

2011 QUALITY ASSESSMENT OF COSPA

Prepared by the Quality Assessment Product Development Team
NOAA/ESRL/Global Systems Division

Steven A. Lack, Michael P. Kay, Geary J. Layne, Melissa A. Petty, and
Jennifer L. Mahoney

16 March 2012

Corresponding Author: J.L. Mahoney (NOAA/ESRL/GSD, 325 Broadway, Boulder,
CO 80305; Jennifer.Mahoney@noaa.gov)

This page intentionally left blank

Table of Contents

1	Introduction	1
2	Approach	1
3	Data.....	2
3.1	CoSPA.....	3
3.2	High Resolution Rapid Refresh (HRRR).....	4
3.3	Collaborative Convective Forecast Product (CCFP)	4
3.4	The Localized Aviation MOS (Model Output Statistics) Program (LAMP) Thunderstorm Product.....	5
3.5	CIWS Observations.....	6
4	Methods.....	6
4.1	Climatology	6
4.2	Upscaling.....	7
4.3	Fractions Skill Score	7
4.4	Flow Constraint Index.....	9
4.5	Forecast Consistency.....	11
4.6	Clustering.....	12
4.7	Statistics	13
4.8	Stratifications	15
5	Results	16
5.1	Climatological Analysis.....	16
5.1.1	Diurnal Convective Signal.....	17
5.1.2	Regional Forecast Differences.....	19
5.2	Performance Analysis	22
5.2.1	Upscaling CoSPA VIL with Lead Time	22
5.2.2	Examination of CoSPA Echo Tops.....	25
5.2.3	Forecast Resolution Analysis	29
5.2.4	Quality Relative to Airspace Flow Constraints	34
5.2.5	Forecast Consistency	36
5.2.6	Performance at CCFP Scales.....	37
5.3	Performance of CoSPA during Winter	41
5.3.1	Winter Climatology	41
5.3.2	Performance Analysis.....	44
6	Summary and Conclusions	49
7	References.....	52
	Acknowledgements.....	53
	Appendix	53

List of Figures

- Figure 3.1. Graphic of CoSPA VIL from 23 June 2011 issued at 1300 UTC, valid at 17 UTC (grey VIP-levels 1 and 2, yellow VIP-level 3 and 4, red VIP-level 5 and greater), CCFP polygons (key in upper-left hand corner of graphic), and echo tops (light purple for tops less than 30 kft, and dark purple for echo tops 40 kft and above at 5-kft intervals).....3
- Figure 3.2. An example of CCFP for 11 May 2011 issued at 1700 UTC and valid at 2300 UTC.4
- Figure 3.3. The operational LAMP Thunderstorm Product on 5 October 2011 issued at 1300 UTC valid at 1600 UTC.....5
- Figure 4.1. An example of upscaling a 6x6 grid to a 2x2 grid. Taking the average of the 4 3x3 calls on the left creates the 4 pixels on the right.....7
- Figure 4.2. Illustration of the FSS. Observation field (top left), deterministic forecast (top right), a uniform forecast (bottom left), and a CCFP-like forecast (bottom right).....8
- Figure 4.3. Fractionalized grid for the 3x3 neighborhood for the observation field (top right), deterministic forecast (top right), uniform forecast (bottom left), and CCFP-like forecast (bottom right).....9
- Figure 4.4. Conceptual model of the FCI. Blue lines represent the corridor boundaries and the red area denotes an area of hazardous convection. Arrow 1 represents the minimum distance across the corridor in the absence of convection. Arrows 2 and 3 show the minimum distance across the available airspace around a hazard..... 10
- Figure 4.5. Illustration of the FCI concept for a hexagonal geometry. The hexagon contains three separate corridors, one for each pair of opposing faces: traffic moving from northeast to southwest, from north to south, and from northwest to southeast. (The FCI is identical for traffic flowing in the opposite directions.). A weather hazard is denoted by the red area. The green arrow (left) shows the mincut distance for the northeast-to-southwest corridor. The length of the red lines (right; as a fraction of the total corner-to-corner distance) represent the FCI value for traffic moving perpendicular to the line..... 11
- Figure 4.6. FFT clustering example. The top panel is the raw observation field and the bottom panel is the observation field in frequency space after the FFT is applied..... 12
- Figure 4.7. The observation and forecast fields (red are observed objects greater than VIP 3) are transformed into three types of clusters that mimic CCFP climatological observed areas for sparse coverage/low confidence (light green), sparse coverage/high confidence (dark green), and medium coverage and higher (yellow)..... 13
- Figure 4.8. An example and description of box and whisker plots that will appear in different results throughout this report..... 14

Figure 4.9. The geographic domains used in the study.....	15
Figure 5.1. Comparison of CIWS analysis at VIP-level 3 valid at 2100 UTC for the summer period in 2010 (left) and 2011 (right).....	17
Figure 5.2. Comparison of the 6-h lead-time CoSPA forecast at VIP-level 3 valid at 2100 UTC for the summer period in 2010 (left) and 2011 (right).	17
Figure 5.3. Plots of VIP-level 3 convection over the CONUS for June, July, August, and September 2010 (left) and 2011 (right). CIWS analysis appears in cyan. The 6-h lead time was used for the CoSPA forecast (magenta) and the equivalent 9-h lead-time forecast of the HRRR (green).	18
Figure 5.4. Same as in Figure 5.3, except for September (left) and October (right) 2011.....	19
Figure 5.5. Same as in Figure 5.3, except for the NE Region.	20
Figure 5.6. Same as in Figure 5.3, except for the SE Region.....	21
Figure 5.7. Same as in Figure 5.3, except for the Western Region.....	22
Figure 5.8. CSI as a function of lead time and resolution during 2011 for June (top left), July (top right), August (bottom left) and September (bottom right) for CoSPA at VIP-level 3 issued at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.	23
Figure 5.9. As in Figure 5.8, but for VIP-level 2.....	24
Figure 5.10. CSI (left) and bias (right) as a function of lead time for CoSPA with a 1500 UTC issuance at VIP-level 3. Resolutions are similar to that of Figure 5.8.	24
Figure 5.11. CSI (left) and bias (right) as a function of lead time for CoSPA with a 1500 UTC issuance at VIP-level 2. Resolutions are similar to that of Figure 5.8.	25
Figure 5.12. CoSPA echo top forecast from 23 June 2011 issued at 1300 UTC, valid at 1700 UTC. CCFP polygons are overlaid. Notice most of the field is forecasting less than 30-kft echo tops.....	26
Figure 5.13. Bias as a function of lead time and resolution during 2011 for June (top left), July (top right), August (bottom left) and September (bottom right) for CoSPA at the 30-kft echo top threshold issued at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.....	27
Figure 5.14. As in Figure 5.13, but for the 2010 Season.....	28
Figure 5.15. CSI (left) and bias (right) as a function of lead time for CoSPA at 1300, 1500, and 1700 UTC issuances at the 30-kft echo top threshold. Resolutions are the same as in Figure 5.13.	28
Figure 5.16. CSI and bias as a function of lead time for the HRRR at 1300, 1500, and 1700 UTC issuances after 3-h latency is applied at the 30-kft echo top threshold for 2010 (left) and for 2011 (right). Resolutions are the same as in Figure 5.13.	29
Figure 5.17. Mean FSS for the NE as a function of resolution for CoSPA (blue), HRRR (green), CCFP (red), re-categorized CCFP (cyan), LAMP (magenta), climatology	

(black), and uniform (gray) for the 1300 UTC issuance valid at 1900 UTC 2011. Results account for the HRRR latency.....	30
Figure 5.18. As in Figure 5.17, but for 2010 in the NE.	30
Figure 5.19. As in Figure 5.17, but for 2011 in the SE.	31
Figure 5.20. As in Figure 5.17, but for 2010 in the SE.	31
Figure 5.21. As in Figure 5.17, but for the top 15 delay days in total minutes from 1 June 2011 to 30 September 2011 in the NE domain for the 1500 UTC issuance and 6-h lead time.	32
Figure 5.22. As in Figure 5.17, but for AFP days (top) and NE terminal GDP days (bottom) from 1 June 2011 to 30 September 2011 in the NE domain for the 1500 UTC issuance and 6-h lead time.	33
Figure 5.23. CSI as a function of FCI (constraint) threshold for 1 June – 30 September 2011 for the NE domain (left) and SE domain (right) at ARTCC scale. CoSPA in blue; HRRR in green; CCFP standard in red; CCFP re-categorized in cyan; LAMP in magenta. The gray-dashed line is the average number of ARTCC hexagons constrained by convection for the given threshold. Right of the yellow-dashed vertical line represents medium constraint; right of the maroon-dashed line represents high constraint. Dotted green, blue, red and cyan lines are confidence intervals.....	35
Figure 5.24. As in Figure 5.19, but for sector scales.....	35
Figure 5.25. CSI as a function of lead time for strategic telecon, pre-convective initiation hours (top 1100, 1300, 1500 UTC issue times) and for post-convective initiation hours (bottom; 1700, 1900, 2100 UTC), following the CCFP criteria of sparse coverage/low confidence and above. CoSPA is in red, LAMP in blue, and CCFP in green.	38
Figure 5.26. As in Figure 5.25, but for CCFP criteria of sparse coverage, high confidence and above.	39
Figure 5.27. Distribution of the area of the phase of precipitation for CIWS analysis, CoSPA VIL forecasts, and HRRR VIL forecasts for VIP-level 3 (left) and VIP-level 2 (right). Green is warm phase, magenta is mixed phase and cyan is cool phase (frozen).....	42
Figure 5.28. A comparison of the CIWS analysis with a VIP-level 3 threshold applied to the winter 2010-2011 study (left) valid at 2100 UTC, and the CoSPA forecast issued at 1500 UTC which is valid at 2011 UTC (right).	43
Figure 5.29. The distribution of echo tops for summer 2010 (left) and winter 2010-2011 (right) for CIWS (blue), CoSPA (green), and HRRR (red).....	43
Figure 5.30. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green). Climatology (black) and uniform (gray) are included for summer 2010 (top) and not included for the winter months (bottom).	45
Figure 5.31. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green). Climatology (black) and uniform (gray) are included for summer 2010 (top) and not included for the ‘like days’ during winter months (bottom).....	46

Figure 5.32. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green) during ‘like days’ in winter 2010-2011 for echo tops greater than or equal to 30 kft (top) and 25 kft (bottom). HRRR latency is accounted for..... 47

Figure 5.33: CSI as a function of FCI threshold (impact) for the summer 2010 (left) and winter 2010-2011 (right) for the SE U.S. at ARTCC scale for forecasts issued at 17 UTC and a 6-h lead time. CoSPA VIL is shown in blue and HRRR VIL is in green. The gray-dashed line referring to the right y-axis gives the average number of ARTCC hexagons impacted at the given threshold. The yellow-dashed vertical line is a medium impact threshold and the maroon-dashed line is a high impact threshold..... 48

Figure 5.34: CSI as a function of lead time and resolution for summer 2010 (left) and significant days in winter 2010-11 (right) for CoSPA at VIP-level 3 at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green. ... 49

Figure 5.35: Bias as a function of lead time and resolution for summer 2010 (left) and significant days in winter 2010-11 (right) for CoSPA at VIP-level 3 at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green. ... 49

List of Tables

Table 3.1. Re-categorization values for the 2-h CCFP for three coverage and confidence combinations, as listed..... 5

Table 3.2. VIP-levels and equivalent VIL values and radar reflectivity values (dBZ). Emphasis for this study is focused on a VIP-level 3 thresholds with additional examinations at VIL-level thresholds 2 and 4..... 6

Table 4.1. A table of dichotomous statistics used in the study with a description of the statistic..... 14

Table 5.1. Consistency for the 1700, 1900, 2100, and 2300 UTC valid times in the NE region when considering the 2-, 4-, and 6-h leads for CCFP re-categorized and CoSPA for thresholds of 0.01 (low constraint and above), 0.1 (moderate constraint and above), and 0.35 (severe constraint and above). 36

Table 5.2. As in Table 5.1, but for the SE region. 36

Table 5.3. Summary of the number of identified objects at the medium and above coverage threshold in the CIWS VIL analysis field and the CoSPA VIL forecast that are coincident in a CCFP sparse coverage, low confidence area for strategic issuance times at the 6-h lead time from 1 June-30 September. 41

Executive Summary

The CoSPA forecast product is an advanced high-resolution automated convective forecast being developed by the FAA to support the management and planning of aviation traffic flow that may be constrained or impacted by the presence of convective weather. The CoSPA algorithm is being considered for transition to FAA operations.

In support of this transition, the Quality Assessment Product Development Team (QA PDT) was tasked with independently assessing the quality of CoSPA with a particular focus on:

- Quality of CoSPA for use in traffic flow management (TFM),
- Quality of the modifications introduced into the 2011 version of the algorithm,
- Quality of CoSPA as a supplement to the Collaborative Convective Forecast Product (CCFP) since TFM planners are currently instructed to use CoSPA in conjunction with CCFP for developing traffic flow plans, and
- Quality during the winter months for capturing convective weather over the southeast U.S., since CoSPA is to be provided to users for decision-making year round.

With some modification, the verification framework used to assess the CoSPA algorithm in 2010 (Lack et al. 2011) is applied in this report. A variety of verification approaches and metrics are utilized for analysis. Results are stratified by region, season, and days with impact to the National Airspace (NAS). Two assessment periods are included in the study: the summer period 1 June – 30 September 2011, and the winter period 20 December 2010 to 28 February 2011. Some of the 2011 results are compared to the 2010 findings from Lack et al. 2011 to illustrate the relative improvement in CoSPA from the algorithm modifications. Note that a rigorous comparative baseline of forecast performance for 2011 as compared to 2010 could not be performed because the CoSPA algorithm elements and the base model (High Resolution Rapid Refresh; HRRR) were not re-run for the 2010 time period; therefore, all results presented in this report are only indicators of forecast improvement.

Summary of Significant Findings

- During high impact weather events, CoSPA provided improved weather quality over all other forecasts for traffic flow planning, particularly for the 0- to 2-h lead times and the 6- to 8-h lead times. This is especially true in the NE U.S. during high impact weather events that were weakly forced, harder-to-forecast events.
- With respect to enhancements to the CoSPA algorithm, the dynamic blending scheme improved CoSPA's overall performance, with greater improvement in the SE U.S. from better utilization of the HRRR information at 6- to 8-h lead times. However, significant underforecasting is evident in both the VIL and echo top fields in the 3- to 5-h time-frame over the CONUS. This underforecasting can manifest itself by decreasing forecast consistency and therefore planning confidence during a forecast cycle.
- As a supplement to CCFP, CoSPA improved the use of CCFP sparse coverage/low confidence polygons for traffic flow planning by decreasing false alarms for dense thunderstorm situations. This allows a forecast planner to act on sparse coverage/low confidence polygons with a higher level of confidence when CoSPA indicates severe convection in the region.
- During winter months, CoSPA has similar skill to that of the 2010 version of the algorithm, including during times of typical convective coverage found in the summer. However, echo tops tend to be lower during winter months. It is important to note that CoSPA currently does not explicitly display binned echo tops below 30 kft, which may hinder air traffic planning at lower flight levels.

1 Introduction

The CoSPA forecast product is an advanced high-resolution automated convective forecast being developed by the FAA to support the management and planning of aviation traffic flow that may be constrained by the presence of convective weather. The CoSPA algorithm is being considered for transition to FAA operations.

The goals of this evaluation are to assess: 1) the quality of CoSPA for use in traffic flow planning with a particular focus on the quality of the modifications introduced into the 2011 version of the algorithm, 2) the quality of CoSPA as a supplement to the Collaborative Convective Forecast Product (CCFP), since TFM planners are currently instructed to use CoSPA in conjunction with CCFP for developing traffic flow plans, and 3) the quality of CoSPA during the winter months for capturing convective weather over the southeast U.S., as CoSPA may be provided to users for decision-making year round.

The report is organized into six sections. Section 2 outlines the assessment approach. Section 3 describes the different data types utilized in this evaluation, while the methods and techniques are detailed in Section 4. The results are presented in Section 5, and the conclusions are highlighted in Section 6.

2 Approach

With some modification, the framework used to assess the CoSPA algorithm for the 2010 evaluation (Lack et al. 2011) is applied in this report. The framework includes an initial investigation of the forecast and observation climatology to determine characteristic differences between forecast products and between the forecast products and the observations. Results from the climatology provide the necessary information to establish meaningful thresholds and to highlight areas of interest for the main assessment. The main assessment includes three primary areas of investigation:

- A relative comparison of CoSPA quality in 2011 versus 2010 in order to evaluate the skill of modifications applied to CoSPA in 2011, with a particular focus on:
 - Forecast consistency
 - Forecast blending
 - Improvement over the HRRR (High Resolution Rapid Refresh, parent model to CoSPA)
 - Quality on high impact days
- Performance of CoSPA as a supplement to CCFP,
- Performance of CoSPA during the winter months for convection over the SE U.S.

A variety of metrics and verification approaches are applied to the assessment of CoSPA in order to meet the goals stated in the introduction. Techniques include:

- **Upscaling** for assessing high-resolution forecasts, as spatial accuracy is difficult to achieve at native resolution,
- **Fractions Skill Score** for assessing forecasts of different temporal and spatial resolutions at the same set of resolutions,
- **Flow Constraint Index** for assessing forecasts at different temporal and spatial resolution after information has been translated to an operationally meaningful constraint, in this case the flow constraint imposed by convective weather,
- **Forecast Consistency** for measuring a forecast's consistency within its issuances and leads, and
- **Clustering** for measuring forecast and observation objects at scales that are meaningful for aviation traffic flow management (TFM) and planning.

The skill scores are stratified by region, season, strategic planning telecon times, aviation impact as measured by Air Space Flow Programs (AFPs) and Ground Delay Programs (GDPs), as well as pre- and post- convective initiation times.

Note: Since a true performance baseline is costly to achieve and was not available from 2010 to 2011, the findings presented in this report should be interpreted as relative measures of forecast quality and indicators of forecast improvement.

