases these regions appear to contain isolated convection that may not meet minimum CCFP criteria to warrant a polygon. More research would need to be done to completely quantify this observation.
- When only CCFP forecasts convection: Rarely – in less than 1% of the area of the domain -- does a CCFP polygon exist without some associated ECF or WRF forecast of convection. These cases almost always contain only "clipped" CCFP polygons that are found along the edge of the verification domain, likely a result of the difference in the product domains and granularity.
- When CCFP and the supplementary forecast agree on a forecast of 'no convection': The forecast combinations appear to accurately predict 'no convection.' For forecasts issued in the study period, the combined regions of 'no convection' have median convective coverage of less than 1%. Rarely did both CCFP and its supplement miss significant convection; together, in this subdomain, they appear very trustworthy for use in air traffic planning.

Acknowledgments

This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

References

Ebert, E.E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appls*., **15**, 51-64.

Federal Aviation Administration, 2008: FAA ConUse Doc for the TRACON Study. Available from http://dist2.extremeforecasting.com/weather/tracon/src/ConUse_New_York_Forecast.pdf.

Kay, M.P., S. Madine, J.L. Mahoney, and J.E. Hart, 2007: 2007 Convective forecast scientific evaluation. Prepared for the FAA System Operations Group. Available from the following URL: http://www-ad.fsl.noaa.gov/fvb/publications/articles/convintercomp_v1.1.pdf.

Megenhardt, D., C. Mueller, S. Trier, D. Ahijevych, and N. Rehak, 2004: NCWF-2 Probabilistic nowcasts. 11th Conf. on Aviation, Range and Aerospace Meteorology, Hyannis, MA, Amer. Meteor. Soc., 23 pp.

National Weather Service, date unknown: Collaborative Convective Forecast Product Product Description Document. Available from http://aviationweather.gov/products/ccfp/docs/pdd-ccfp.pdf.

Phaneuf, M., 2008: New York 2008 Convective Weather Project Plan, 9 pp. Available from the author at mphaneuf@avmet.com.

Wilks, D.S., 2006: Statistical Methods in the Atmospheric Sciences. Second Edition. Academic Press, San Diego, 627 pp.

Appendix

The figures in this section refer to material in the YY sub-domain results section (Section 4.3.2). These plots offer a comparison of mean and median object areas in the supplementary forecasts (ECF and WRF) compared to the observed (NCWD) field in an effort to portray skill in adding structure to CCFP.

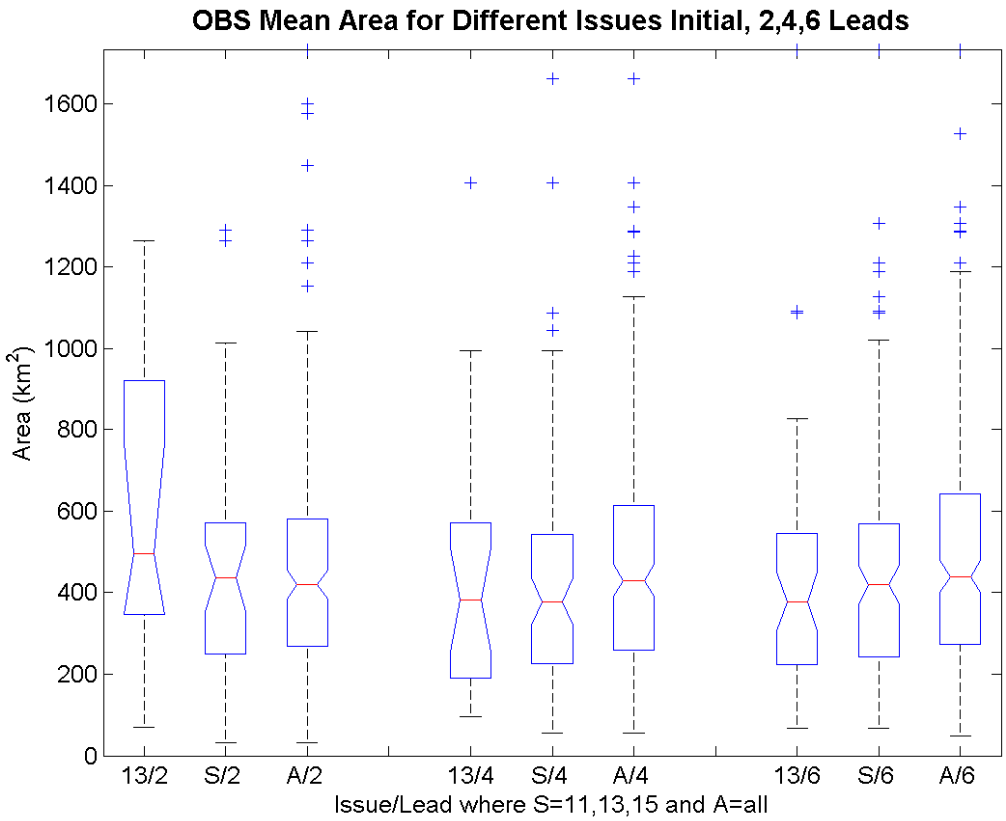

Figure A1. Boxplots of NCWD object mean size.