3 Data

Data were collected for analysis from 1 June to 30 October 2011 for the summer (hereafter, summer 2011) and from 20 December 2010 to 28 February 2011 for the winter 2010-2011 (hereafter, winter 2010-11). Although the HRRR model changed slightly between 1 June and 7 July 2011, results indicate little change in forecast skill in the CoSPA product between June and the other months of the study (see Figure 5.8 and Figure 5.9 for monthly plots of skill). Therefore, the period of performance for this assessment included June so that relative comparisons could be performed on the results computed in 2010. The forecasts included in the assessment are CoSPA, CCFP, and the Localized Aviation MOS (Model Output Statistics) Program (LAMP). LAMP is included in this study for comparison purposes, as it was used by the TFM planners at the Air Traffic Control Systems Command Center (ATCSCC). The 'truth' field is represented by CIWS (Corridor Integrated Weather System).

3.1 CoSPA

CoSPA is an automated convective forecast produced over the Contiguous United States (CONUS). Convective forecasts of vertically integrated liquid water (VIL) and echo tops at 1 km resolution from 0- to 8-h are provided. The forecasts are produced every 15-min (Wolfson et al. 2008), but for this assessment only hourly forecasts are evaluated. CoSPA consists of three main components: (1) an extrapolation forecast provided by CIWS; (2) a high-resolution numerical weather prediction (NWP) model provided by the HRRR, and (3) a blending algorithm. It is important to note that the 0- to 2-h CoSPA forecast is simply the extrapolation forecast produced by CIWS. An example of the CoSPA display is shown in Figure 3.1. Changes introduced into the 2011 version of CoSPA are discussed in detail by Iskenderian (2011). However, highlights impacting the assessment are listed below:

- Modifications to the blending algorithm to include dynamic weights used for combining the 0- to 2-h CoSPA extrapolation and the HRRR forecasts to form the final CoSPA forecast
- Improvements in CIWS storm extrapolation and echo top decay parameters
- Changes to the HRRR

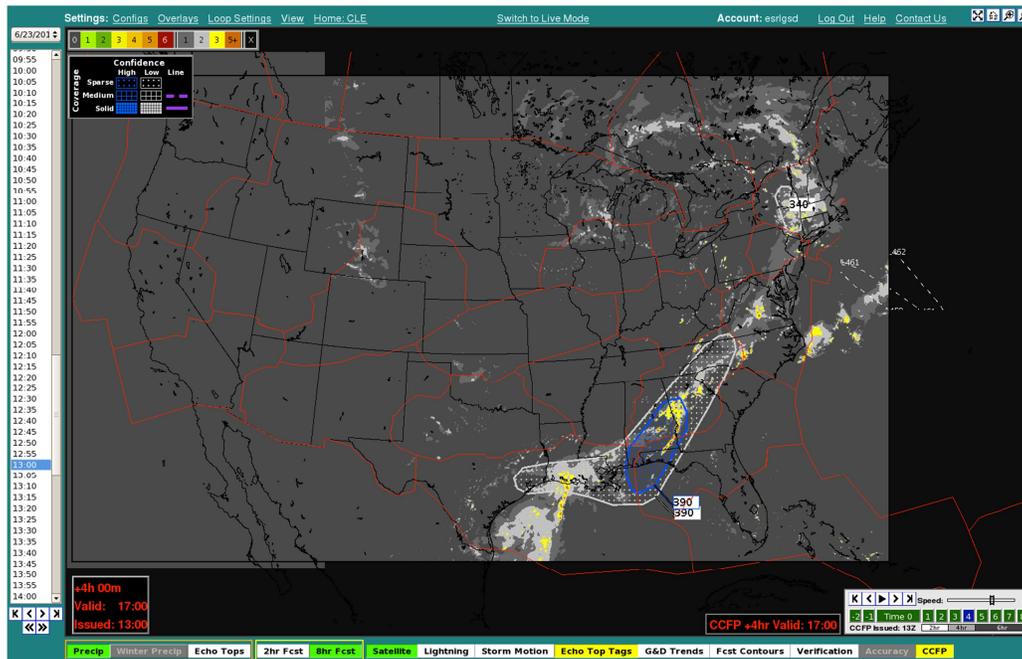


Figure 3.1. Graphic of CoSPA VIL from 23 June 2011 issued at 1300 UTC, valid at 17 UTC (grey VIP-levels 1 and 2, yellow VIP-level 3 and 4, red VIP-level 5 and greater), CCFP polygons (key in upper-left hand corner of graphic), and echo tops (light purple for tops less than 30 kft, and dark purple for echo tops 40 kft and above at 5-kft intervals).

3.2 High Resolution Rapid Refresh (HRRR)

The HRRR model is the parent model used to bridge from the 2-h CIWS forecast to the 4- to 8-h forecast provided by CoSPA (Weygandt et al. 2010). The HRRR model is available hourly with 15-min lead-time increments and it provides both VIL and echo top fields. Changes to the HRRR between 2010 and 2011 were significant. The boundary conditions providing basic information to the HRRR switched between 2010 and 2011 from the Rapid Update Cycle (RUC) model to the WRF Rapid Refresh (RAP) model. In addition, a moisture nudging routine was added to the HRRR during the evaluation, but it was deemed to have little impact on aggregate statistics as measured before and after the change.

3.3 Collaborative Convective Forecast Product (CCFP)

The CCFP is the primary forecast used by Air Traffic Control System Command Center (ATCSCC) traffic flow managers for planning routes in response to convective weather impacts. Therefore, CCFP is used in this assessment as a standard of reference or ‘performance bar’ for judging the quality of CoSPA.

A depiction of the CCFP is shown in Figure 3.2. In the assessment, CCFP forecasts are evaluated in two ways: 1) strictly according to the forecast definition and 2) as a re-categorized (sometimes referred to as calibrated) forecast. In the second case, the forecast coverage categories provided by the CCFP are re-categorized to closely align with climatological findings. The re-categorized values for the 2-h CCFP for the various coverage/confidence thresholds are listed in Table 3.1. Since there are few changes in observed coverages between 2010 and 2011, the re-categorized values computed in 2010 are also used in the 2011 assessment. It is important to note that no changes were applied to the CCFP product definition from 2010 to 2011.

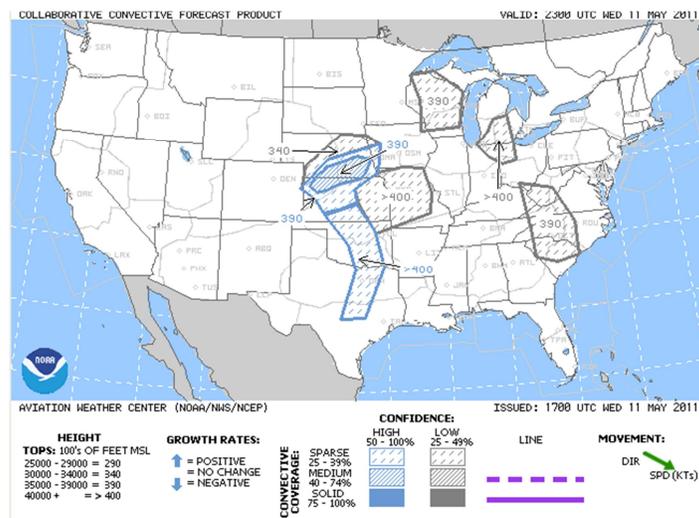


Figure 3.2. An example of CCFP for 11 May 2011 issued at 1700 UTC and valid at 2300 UTC.

Table 3.1. Re-categorization values for the 2-h CCFP for three coverage and confidence combinations, as listed.

Area Category	Coverage Definition	2010 Calibrated Coverage	2011 Calibrated Coverage
Sparse Coverage/ Low Confidence	25-39%	4.64%	5.16%
Sparse Coverage/ High Confidence	25-39%	9.77%	11.66%
Medium Coverage and higher	40-74% 75-100%	20.38%	21.68%

3.4 The Localized Aviation MOS (Model Output Statistics) Program (LAMP) Thunderstorm Product

Because LAMP was a part of the 2010 study, it was introduced into the current study for consistency purposes only. 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 National Weather Service’s (NWS) National Digital Forecast Database (NDFD) 5-km grid, with hourly updates, and forecast lead times from 1 to 25 h. An example of the LAMP probabilistic product is shown in Figure 3.3.

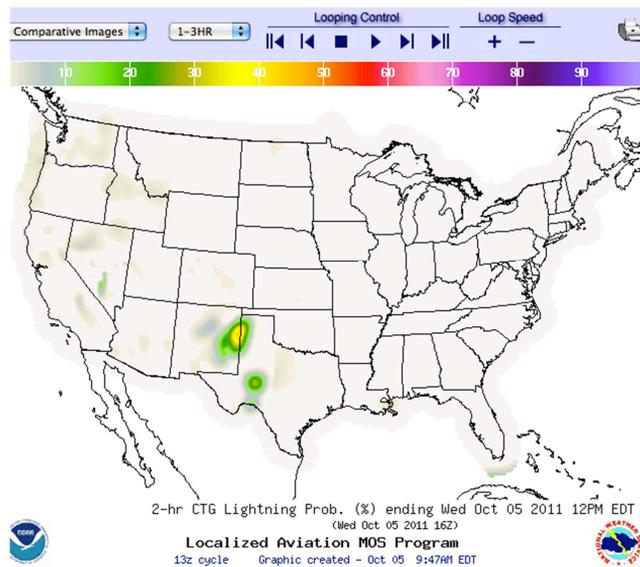


Figure 3.3. The operational LAMP Thunderstorm Product on 5 October 2011 issued at 1300 UTC valid at 1600 UTC.

3.5 CIWS Observations

The CIWS analysis is used as the truth field to verify the quality of the weather forecasts. CIWS has a 2.5-min update cycle, is available at 1-km horizontal resolution, and includes an analysis of VIL and echo top data (Dupree et al. 2009). The VIL and echo top fields will be evaluated independently. Values of VIP-level 3 and greater are considered to represent locations of significant convection, and are therefore primary to this assessment. Equivalent radar reflectivity and VIL values for a given VIP-level are shown in Table 3.2. Echo top information is visualized in the CoSPA displays in 5-kft bins beginning at 30 kft and ending at greater than 40 kft; therefore, for much of the study, echo tops will be examined using these bins. It is important to note that additional thresholds were applied throughout this study for both VIL and echo tops. The additional thresholds not appearing in this report are available upon request.

Table 3.2. VIP-levels and equivalent VIL values and radar reflectivity values (dBZ). Emphasis for this study is focused on a VIP-level 3 threshold with additional examinations at VIL-level thresholds 2 and 4.

VIP-level	VIL (kg m ⁻²)	dBZ
0	0.05	<18
1	0.15	18
2	0.76	30
3	3.5	40
4	6.9	46
5	12.1	50
6	>31.6	57

4 Methods

The following sections will discuss the several components that are included in this 2011 CoSPA evaluation. Discussions of diagnostic techniques and advanced metrics for indicating forecast quality will be included, as well as an introduction to the plots and statistics that will be included in the results.

4.1 Climatology

The climatological overview of the forecasts and observations has significant value when qualitatively assessing the coarse spatial and temporal performance of the products, allowing one to gain insight into large-scale differences between forecasts and observations. The climatological grids are created by averaging the occurrence of convection at each grid box for the set of days used in the study, for each forecast product and for the observation set (CIWS). A Gaussian smoothing operator is applied to the observations and to the forecast grids of averages to retain the systematic signal. The grids are then normalized to a common color scale for ease of comparison.

4.2 Upscaling

The primary use of the upscaling technique is to diagnose changes in forecast skill (represented by CoSPA, CCFP and LAMP) with changes in threshold and forecast lead time. This technique is applied to each independent forecast for both the VIL and echo top fields (if both fields exist). The upscaling technique is most useful for assessing high-resolution forecasts where co-location of forecasts and observations is difficult to achieve. The basic mechanics of upscaling includes coarsening a high-resolution forecast and observation by using a representative characteristic of the points within a neighborhood, typically by applying the mean, median, or maximum. An example of upscaling appears in Figure 4.1.

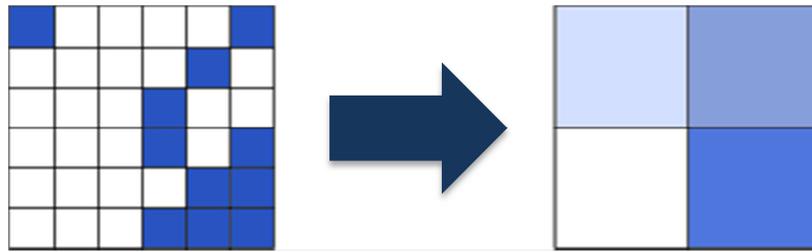


Figure 4.1. An example of upscaling a 6x6 grid to a 2x2 grid. Taking the average of the 4 3x3 calls on the left creates the 4 pixels on the right.

4.3 Fractions Skill Score

The Fractions Skill Score (FSS), described by Roberts and Lean (2005) is used in this study as a meteorological translation evaluation tool. Similar to the upscaling technique, the FSS is commonly used to assess the skill of high-resolution numerical weather prediction (NWP) models at various resolutions. Unlike the upscaling technique, the FSS allows for the comparison of both deterministic and probabilistic forecasts, placing each on a level playing field. The FSS allows for the comparison of the percent coverage of the forecast to the percent coverage of the observations for a given neighborhood about a reference pixel for all pixels in the forecast field. The FSS is given by equation (1), and is defined as the average sum squared difference of the percent coverage in the forecast and observations, divided by the average sum of the squares of the percent coverage of the forecast and observations. The FSS has a valid range between 0 (worst) and 1 (best), where values over a defined baseline are said to have skill.

$$FSS = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (P_{fct} - P_{obs})^2}{\frac{1}{N} \sum_{i=1}^N P_{fct}^2 + \frac{1}{N} \sum_{i=1}^N P_{obs}^2} \quad (1)$$

An example of how to calculate percent coverage in a domain is shown in Figure 4.2. For this example, a 5x5 neighborhood is created around the center pixels for the upper left and upper right images in Figure 4.2. The observation at the center pixel receives a value of 0.32 (P_{obs} , upper left) and the forecast at the center pixel receives a value of 0.44 (P_{fcst} , upper right). This procedure is repeated for all pixels in the native domain and the results are input into equation 1 for the calculation of the FSS for a 5x5 neighborhood.

Figure 4.2 also shows a forecast of constant probability, referred to as a uniform forecast (lower left), and a CCFP-like forecast (lower right). For demonstration purposes, the uniform forecast and CCFP-like forecast are such that the bias of each new forecast matches the original bias created for the deterministic forecast (bias=19/21). Additionally, the coarse CCFP-like forecast and deterministic forecast represent approximately the same region of the domain. The fractionalized grid for each of the forecasts is shown in Figure 4.3 for a 3x3 neighborhood. Transforming both deterministic and probabilistic forecasts into fractionalized space allows for direct comparisons to be made at varying neighborhood radii.

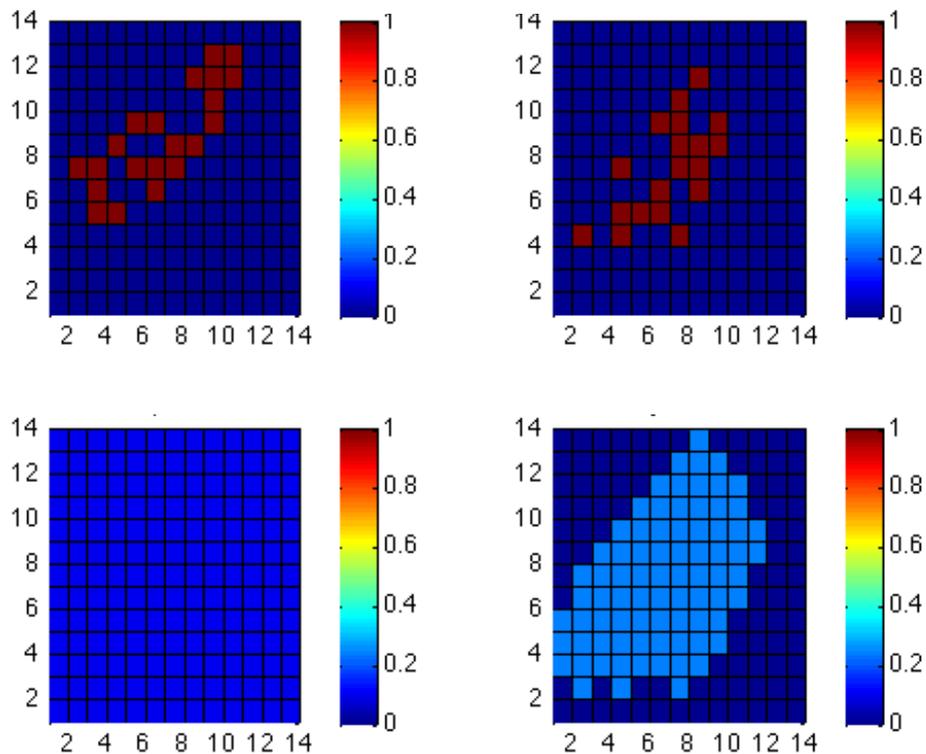


Figure 4.2. Illustration of the FSS. Observation field (top left), deterministic forecast (top right), a uniform forecast (bottom left), and a CCFP-like forecast (bottom right).

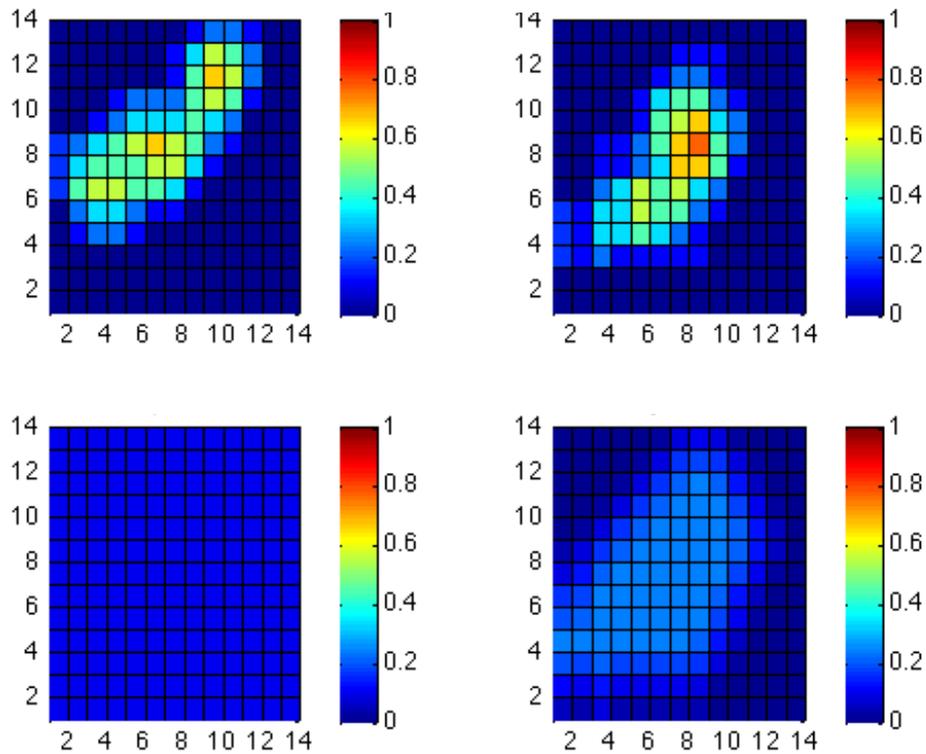


Figure 4.3. Fractionalized grid for the 3x3 neighborhood for the observation field (top left), deterministic forecast (top right), uniform forecast (bottom left), and CCFP-like forecast (bottom right).

4.4 Flow Constraint Index

The FCI is considered to be an operationally relevant translation metric since it has been shown to have a relationship to strategic traffic management initiatives (TMIs; Layne et al. 2012). The FCI methodology was adapted by Layne and Lack (2010) from the Mincut-Bottleneck technique introduced for TFM by Krozel et al. in 2004.

To begin, consider a constraint field representing potential traffic flow restriction through a portion of the airspace due to the presence of a particular weather hazard, such as convection. The traffic flow constraint is determined using a class of mathematical algorithms known as the Mincut Max-flow (MCMF), developed as a part of graph theory (Ford and Fulkerson, 1956). The FCI is a specific implementation of the MCMF approach for weather, where weather can be either forecast or observed. Any given portion of the airspace can be treated as a corridor through which air traffic travels; the sides of the corridor comprise one or more connected line segments as part of a geometric shape (Figure 4.4). Significant weather located within the corridor will impact the flow of traffic through the corridor. The FCI is a measure of the reduction in the potential flow through the corridor, and is independent of the actual traffic flow.

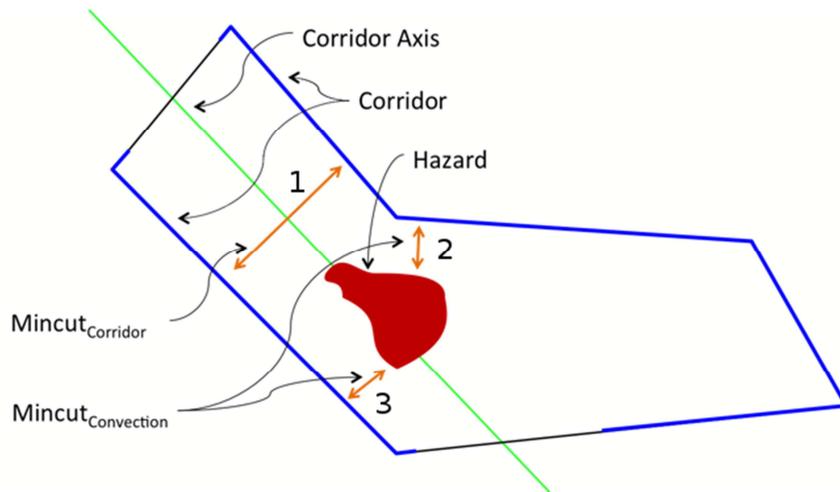


Figure 4.4. Conceptual model of the FCI. Blue lines represent the corridor boundaries and the red area denotes an area of hazardous convection. Arrow 1 represents the minimum distance across the corridor in the absence of convection. Arrows 2 and 3 show the minimum distance across the available airspace around a hazard.

To calculate FCI given a polygon defining the bounds of a corridor, Mincut calculations are performed for the corridor itself and for the corridor with hazards included. These two Mincut values are then combined to produce the FCI, according to (2).

$$FCI=1-\frac{Mincut_{convection}}{Mincut_{corridor}} \quad (2)$$

For this study, two hexagon geometries of size 75 NM and 300 NM are used to compute the FCI. The 75-NM hexagon approximates the size of the average super-high altitude sector and the 300-NM hexagon approximates the size of Air Route Traffic Control Centers (ARTCCs). Figure 4.5 shows an example of the hexagonal shape. Removing a pair of opposing sides of the hexagon creates a corridor; the flow restriction is determined for each of the three corridors, yielding three FCI values for the hexagon. The elongated area of convection, shown in red in Figure 4.5 and oriented from northwest to southeast, restricts 75% of the airspace for planes attempting to travel from the southwest the northeast. Because of the northwest-southeast orientation and location of the convection, less than half of the potential flow of the north-south corridor is constrained, and nearly zero constraint is found for traffic moving from northwest to southeast. Each of the three FCI values are represented by the length of the lines, as a fraction of the distance from opposing corners plotted within the hexagon (see right side of Figure 4.5). FCI can easily be calculated for both probabilistic and deterministic forecasts and observations (Layne and Lack 2010).

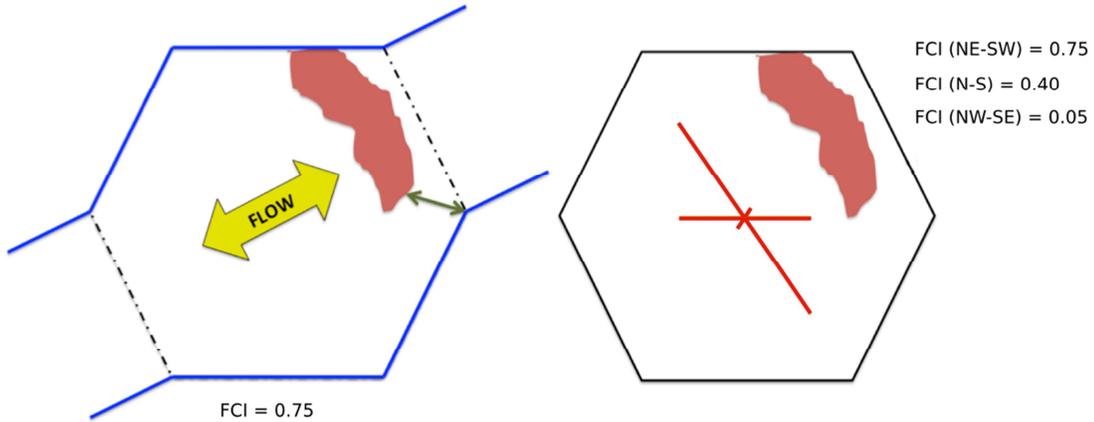


Figure 4.5. Illustration of the FCI concept for a hexagonal geometry. The hexagon contains three separate corridors, one for each pair of opposing faces: traffic moving from northeast to southwest, from north to south, and from northwest to southeast. (The FCI is identical for traffic flowing in the opposite directions.). A weather hazard is denoted by the red area. The green arrow (left) shows the mincut distance for the northeast-to-southwest corridor. The length of the red lines (right; as a fraction of the total corner-to-corner distance) represent the FCI value for traffic moving perpendicular to the line.

4.5 Forecast Consistency

Comments during the 2010 operational evaluation period prompted the Quality Assessment Product Development Team (QA PDT) to provide a quantitative metric for measuring forecasting consistency. The Correspondence Ratio (*CR*; equation 3; Stensrud and Wandishin (2000)) as applied here is the ratio of intersection and union over a set of a gridded forecast issuance and lead times, associated to specific valid times. The *CR* applied in this analysis measures the consistency between the issuance and lead times of a forecast, and is not a measure of accuracy because it does not utilize observational data.

$$CR = \frac{I}{U} = \frac{\sum_{i=1}^n f_{1,i} \cap f_{2,i} \cap \dots \cap f_{m,i}}{\sum_{i=1}^n f_{1,i} \cup f_{2,i} \cup \dots \cup f_{m,i}} \quad (3)$$

The *CR* was computed for multiple FCI thresholds (i.e., any impact, medium impact, and high impact) for issue and lead times relevant to traffic flow planning for CCFP, CoSPA, HRRR, and LAMP. For example, the *CR* will be calculated using three CCFP forecasts with 2-, 4-, and 6-h lead times all valid at 2100 UTC, using the hexagonal grid of FCI values exceeding the medium impact threshold of 0.10. If all medium impact threshold hexagons overlap for the three forecasts valid at 2100 UTC, the *CR* will have a value of 1, indicating perfect consistency of the forecast. Several variations were applied to the consistency formulation: 1) a strict definition requiring all forecasts to exceed a selected threshold at a given valid time; 2) a looser definition requiring 2/3 of the forecasts to exceed a selected threshold for a

specific valid time, and; 3) the strict definition combined with credit for being perfectly consistent for non-events (e.g. forecasts of no constraint, which is the correct negative case).

4.6 Clustering

Closely following the work of Lack et al. (2010a), CoSPA, LAMP, and CIWS clustering to the size of CCFP objects is done using a Fast Fourier Transform (FFT) methodology. FFT band passes are used to convert spatial intensity to spatial frequency (Lack et al. 2010b). An example of this clustering technique for radar reflectivity over Texas is shown in Figure 4.6. For each cluster exceeding a minimum size criterion of 3000 sq mi (min. size criteria for a CCFP polygon), the percent of convective coverage within the cluster is calculated. The amount of coverage is assigned to one of three coverage categories (sparse/low, sparse/high and medium and above) coinciding with the coverage category definitions for CCFP to allow direct comparisons between forecasts and observations.

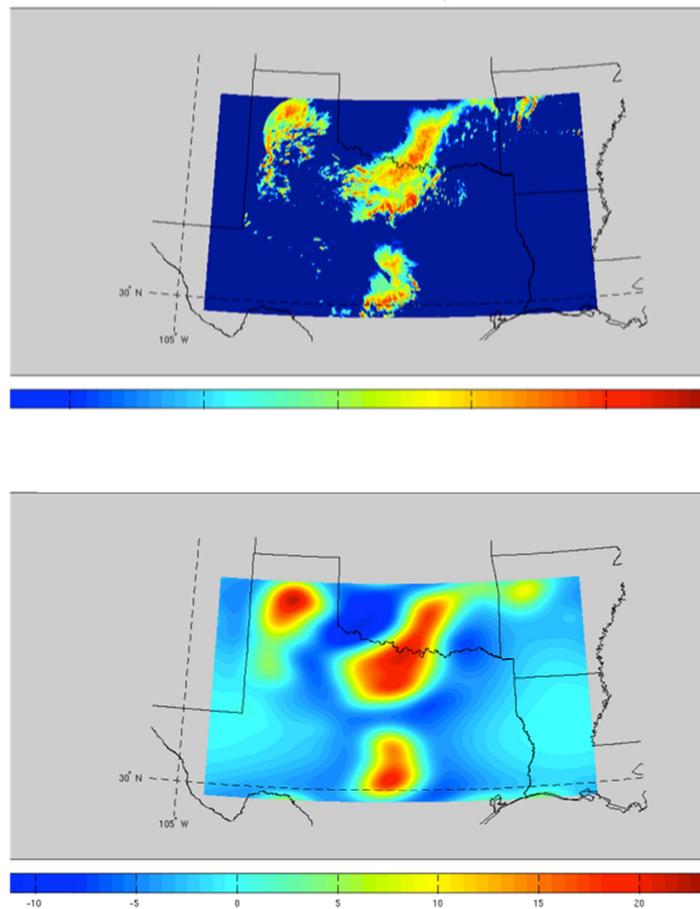


Figure 4.6. FFT clustering example. The top panel is the raw observation field and the bottom panel is the observation field in frequency space after the FFT is applied.

An example of the CIWS analysis field, after the clustering technique has been applied, is shown in Figure 4.7. Using this technique, areas with a strong frequency signal above the VIP 2 threshold, as measured by the FFT, are identified into one of the three coverage categories listed above.

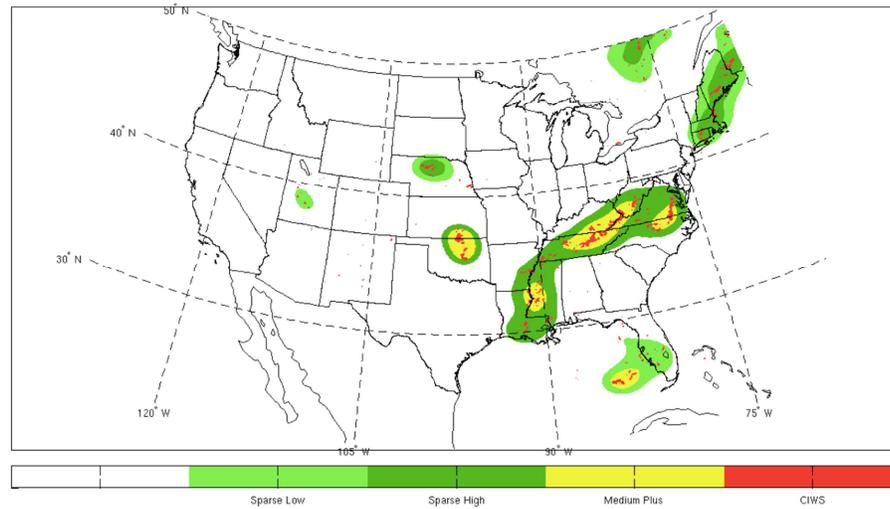


Figure 4.7. The observation and forecast fields (red are observed objects greater than VIP 3) are transformed into three types of clusters that mimic CCFP climatological observed areas for sparse coverage/low confidence (light green), sparse coverage/high confidence (dark green), and medium coverage and higher (yellow).

4.7 Statistics

Table 4.1 lists the dichotomous statistics calculated for the techniques described in the previous sections. The statistics are derived from a standard 2x2 contingency table, and include probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and bias.

Table 4.1. A table of dichotomous statistics used in the study with a description of the statistic.

Statistic	Formula	Description
POD	$\frac{hits}{hits + misses}$	The fraction of the observed "yes" events that were correctly forecast.
FAR	$\frac{false\ alarms}{false\ alarms + hits}$	The fraction of the predicted "yes" events actually did not occur (i.e., were false alarms).
CSI	$\frac{hits}{hits + misses + false\ alarms}$	Measures the fraction of observed and/or forecast events that were correctly predicted.
Bias	$\frac{hits + false\ alarms}{hits + misses}$	Indicates whether the forecast system has a tendency to underforecast (BIAS<1) or overforecast (BIAS>1) events.

Many of the statistics presented in this report are conveyed through the box and whisker plot, which is described in Figure 4.8.

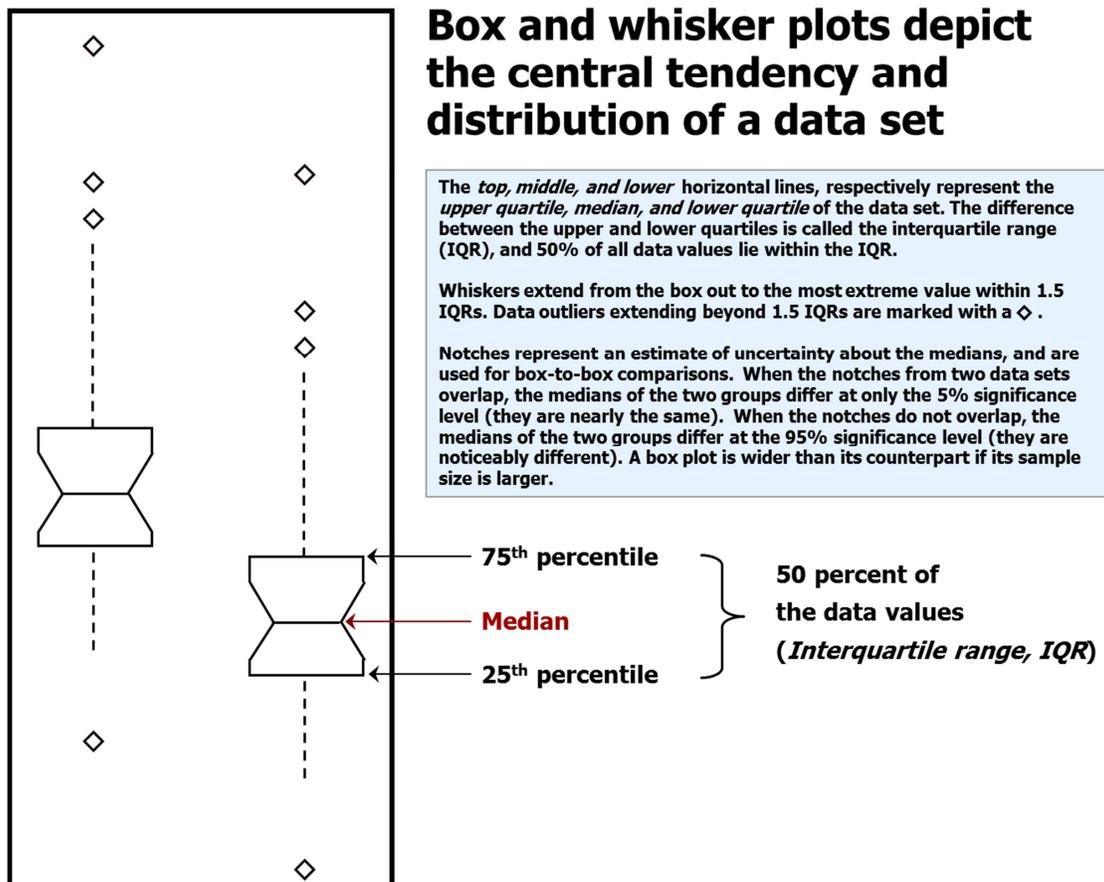


Figure 4.8. An example and description of box and whisker plots that will appear in different results throughout this report.