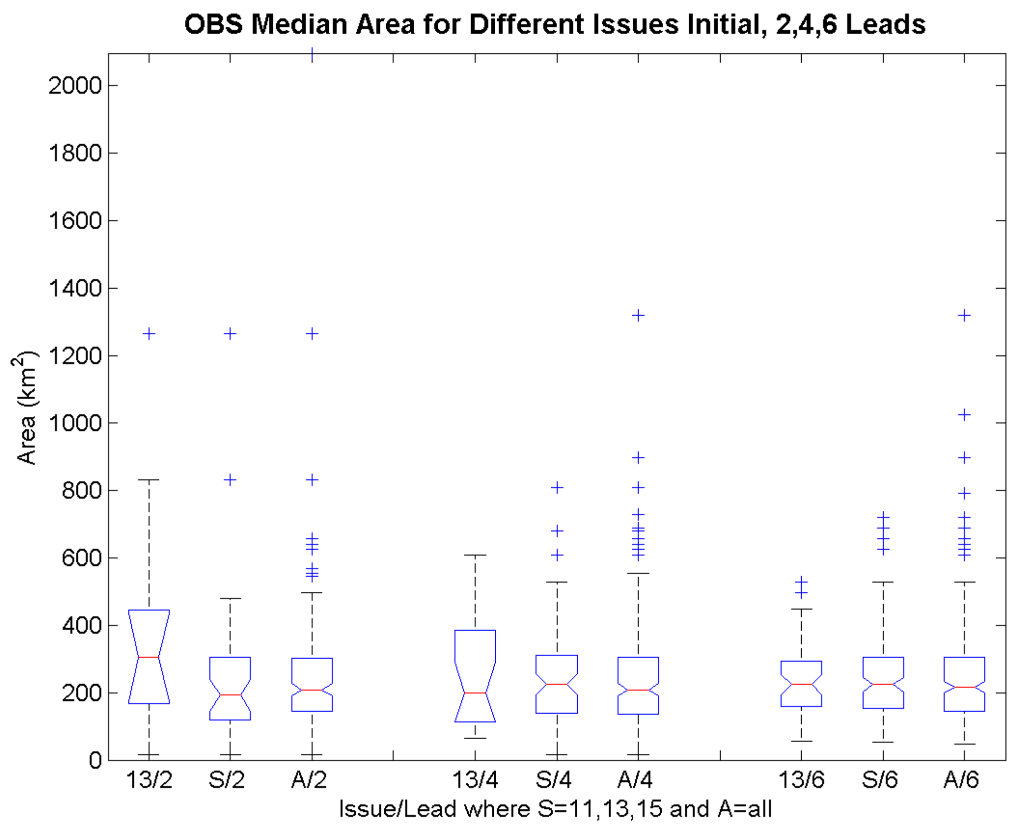

Figure A2. Boxplots of NCWD object mean size.