4.8 Stratifications

Primary stratifications used in this study include:

- Strategic issuances and lead times with a particular emphasis on the 1100, 1300, and 1500 UTC issuances and the 4- to 8-h lead times. Other thresholds and issuance/lead times were examined for completeness and are available upon request.
- Hazardous convection identified by VIP values greater than or equal to 3 (equivalent to 40 dBZ or VIL 3.5 kg m^{-2}).
- Geographic stratifications as shown in Figure 4.9. Regions are divided into three main areas: northeast (NE) with the highest traffic density and frontally forced convection, southeast (SE) with airmass-type convection, and west (W) with convection driven by large-scale circulations (i.e. AZ/NM monsoon).
- Seasonal stratification.
- Days with high impact to the NAS, as measured by Traffic Management Initiatives, including Airspace Flow Programs (AFPs) and Ground Delay Programs (GDPs).

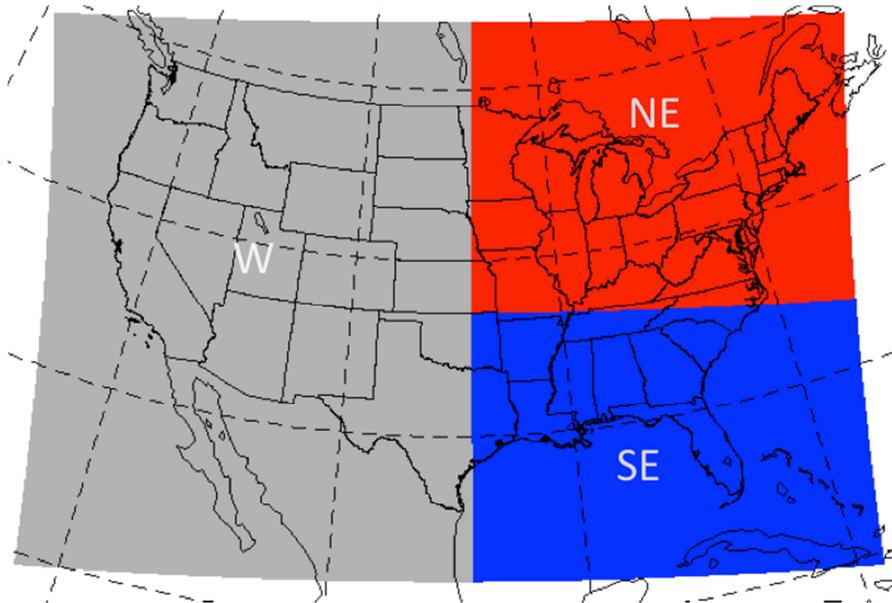


Figure 4.9. The geographic domains used in the study.

5 Results

The analysis of results strives to address the following themes:

- Characteristics of CoSPA as represented by climatology.
- Performance of the 2011 version of CoSPA as it compares to the 2010 version to evaluate modifications applied to CoSPA in 2011 with a focus on:
 - Forecast consistency
 - Forecast blending
 - Improvement over the HRRR (High Resolution Rapid Refresh, parent model to CoSPA)
 - Quality on high impact days
- Performance of CoSPA as a supplement to CCFP.
- Performance of CoSPA during the winter months for convection over the SE United States.

5.1 Climatological Analysis

Prior to beginning the analysis of forecast improvements, it is necessary to investigate seasonal convective activity to provide additional context and understanding for the forecast comparisons. Figure 5.1 presents a comparison of CIWS analyses for 2010 (left) vs. 2011 (right), each with a VIP-level 3 threshold and valid at 2100 UTC, considered to be the time at which convection reaches a maximum. The broad-scale convective picture for the CONUS between the two years is nearly identical, particularly over the SE U.S. There is a slight shift in observed convective activity eastward away from the Central Plains from 2010 to 2011. However, significant differences from 2010 to 2011 in the amount of convection produced by CoSPA are evident in Figure 5.2. Although CoSPA underforecasts the overall convective activity in both 2010 and 2011, the placement and amount of convection produced by CoSPA in 2011 was more representative of the observed convection.

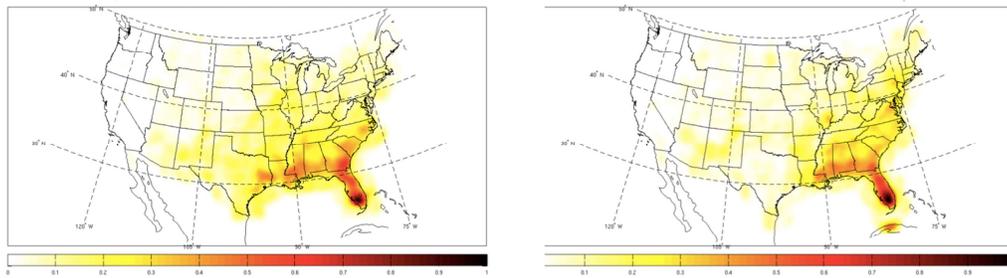


Figure 5.1. Comparison of CIWS analysis at VIP-level 3 valid at 2100 UTC for the summer period in 2010 (left) and 2011 (right).

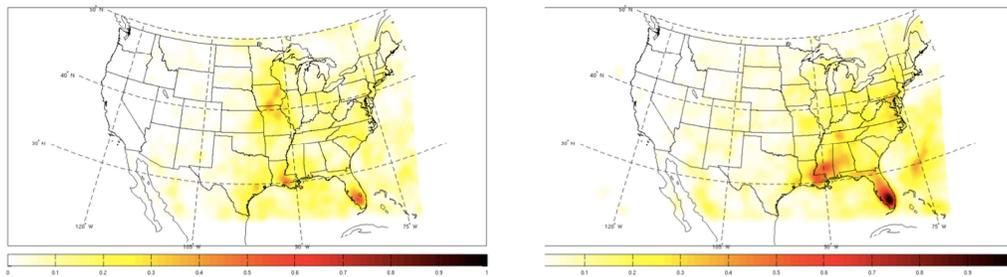


Figure 5.2. Comparison of the 6-h lead-time CoSPA forecast at VIP-level 3 valid at 2100 UTC for the summer period in 2010 (left) and 2011 (right).

5.1.1 Diurnal Convective Signal

Monthly analyses of convective coverage for 2010 and 2011 are presented in the following section. Figure 5.3 shows the coverage of convection at VIP-level 3 from CoSPA, HRRR, and CIWS observations for 2010 (left) and 2011 (right). The 6-h lead time was selected for the CoSPA forecast and the equivalent 9-h lead-time forecast was selected for the HRRR, given its 3-h latency. This plot shows that overall bias for both CoSPA and the HRRR improved significantly from 2010 to 2011 for June-September. The difference in convective coverage between CoSPA and CIWS is reduced for each month in 2011 as compared to 2010. In addition, the convective lag evident during initiation time periods in 2010 decreased in 2011 for CoSPA, indicating that forecast convection in 2011 was more coincident with the onset of convection than it was in 2010. Results for September and October 2011 are presented in Figure 5.4. Notice the overall decrease in convective coverage from September to October 2011 in all regions. With little convective activity in October, results for October 2011 are excluded from further analyses in this report.

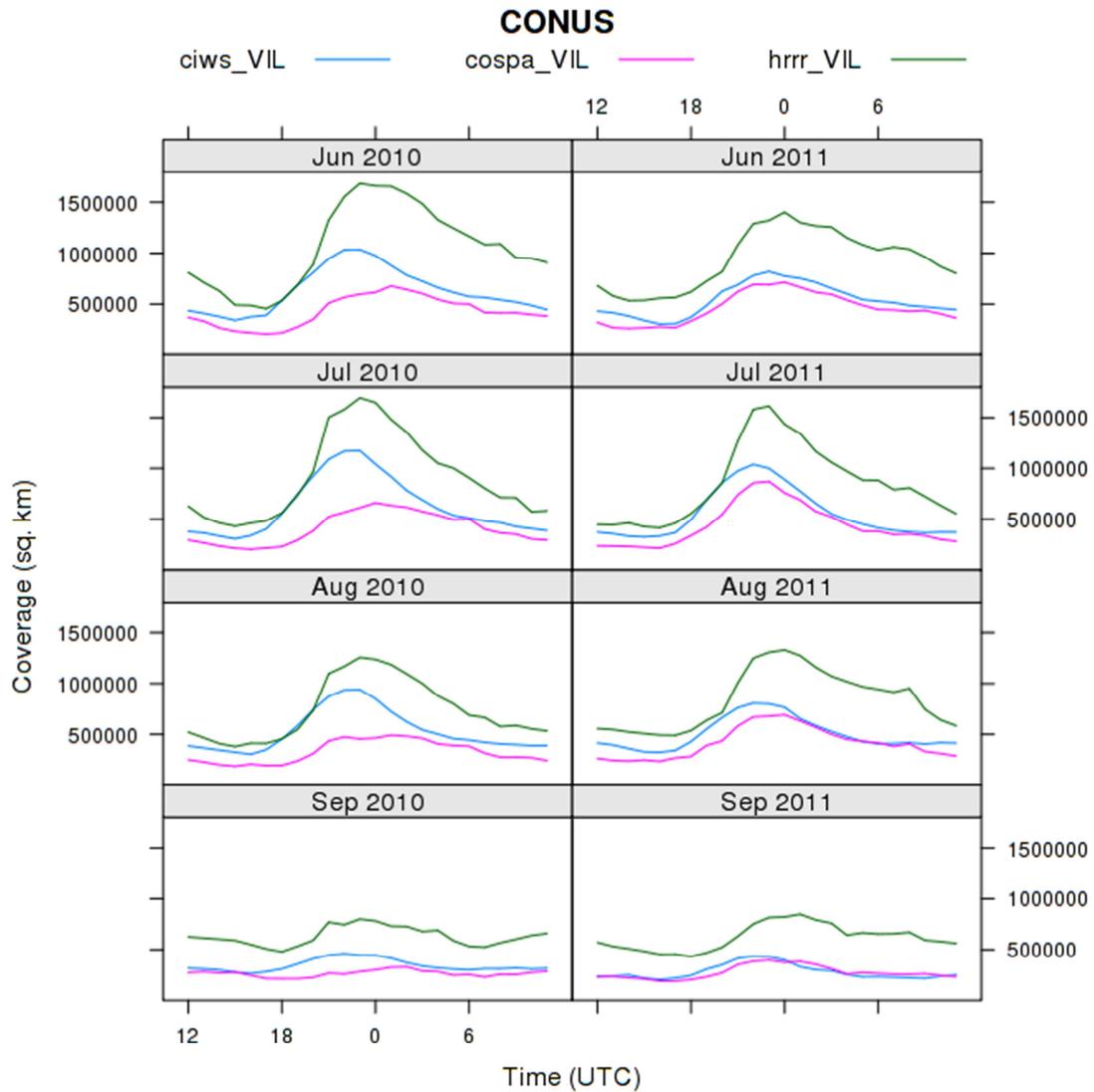


Figure 5.3. Plots of VIP-level 3 convection over the CONUS for June, July, August, and September 2010 (left) and 2011 (right). CIWS analysis appears in cyan. The 6-h lead time was used for the CoSPA forecast (magenta) and the equivalent 9-h lead-time forecast of the HRRR (green).

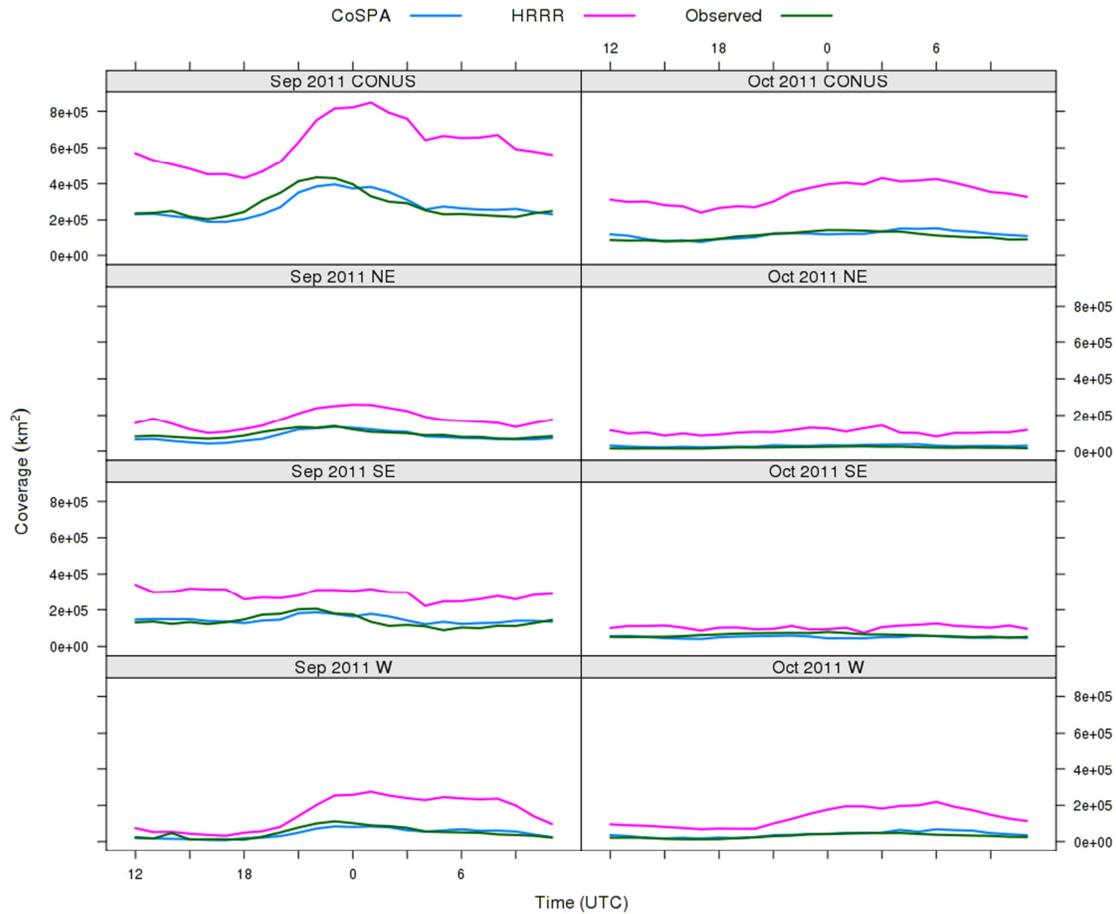


Figure 5.4. Plots of VIP-level 3 convection over the CONUS for September (left) and October (right) 2011. CIWS analysis appears in green. The 6-h lead time was used for the CoSPA forecast (cyan) and the equivalent 9-h lead-time forecast of the HRRR (magenta).

5.1.2 Regional Forecast Differences

In the 2010 assessment, an underforecasting weakness was identified in CoSPA over the Southeast (SE) U.S., while CoSPA convection in the Northeast (NE) was nearly the same as that which was observed. In order to improve the underforecasting in the SE, CoSPA developers modified the blending scheme from static to dynamic in the 2011 version of CoSPA. To investigate these enhancements, analysis of regional differences in the forecasts from 2010 to 2011 is presented in the following section.

Overall results, shown in Figure 5.5 and Figure 5.6, for 2011 CoSPA stratified by region (NE and SE) indicate that while CoSPA keeps the convective coverage nearly the same for the NE region, improvements in the underforecasting of the convection in CoSPA in the SE did occur (CoSPA now more similar to CIWS). Additionally, it is noted that the high NE bias in the HRRR during June and July 2010 was reduced in 2011 (Figure 5.5), but the peak of convection produced by the HRRR in the NE often

lagged the onset of observed convection, most notably in August 2011. CoSPA also exhibited a small lag in the onset of convection, particularly during July and August 2011 in the NE. The lag noted in the SE for the 2010 CoSPA was greatly reduced in 2011.

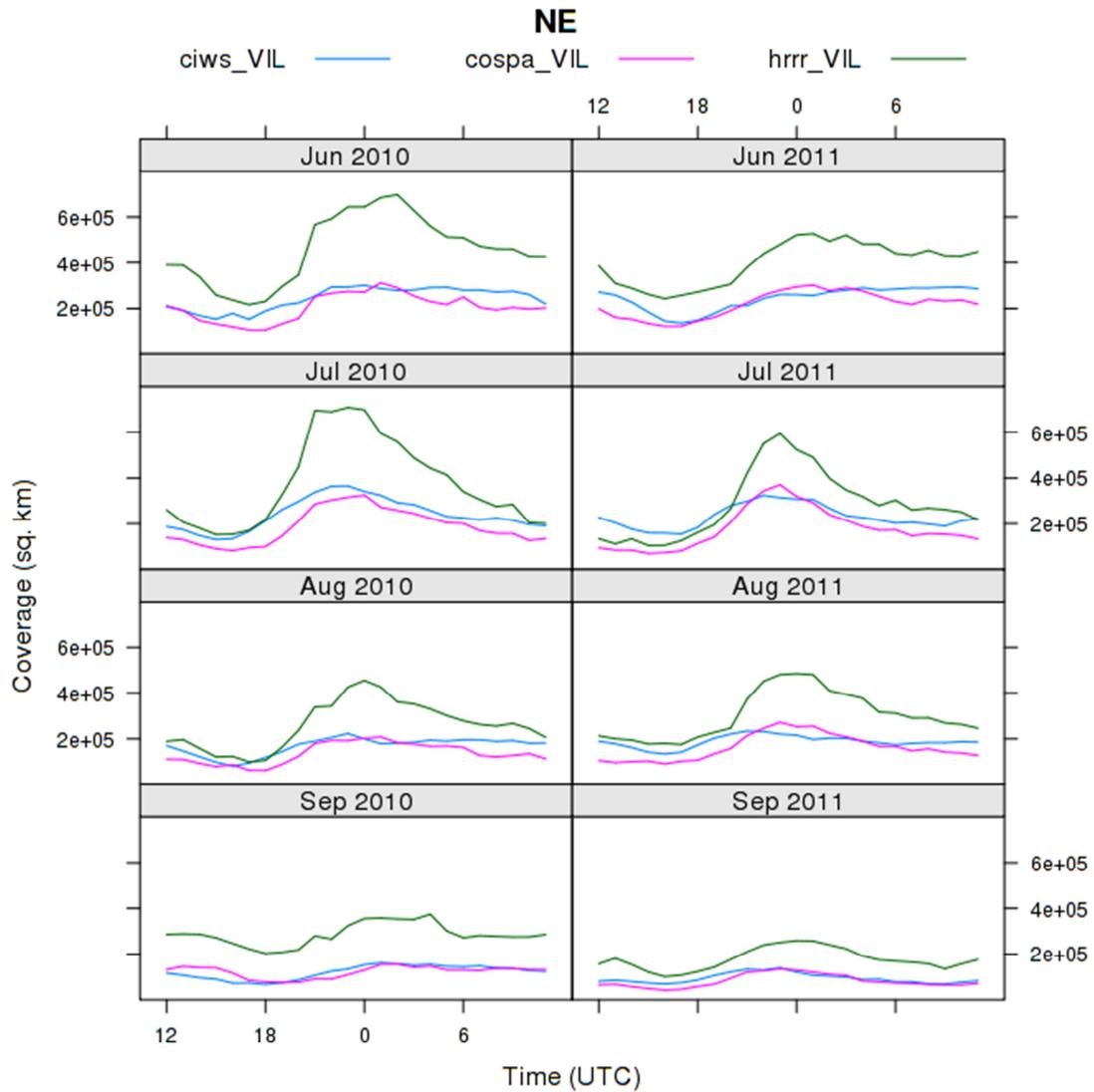


Figure 5.5. Same as in Figure 5.3, except for the NE Region.

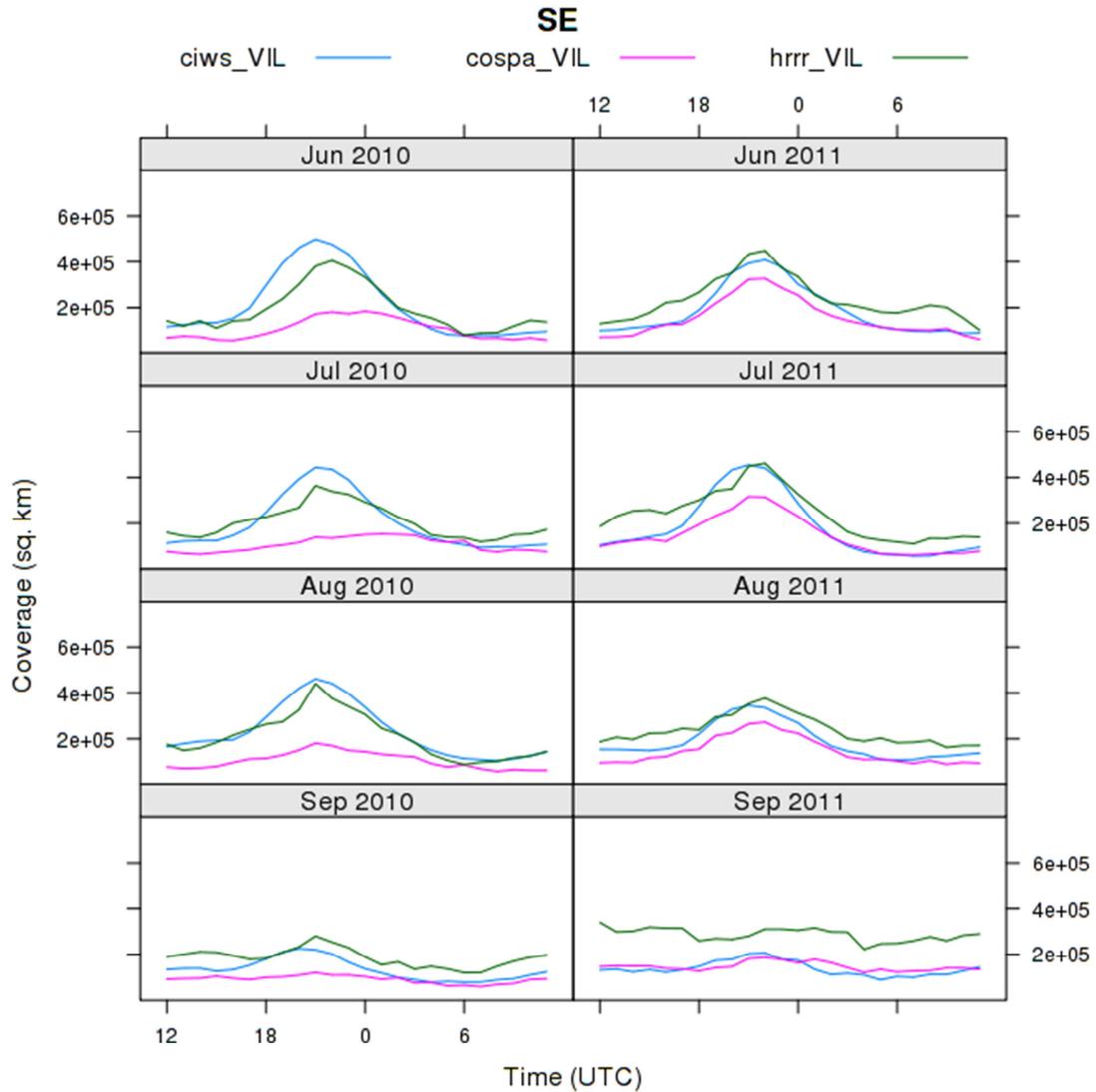


Figure 5.6. Same as in Figure 5.3, except for the SE Region.

The results for the Western U.S. are shown in Figure 5.7. Although convection over the Western domain has little impact on air traffic for the NAS, the results are worth mentioning here. The convective coverage produced by CoSPA improved from 2010 to 2011, and is nearly identical to that which was observed by CIWS. In addition, the lag in convective onset noted in 2010 in CoSPA was reduced in the West in 2011. Overforecasting continued to occur in the HRRR from 2010 to 2011.

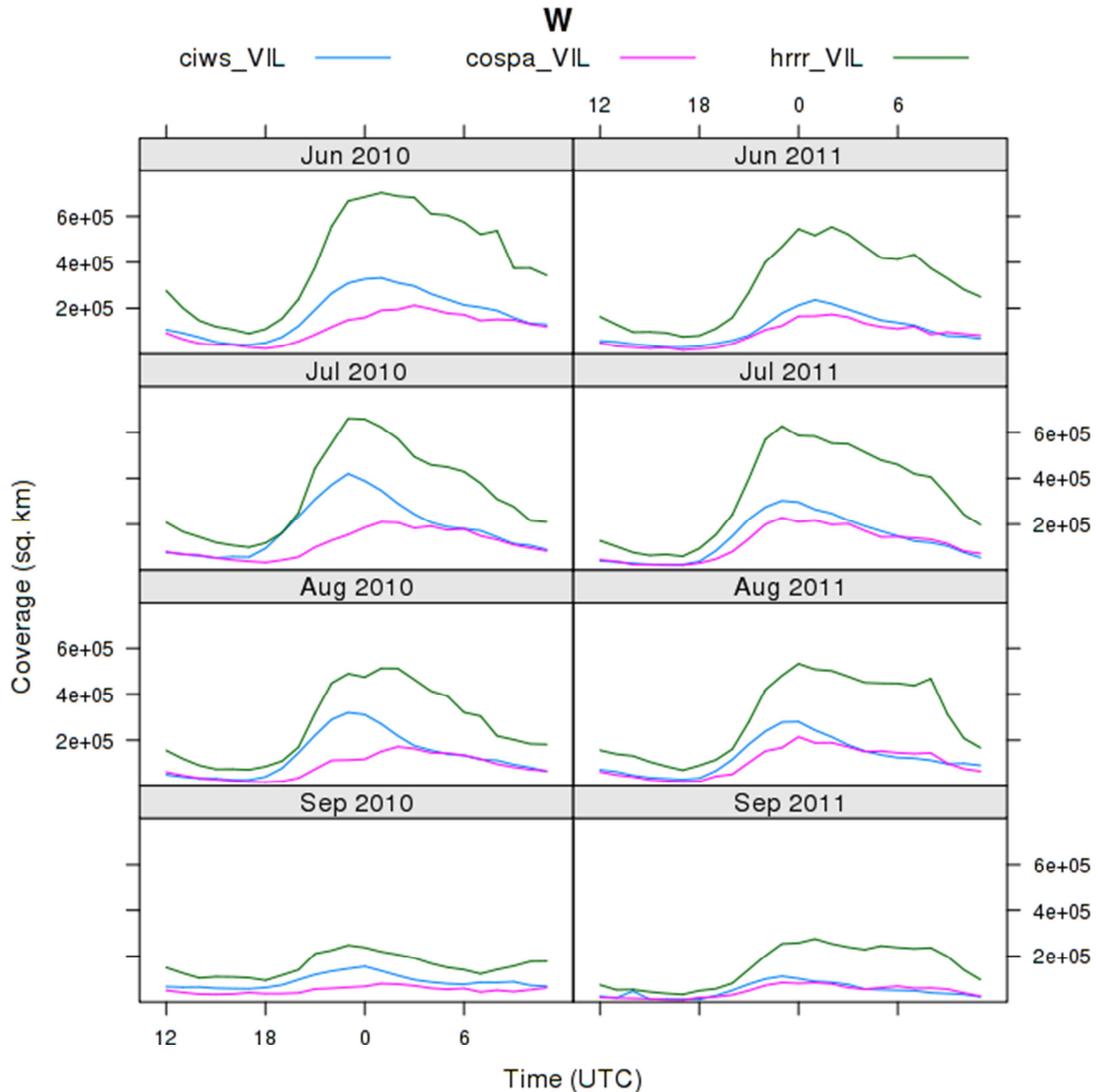


Figure 5.7. Same as in Figure 5.3, except for the Western Region.

5.2 Performance Analysis

5.2.1 Upscaling CoSPA VIL with Lead Time

Two specific aspects of CoSPA are investigated in this section: 1) the quality of CoSPA by lead time with a focus on the 2- to 5-h time periods when modifications to the blending scheme are evident and; 2) the resolution of information provided by CoSPA for supporting operational decisions. An examination of both VIL and echo top fields will be presented. The CSI results are presented in the form of boxplots; see Section 4.3 for an explanation of the boxplot.

The quality of CoSPA for 1- to 8-h lead times issued at 1500 UTC for three resolutions (native, 20-km, and 60-km) for June, July, August, and September 2011 at VIP-level 3, or hazardous convection, is shown in Figure 5.8. The CSI results

indicate a relative decrease in performance at all resolutions for all summer months in 2011 at the 3-h lead time. This decrease is most notable at 60-km resolution (green). This decrease in performance at the 3-h lead time, which is an important strategic period, represents a pattern that is indicative of less-than-optimal blending for the extrapolation forecast and for the model forecast at VIP-level 3. It is interesting to note that although the relative decrease in performance is evident at VIP-level 3, this is not the case at VIP-level 2 (Figure 5.9), which indicates the blending may have been optimized at this lower threshold.

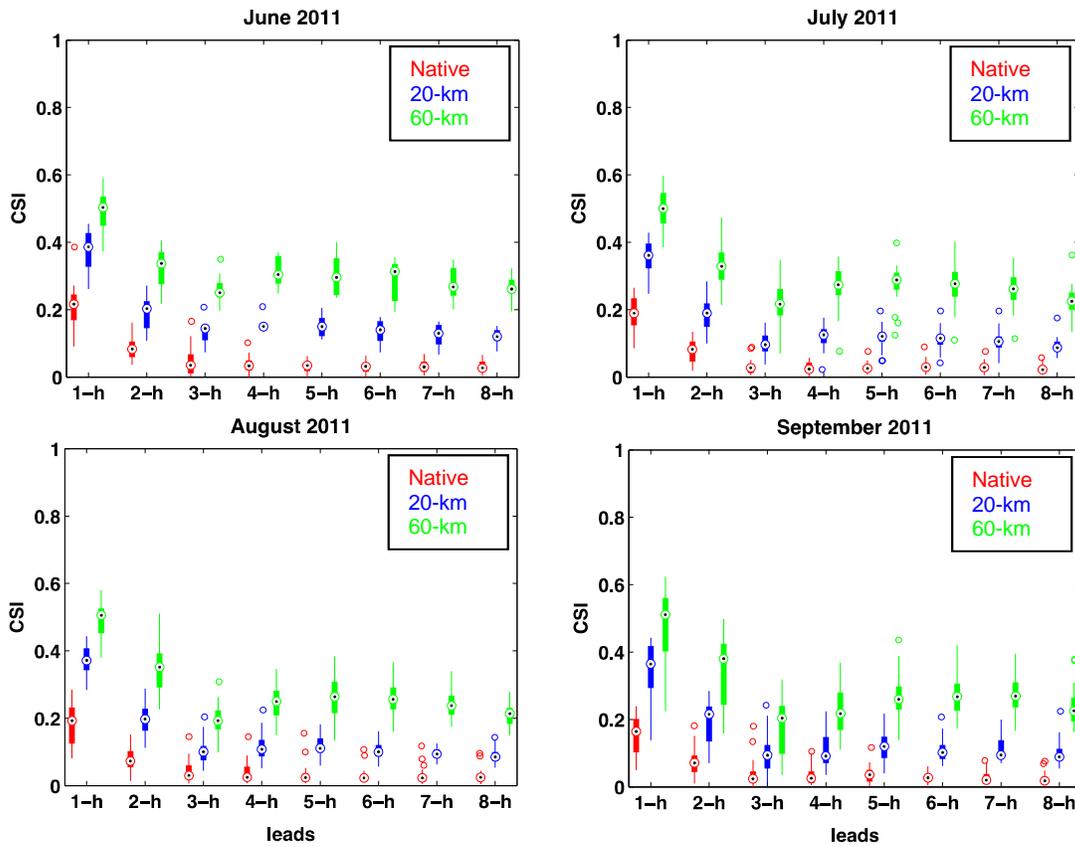


Figure 5.8. CSI as a function of lead time and resolution during 2011 for June (top left), July (top right), August (bottom left) and September (bottom right) for CoSPA at VIP-level 3 issued at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.

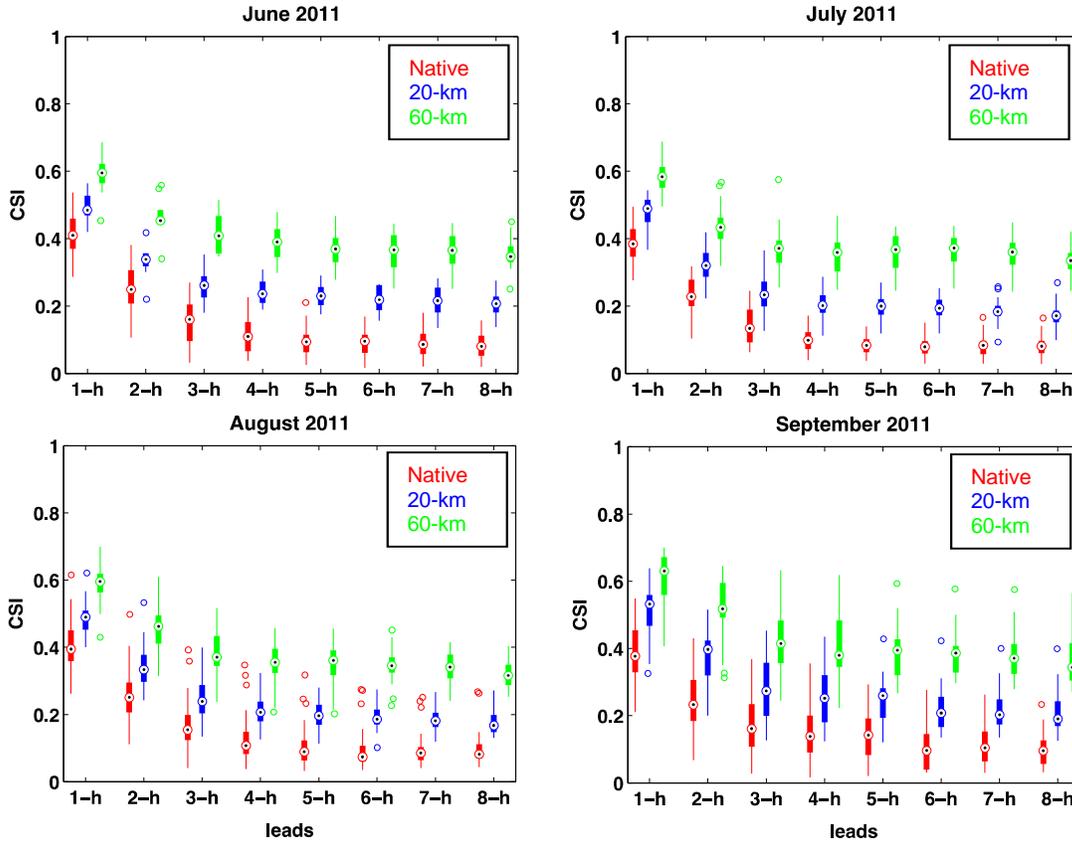


Figure 5.9. As in Figure 5.8, but for VIP-level 2.

Due to the apparent reduction in skill at VIP-level 3, it is necessary to examine the bias behavior as a function of lead time for the 2011 summer period. Figure 5.10 presents an aggregate of all months at the 1500 UTC issuance time for both CSI and bias. There is noticeable underforecasting at the 3-h lead time at VIP-level 3 corresponding to the reduction in skill. Figure 5.11 presents a similar result, but for VIP-level 2. No reduction is present at this threshold and there is actually a slight overforecasting signal in the 3- to 5-h lead-time frame.