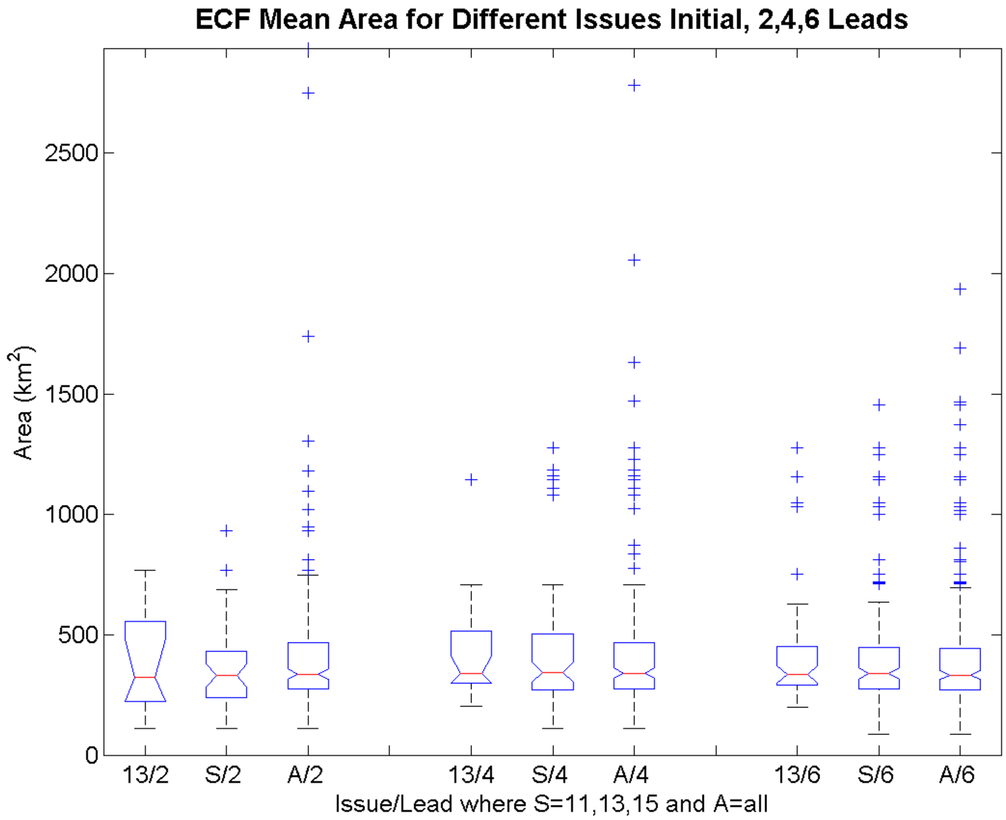

Figure A3. Boxplots of ECF object mean size.

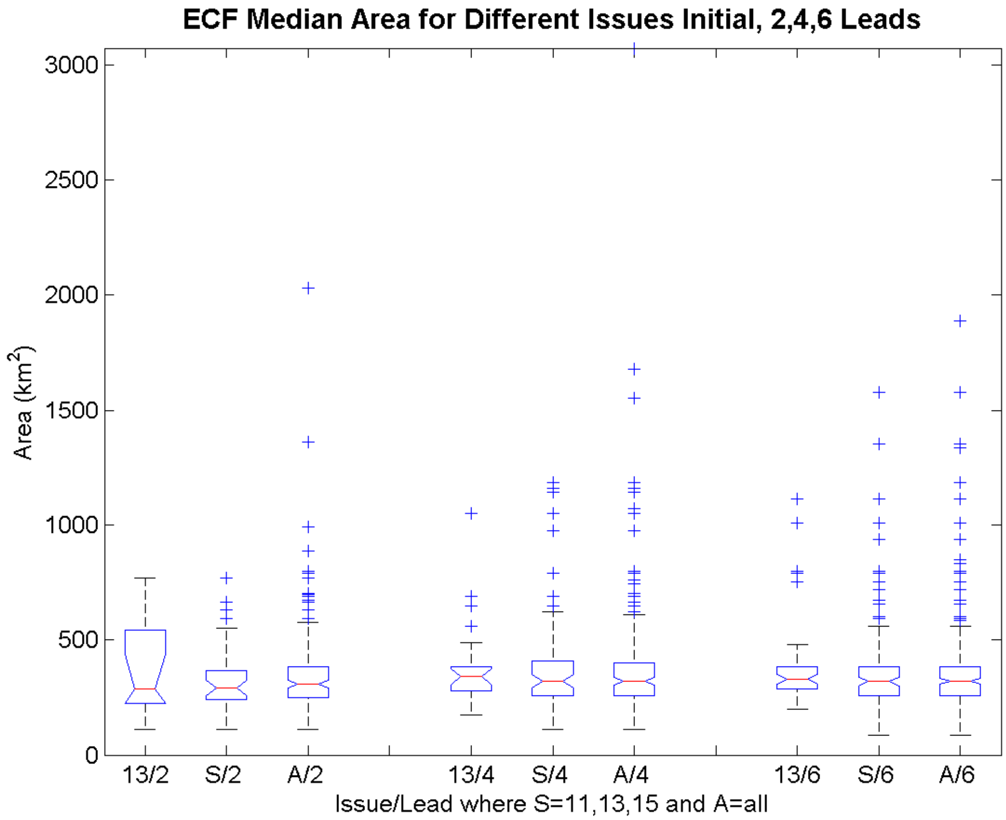

Figure A4. Boxplots of ECF object median size.