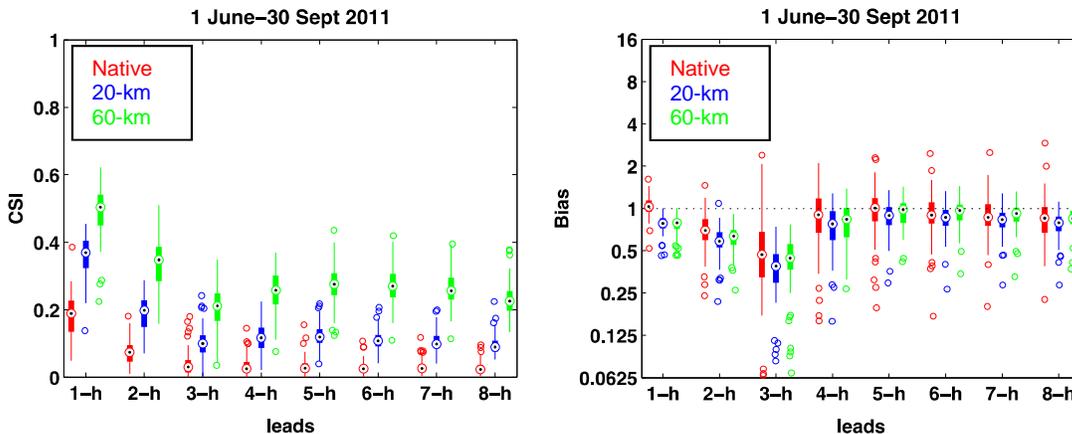


Figure 5.10. CSI (left) and bias (right) as a function of lead time for CoSPA with a 1500 UTC issuance at VIP-level 3. Resolutions are similar to that of Figure 5.8.

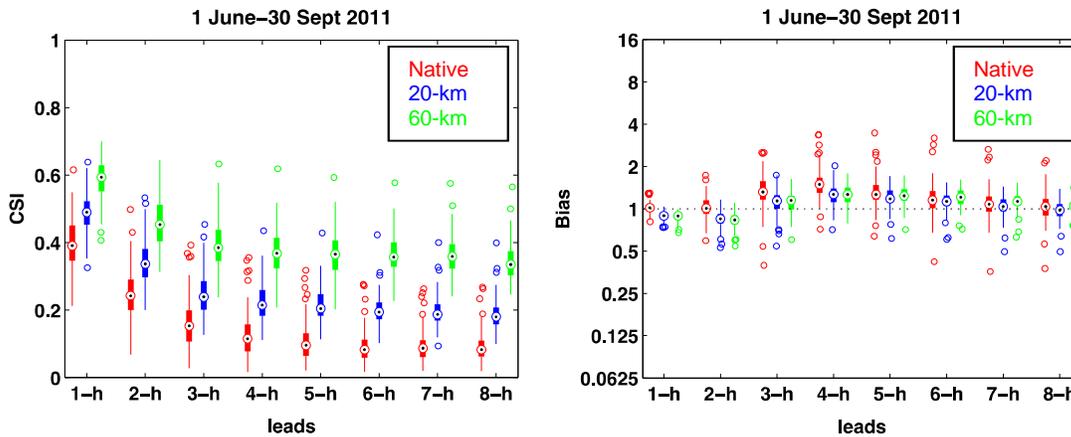


Figure 5.11. CSI (left) and bias (right) as a function of lead time for CoSPA with a 1500 UTC issuance at VIP-level 2. Resolutions are similar to that of Figure 5.8.

5.2.2 Examination of CoSPA Echo Tops

Cloud echo top information is critical to flight planning and is used to determine if planes can fly over a thunderstorm or whether re-routing around the storm is needed. An initial subjective assessment of CoSPA echo tops indicated a performance anomaly where forecast echo tops appeared to be significantly low across the CONUS. In Figure 5.12 this problem is illustrated for the 1300 UTC issuance on 23 June 2011, where it is seen that the CoSPA 4-h lead-time echo tops lack dimensionality and were forecast across the entire domain to be below 30 kft, while CCFP echo tops were forecast to exceed 39 kft over the SE U.S. and 34 kft in the NE. Recall from Figure 3.1 that the CoSPA VIL forecast for this same time period had embedded echoes in the SE that exceed VIP-level 3, which should also exceed 30 kft in the echo top forecast for that time of year.

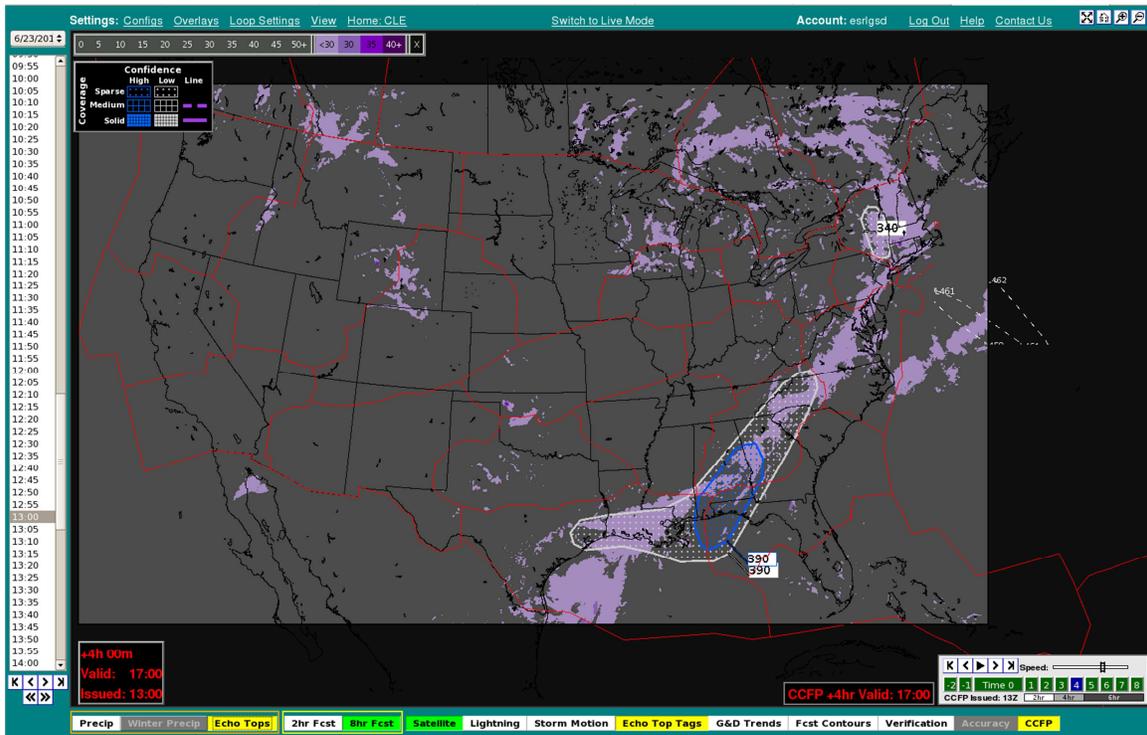


Figure 5.12. CoSPA echo top forecast from 23 June 2011 issued at 1300 UTC, valid at 1700 UTC. CCFP polygons are overlaid. Notice most of the field is forecasting less than 30-kft echo tops.

An aggregate plot of the bias of echo tops at the 30-kft threshold by lead time for 1500 UTC strategic issuances during 2011 is presented in Figure 5.13. The CoSPA echo top forecasts exhibited low echo top bias during most months at the 3- to 5-h lead time; however, June and August are the most extreme examples of the underforecasting of echo tops. This behavior was not evident in 2010. Biases for the 2010 CoSPA echo tops were relatively well-behaved at the 1500 UTC issuance (Figure 5.14). When aggregating additional issuance times for bias measurements by lead time for 2011, the underforecasting signal is still evident; however, it appears that the bias at 1500 UTC is where the signal is most prominent (Figure 5.15). The skill of the 30-kft “echo top and above” CoSPA forecast is low compared to skill at the VIP-level 3 and above CoSPA forecast.

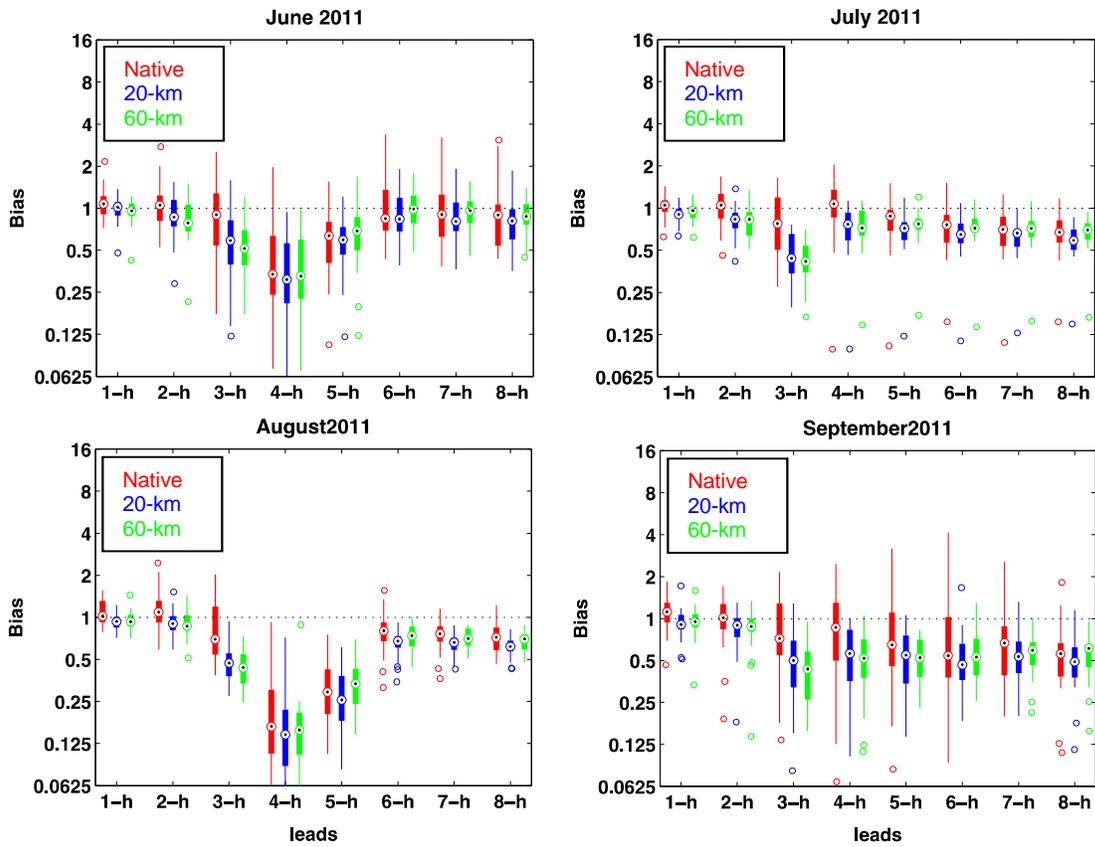


Figure 5.13. Bias as a function of lead time and resolution during 2011 for June (top left), July (top right), August (bottom left) and September (bottom right) for CoSPA at the 30-kft echo top threshold issued at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.

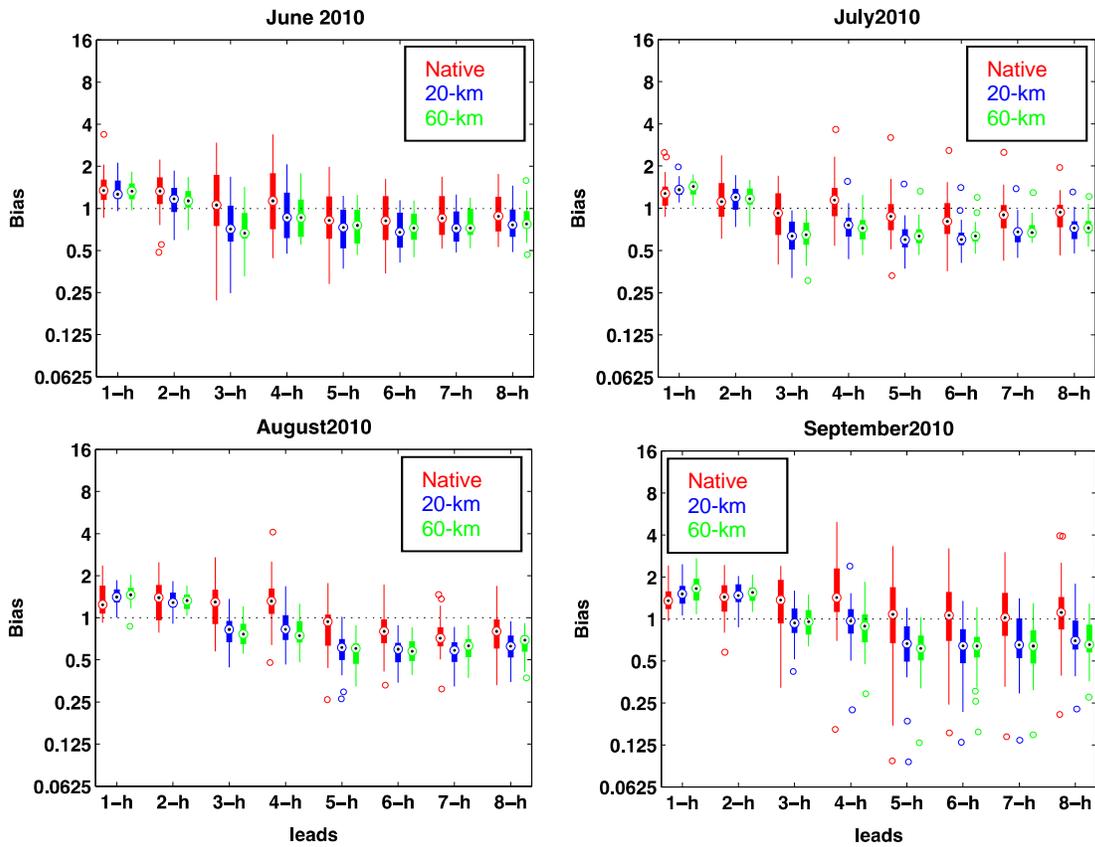


Figure 5.14. As in Figure 5.13, but for the 2010 Season.

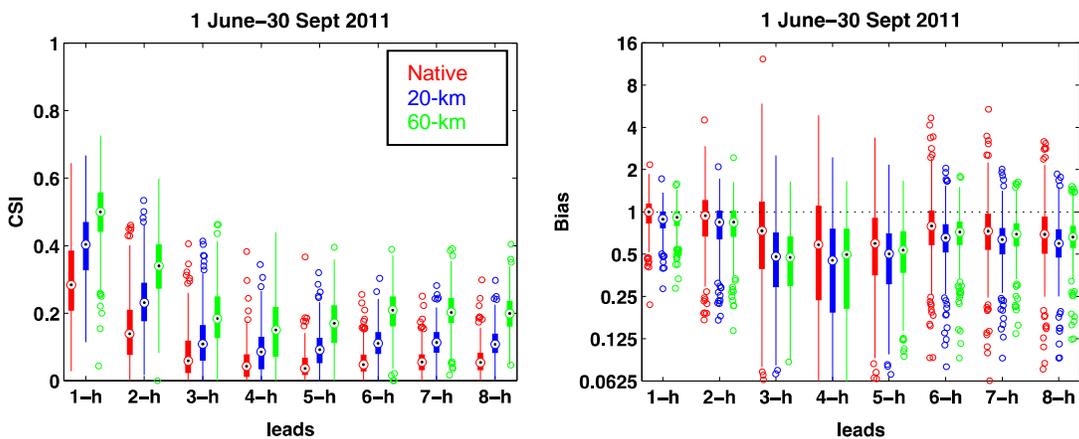


Figure 5.15. CSI (left) and bias (right) as a function of lead time for CoSPA at 1300, 1500, and 1700 UTC issuances at the 30-kft echo top threshold. Resolutions are the same as in Figure 5.13.

When examining the underlying HRRR, the signal in bias is nearly the same for all months in 2010 and 2011, indicating that the blending algorithm is the likely cause of this systematic behavior (Figure 5.16). The HRRR's 3-h latency is accounted for in the plots to match the input to the CoSPA 1300, 1500, and 1700 UTC issuances and to the 1- to 8-h forecast lead times.

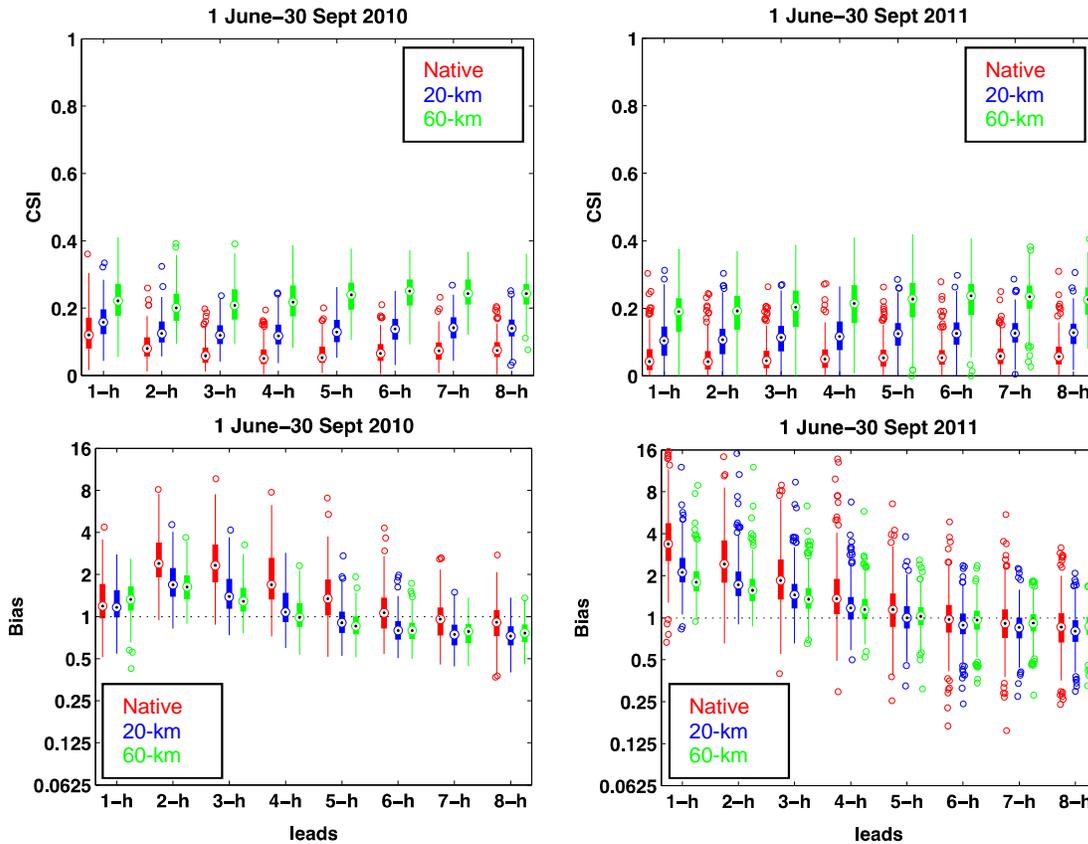


Figure 5.16. CSI and bias as a function of lead time for the HRRR at 1300, 1500, and 1700 UTC issuances after 3-h latency is applied at the 30-kft echo top threshold for 2010 (left) and for 2011 (right). Resolutions are the same as in Figure 5.13.

5.2.3 Forecast Resolution Analysis

Use of the Fractions Skill Score (FSS) allows for a meteorological comparison of skill at multiple spatial resolutions for forecasts of different types, including deterministic, probabilistic, and categorical forecasts. The FSS also provides a common approach for directly comparing the quality of the forecasts. Results in this section will be stratified by region and by high impact days. High impact days were identified by: 1) total delays in minutes due to weather across the NAS, and 2) days that included an AFP. The 6-h lead is of particular interest to traffic flow planning, so results for that time period are highlighted. Results for other leads are available upon request.

Regional Analysis

Figure 5.17 shows the FSS results for 1300 UTC issuance 6-h lead for CoSPA, HRRR, CCFP, CCFP-re-categorized, LAMP, Climatology, and Uniform for the NE domain. Improvement in CoSPA performance was noticeable in 2011 in the NE where CoSPA outperformed all forecasts at resolutions greater than 45 km. This improvement was significant when compared to results from 2010, where a coarser CoSPA resolution (greater than 100 km) was needed to outperform the other forecasts at the 1300 UTC issuance time (Figure 5.18). These results indicate for this strategic issuance time, that forecast information was available in 2011 from CoSPA at higher resolutions for the NE than was available in 2010.

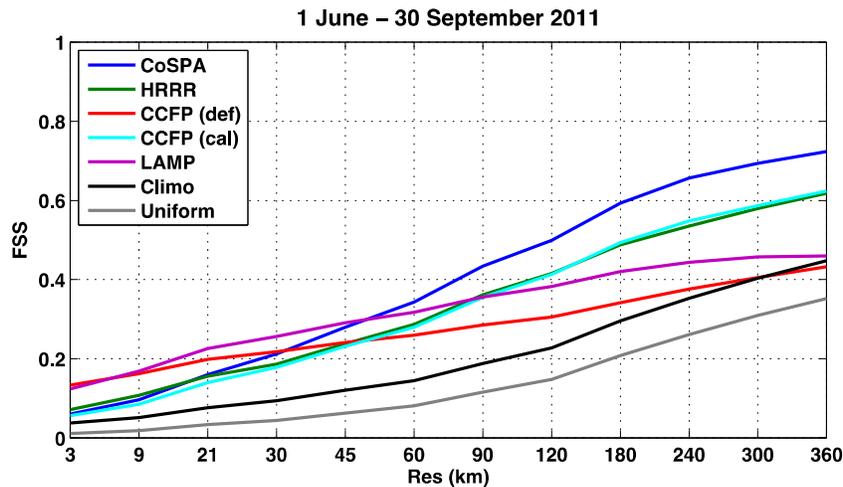


Figure 5.17. Mean FSS for the NE as a function of resolution for CoSPA (blue), HRRR (green), CCFP (red), re-categorized CCFP (cyan), LAMP (magenta), climatology (black), and uniform (gray) for the 1300 UTC issuance valid at 1900 UTC 2011. Results account for the HRRR latency.

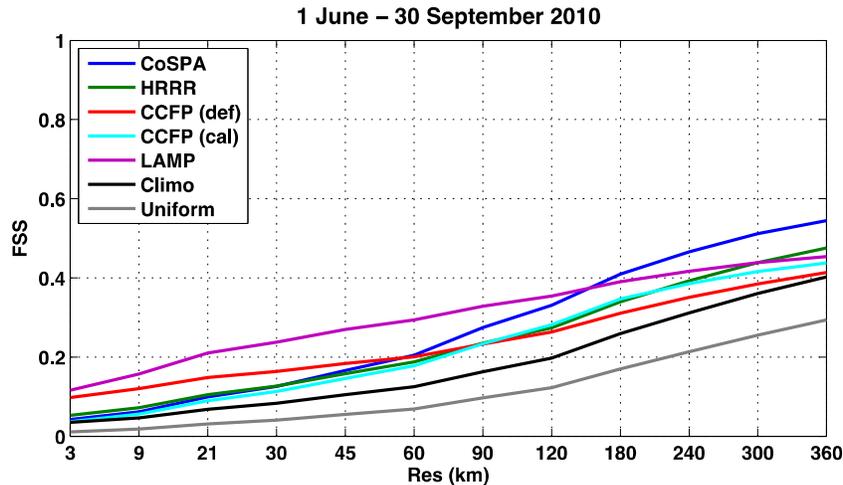


Figure 5.18. As in Figure 5.17, but for 2010 in the NE.

In the SE domain (Figure 5.19), the quality of CoSPA improved significantly over the quality measured during the 2010 evaluation (Lack et al. 2011), and is as accurate as the forecasts in the NE domain for 2011 (Figure 5.17). In the SE, it is notable that the re-categorized CCFP is performing similarly to CoSPA for the 1300 UTC issuance, 6-h lead time (Figure 5.19). Note that the performance of the HRRR in 2010 (Figure 5.20) outperformed CoSPA, indicating the blending algorithm performed less than optimally. The increase in CoSPA's skill in the SE over its parent model from 2010 to 2011 indicates that the addition of the dynamic blending algorithm in 2011 resulted in more effective use of the HRRR as a component of the CoSPA product.

The skill of CoSPA in terms of FSS at the 1500 UTC issuance for 6-h lead (not shown) is similar to the 1300 UTC issuance with CoSPA performing slightly better than the other products in both the NE and SE. In addition, the 2-h lead-time CoSPA product significantly outperformed all other products, similar to the 2010 results. A discontinuity still exists between the 2:00 lead time and the 2:15 lead time in 2011, since the 0 to 2-h lead-time forecast is simply the CIWS extrapolation product.

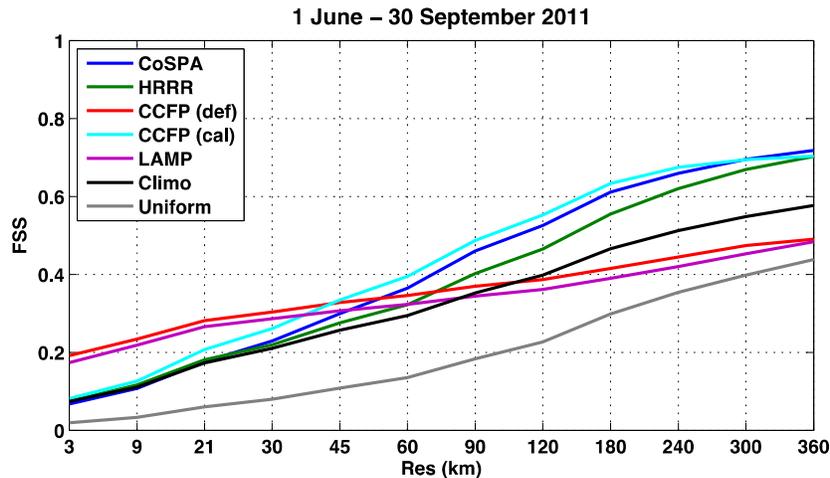


Figure 5.19. As in Figure 5.17, but for 2011 in the SE.

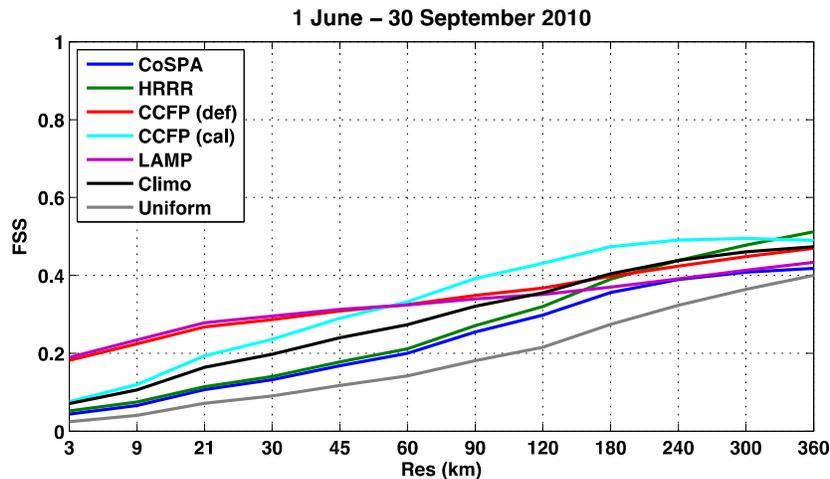


Figure 5.20. As in Figure 5.17, but for 2010 in the SE.

High Impact Days

Figure 5.21 presents results from 1300 UTC 6-h lead time for CoSPA, HRRR, CCFP, CCFP-re-categorized (calibrated), and LAMP from 1 June – 30 September 2011 for the top 15 delay days based on ground delay in total minutes. During this period, CoSPA continued to outperform all other forecasts from a resolution of 70 km and greater. CCFP re-categorized retains high skill on these days as these situations were most likely strongly forced (frontal) events.

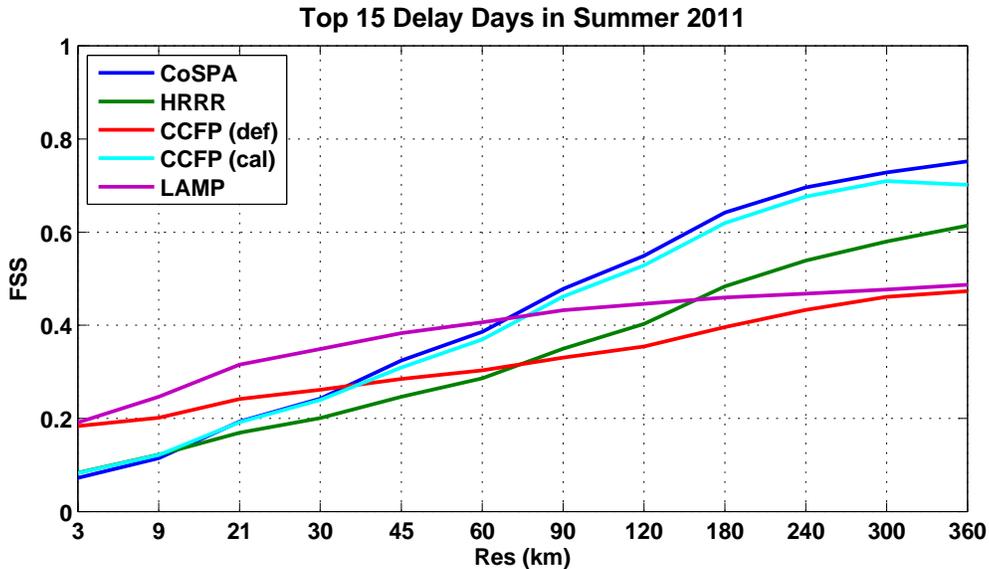


Figure 5.21. As in Figure 5.17, but for the top 15 delay days in total minutes from 1 June 2011 to 30 September 2011 in the NE domain for the 1500 UTC issuance and 6-h lead time.

The type of Traffic Management Initiative enacted during an impactful weather event (e.g., an AFP or GDP) depends on the type of convective weather that is present. For instance, most AFPs are associated with strongly forced cold fronts and weather events that are more easily forecast with respect to their location, orientation, and strength. In contrast, GDPs are often associated with isolated air mass thunderstorms where their location, movement, and intensity are more difficult to identify. Therefore, it was necessary to investigate the quality of CoSPA relative to both AFP and GDP impact days.

Figure 5.22 shows a comparison between the forecasts on days where AFPs were issued (top) and days where GDPs were issued that affected NE terminals (bottom) for the period 1 June – 30 September 2011. The results here show that CoSPA performed equally well on AFP and GDP days, and better than all other forecasts at a resolution of greater than 75 km on AFP days and 45 km on GDP days. These results suggest that when weather features are in the form of isolated convection and are more difficult to forecast, CoSPA provides an added advantage to air traffic flow planners over coarser products that rely on convective parameterization. At

resolutions less than 45 km, LAMP and CCFP provide better overall performance, but frequently lack convective structure that is often found in the high resolution CoSPA forecasts. Interestingly, the HRRR performance varies by nearly 10% between AFP days and GDP days. The consistent, high performance of CoSPA for both types of events suggests that the blending algorithm is appropriately accounting for the HRRR variation.

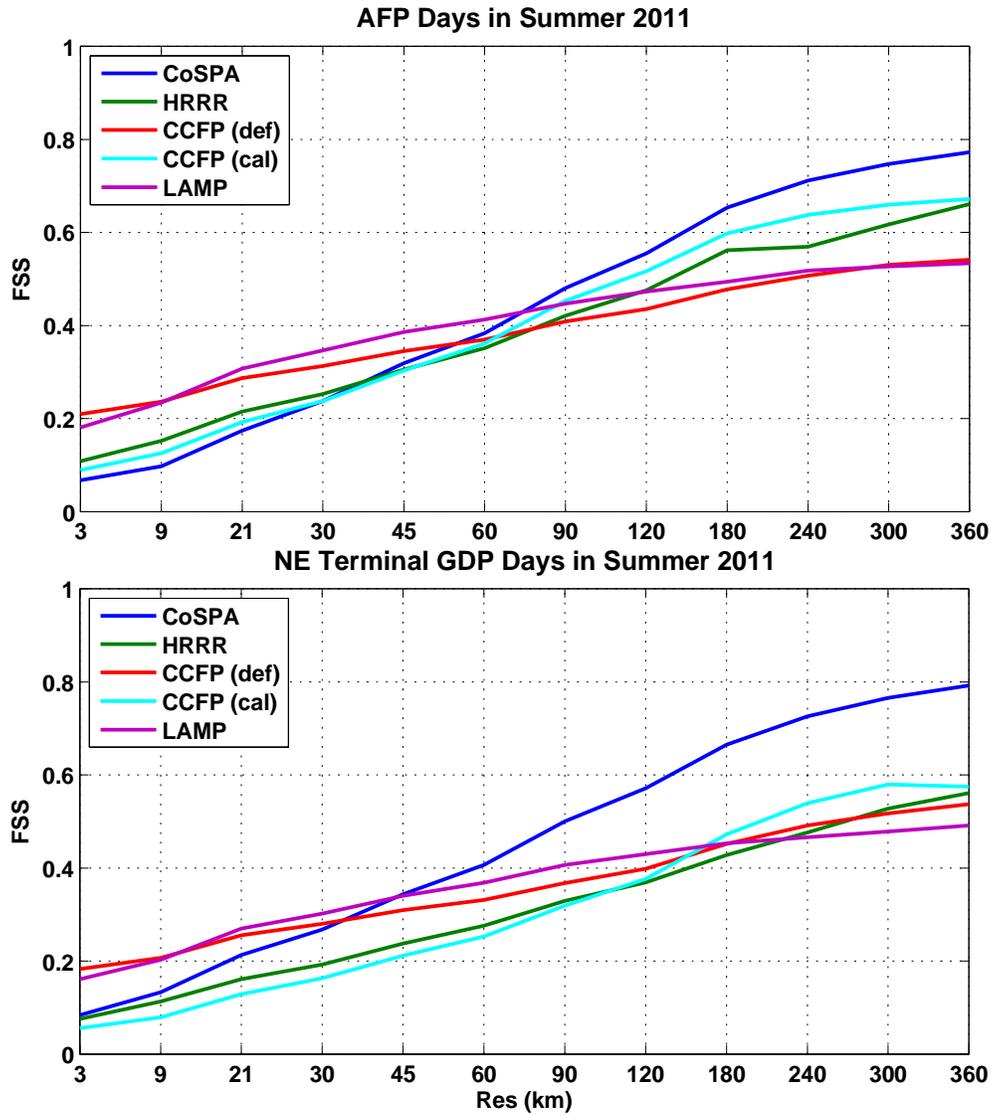


Figure 5.22. As in Figure 5.17, but for AFP days (top) and NE terminal GDP days (bottom) from 1 June 2011 to 30 September 2011 in the NE domain for the 1500 UTC issuance and 6-h lead time.

5.2.4 Quality Relative to Airspace Flow Constraints

An investigation of the performance of CoSPA as a predictor of airspace constraint is presented in this section. The Flow Constraint Index (FCI) is the measure used to quantify the performance of CoSPA in this regard.

The CSI as a function of FCI threshold for the period 1 June – 30 September 2011 is presented in Figure 5.23 at the ARTCC scale, and in Figure 5.24 at the sector scale, for forecasts issued at 1500 UTC with a 6-h lead time. CSI values to the left of the vertical dotted yellow line (0.1) coincide with little to no convective-related constraint throughout the NAS. CSI values between the dotted yellow and the dotted red line (0.35) coincide with moderate constraint throughout the NAS, and CSI values to the right of the vertical dotted red line coincide with significant constraint.

The results at ARTCC scale indicate that in the NE, the performance of CoSPA is nearly identical to the performance of the HRRR and the re-categorized CCFP. The performance of CoSPA and the HRRR drops slightly for significant constraints, as indicated by the lower CSI values for FCI thresholds greater than 0.35. In the SE for moderate or greater constraints, the performance of CoSPA is equivalent to or exceeds all other forecasts. The quality of CoSPA in the SE was nearly identical to the quality measured in 2010 (Lack et al. 2011). However, with the change in the CoSPA blending scheme introduced in 2011, CoSPA is now able to outperform its parent model (HRRR) in the SE for events that impose a significant constraint on the NAS.