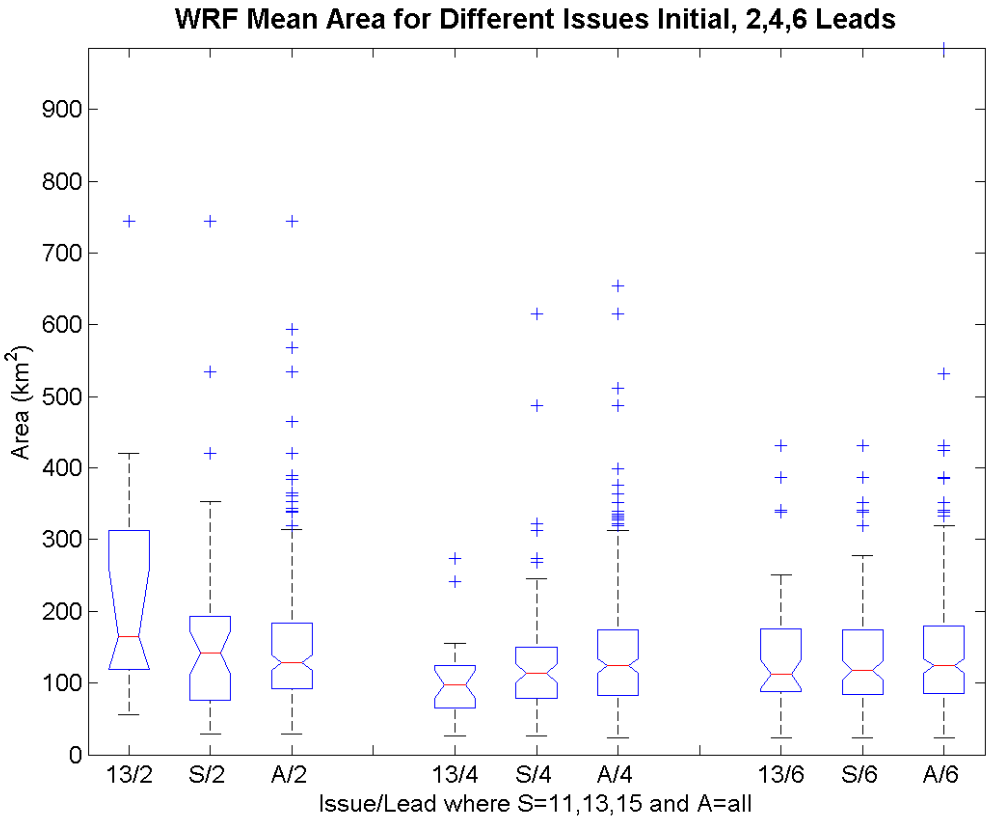

Figure A5. Boxplots of WRF object mean size.

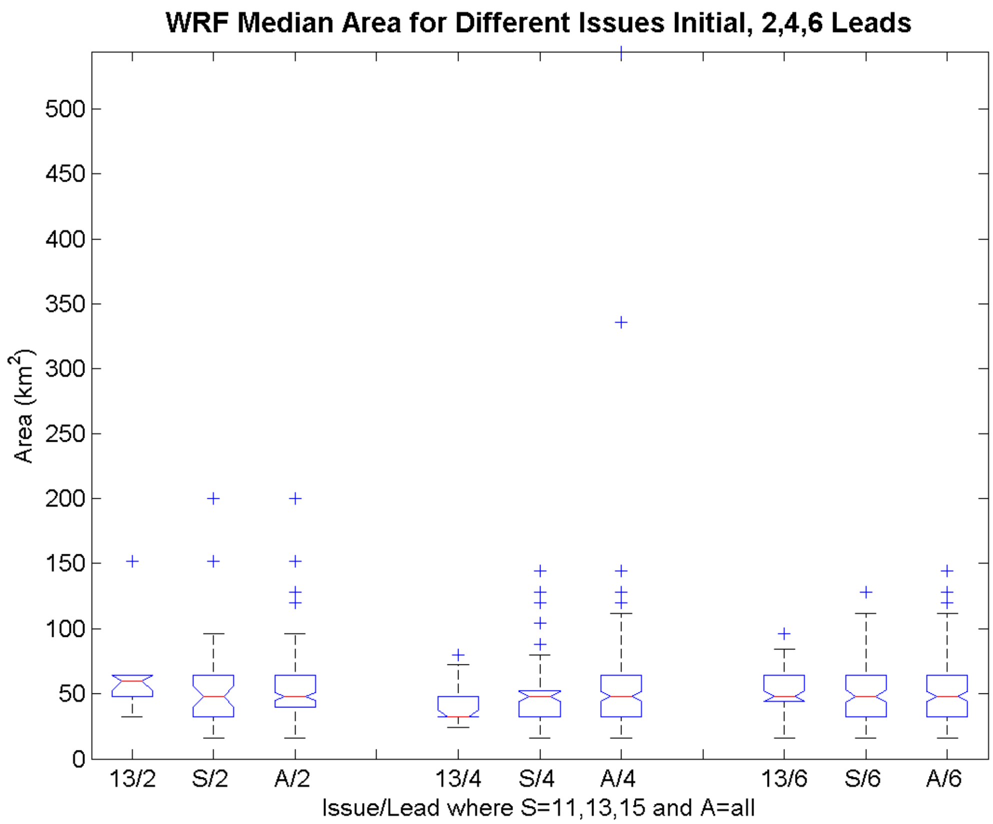

Figure A6. Boxplots of WRF object median size.