When considering results at the higher-resolution (sector size; Figure 5.24) the performance of CoSPA and all other models is reduced, as reflected by the lower CSI values. However, the performance of CoSPA, HRRR, and re-categorized CCFP are nearly identical for moderate and greater constraints and are better than both the LAMP and standard CCFP, as these forecasts have limited sharpness.

Other strategic issuance times (not shown) for the 6-h lead time perform similarly at each spatial scale. The SE seems to perform better at both resolutions due to more frequent convection in the SE than in the NE. In other words, significant convection in the NE is much less likely than it is in the SE.

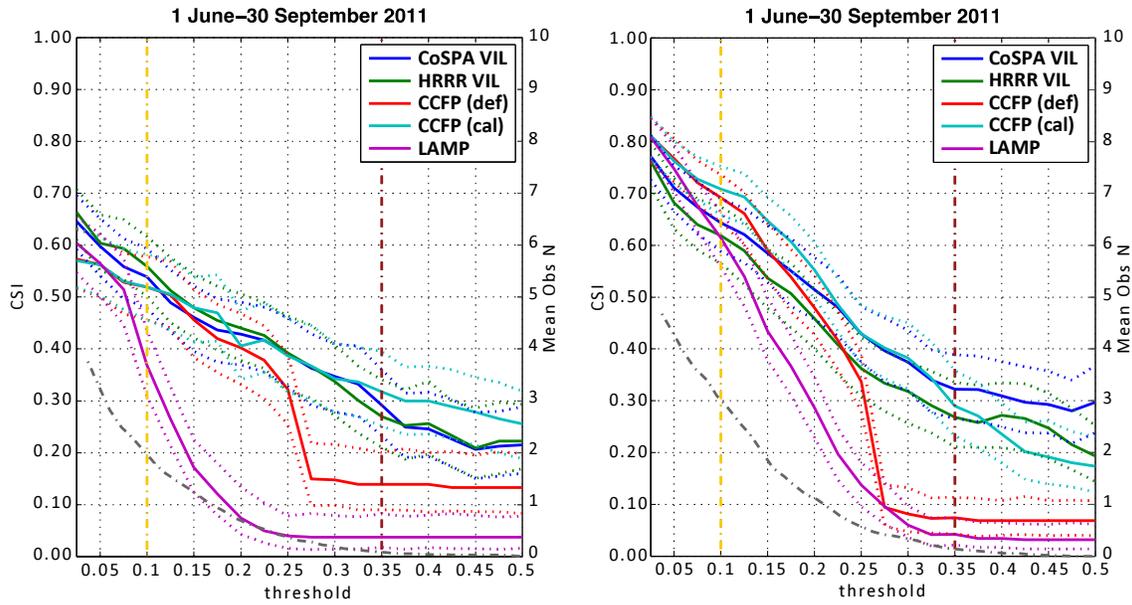


Figure 5.23. CSI as a function of FCI (constraint) threshold for 1 June – 30 September 2011 for the NE domain (left) and SE domain (right) at ARTCC scale. CoSPA in blue; HRRR in green; CCFP standard in red; CCFP re-categorized in cyan; LAMP in magenta. The gray-dashed line is the average number of ARTCC hexagons constrained by convection for the given threshold. Right of the yellow-dashed vertical line represents medium constraint; right of the maroon-dashed line represents high constraint. Dotted green, blue, red and cyan lines are confidence intervals.

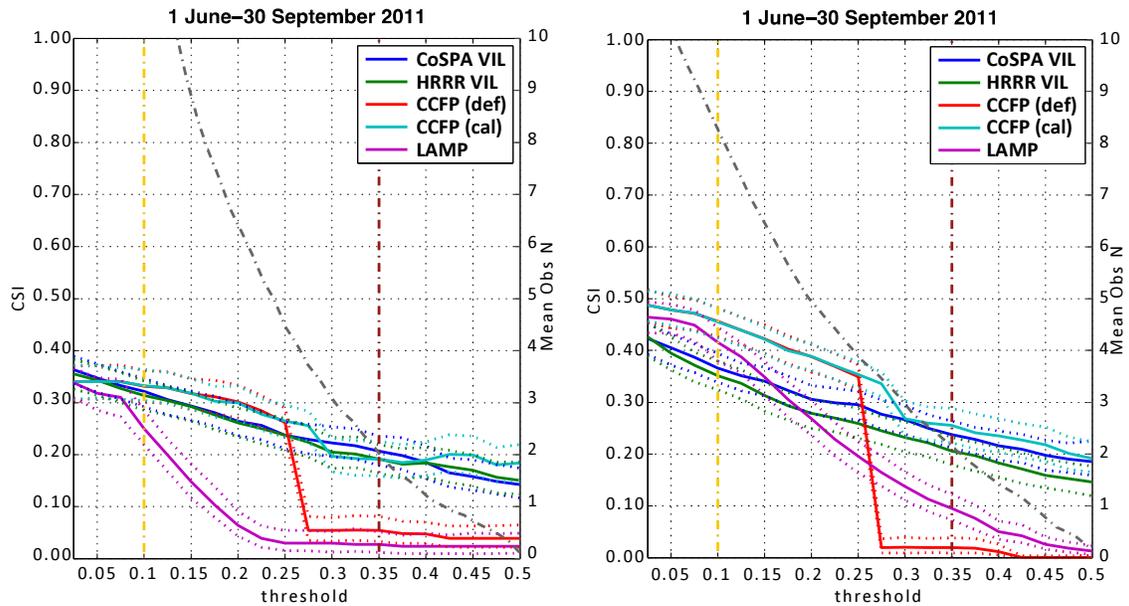


Figure 5.24. As in Figure 5.23, but for sector scales.

5.2.5 Forecast Consistency

A consistent forecast message is critical for effective ATM strategic planning. To measure forecast consistency between the issue/leads of a forecast suite, the correspondence ratio (CR) applied to FCI-translated forecast products is used. CR values of 1.0 indicate perfect consistency across the valid times, while 0.0 indicates no consistency. The results for the NE region for ARTCC-sized hexagons, presented in Table 5.1, indicate that CCFP is slightly more consistent than CoSPA for all levels of impact and for the times of day when constraints are more frequent. However, CoSPA does maintain relatively high consistency at the severe constraint threshold of 0.35 across most valid times. It is interesting to note that the consistency values for each of the thresholds are similar across the four valid times chosen.

Table 5.1. Consistency for the 1700, 1900, 2100, and 2300 UTC valid times in the NE region when considering the 2-, 4-, and 6-h leads for CCFP re-categorized and CoSPA for thresholds of 0.01 (low constraint and above), 0.1 (moderate constraint and above), and 0.35 (severe constraint and above).

NE Region	CCFP			CoSPA		
	<i>consistency by level of constraint</i>			<i>consistency by level of constraint</i>		
Valid Time	Low	Moderate	Severe	Low	Moderate	Severe
17 UTC	0.70	0.68	0.89	0.58	0.60	0.81
19 UTC	0.75	0.71	0.88	0.65	0.58	0.78
21 UTC	0.76	0.74	0.85	0.67	0.56	0.77
23 UTC	0.70	0.69	0.86	0.62	0.57	0.78

Table 5.2 shows consistency results for the SE for both forecasts. Unlike the NE region, the skill varies quite significantly across valid time with a considerable drop in consistency at early pre-initiation times (1700 UTC) and post-initiation times (2300 UTC).

Table 5.2. As in Table 5.1, but for the SE region.

SE Region	CCFP			CoSPA		
	<i>consistency by level of constraint</i>			<i>consistency by level of constraint</i>		
Valid Time	Low	Moderate	Severe	Low	Moderate	Severe
17 UTC	0.65	0.51	0.74	0.65	0.51	0.71
19 UTC	0.80	0.61	0.75	0.73	0.50	0.69
21 UTC	0.68	0.50	0.72	0.71	0.48	0.70
23 UTC	0.39	0.43	0.81	0.59	0.51	0.72

5.2.6 Performance at CCFP Scales

The CoSPA User's Guide (FAA 2011) states that "CoSPA is intended to be used in conjunction with CCFP, which remains the official product for TFM decision-making." It is therefore important to measure the quality of CoSPA as it relates to CCFP. Two types of assessments are presented in this section: 1) a direct comparison between CoSPA and CCFP at CCFP temporal and spatial scales, and 2) an assessment of CoSPA when used as a supplement to CCFP.

The clustering technique (described in Section 4.6) is used to coarsen CoSPA and LAMP to the spatial resolution of the CCFP, and the CSI statistic is computed at issuances and leads corresponding to those of the CCFP. The CSI results are presented in the form of boxplots, see Section 4.7 for an explanation of the boxplot.

Comparison of CoSPA and CCFP

Figure 5.25 illustrates the performance of CoSPA relative to CCFP and LAMP at CCFP temporal and spatial scales, and corresponding to the sparse coverage, low confidence and above criteria for CCFP areas. The results are broken down by 2-, 4-, 6-, and 8-h lead times. Because strategic planning is critical 8 hours prior to the onset of convection, the quality of the forecasts at this lead time is of importance. Therefore, the CCFP 6-h forecast was persisted and used for comparison in this analysis. Results are presented for strategic, pre-convective initiation time periods by combining the 1100, 1300, and 1500 UTC issuance times, and for post-convective initiation by combining the 1700, 1900, and 2100 UTC issuance times.

The results in Figure 5.25 indicate that for strategic time periods, CoSPA outperforms LAMP and CCFP at the 2-h lead time by a statistically significant margin. CoSPA performance at the 4-h lead time is still greater than that of the other products, but dips below that of CCFP at the 6- and 8-h lead times. It is important to note that CoSPA does not suffer a degradation in skill at the 4-h lead time, as was shown in the upscaling plots at VIP-level 3. This is due to the clustering technique, which uses information at both VIP-level 3 and VIP-level 2. When examining longer lead times, it is apparent that the CCFP outperforms CoSPA and LAMP for the strategic time period. It is also worth noting that a persistent CCFP 6-h forecast provides considerable skill at 8 h.

When comparing the pre- and post- convective initiation results (Figure 5.25 (top) and (bottom)), performance for all forecasts is nearly the same with only a slight reduction in quality for the post-initiation period at longer lead times. This indicates a potential decrease in accuracy due to the cessation of convective activity.

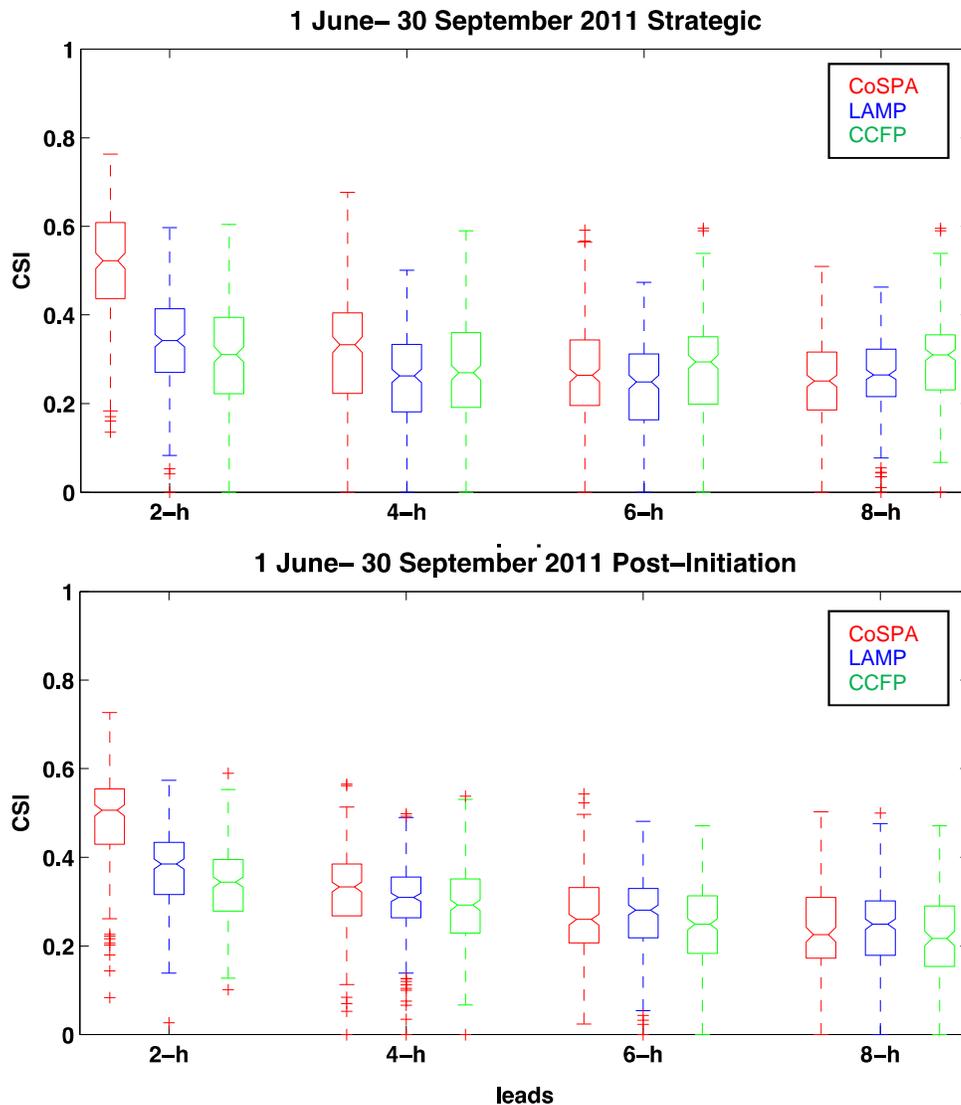


Figure 5.25. CSI as a function of lead time for strategic telecon, pre-convective initiation hours (top 1100, 1300, 1500 UTC issue times) and for post-convective initiation hours (bottom; 1700, 1900, 2100 UTC), following the CCFP criteria of sparse coverage/low confidence and above. CoSPA is in red, LAMP in blue, and CCFP in green.

The performance of CoSPA, CCFP, and LAMP for the sparse coverage, high confidence and above criteria is shown in Figure 5.26. Comparisons between Figure 5.25 and Figure 5.26 indicate a decrease in performance at the sparse coverage, high confidence and above criteria for all forecasts. However, the performance of CoSPA is higher than CCFP for all time periods, and is higher than LAMP at the 2- and 4-h lead times. Similar results are measured for thresholds at medium coverage CCFP areas, where CoSPA retains some skill, especially at the 2-h lead time as compared to the other forecasts assessed.

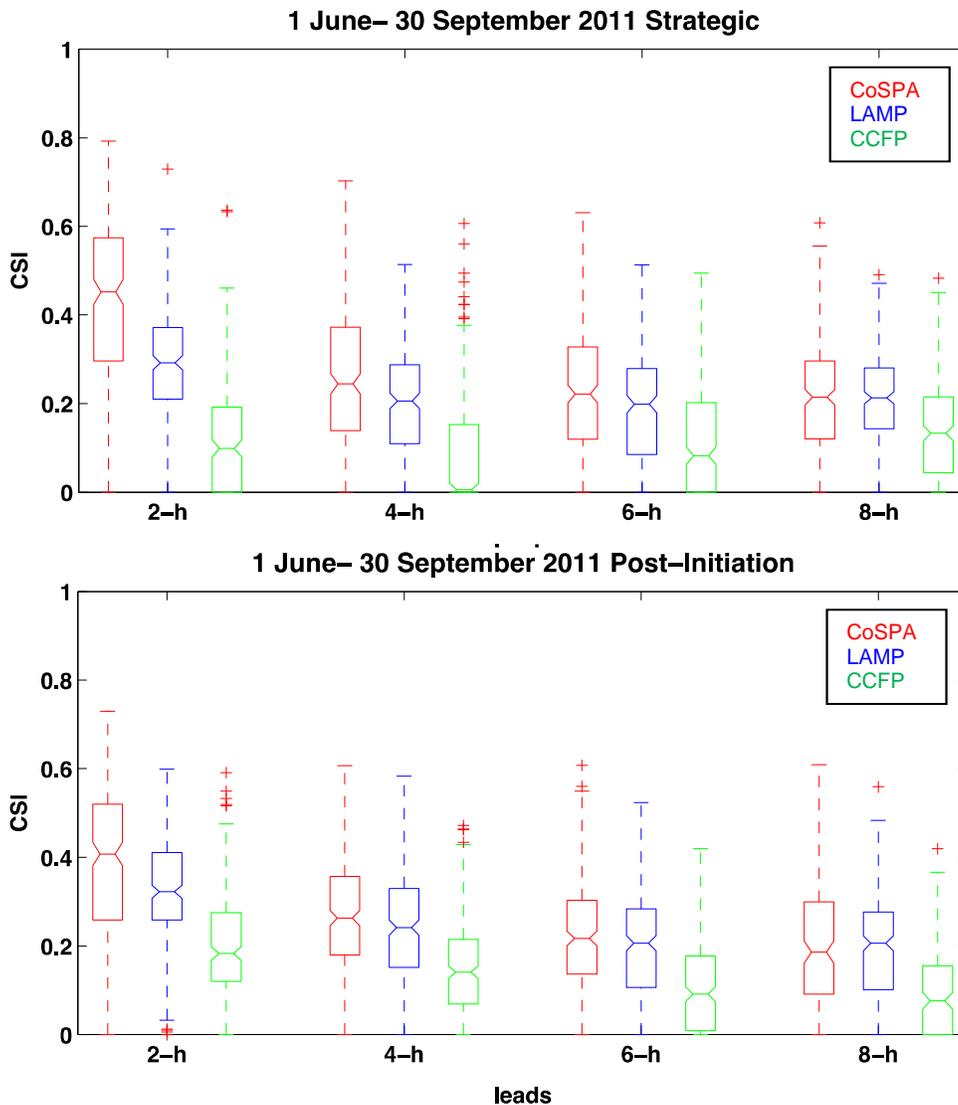


Figure 5.26. As in Figure 5.25, but for CCFP criteria of sparse coverage, high confidence and above.

CoSPA as a Supplement to CCFP

The goal of this section is to determine if CoSPA can be used as a supplement to the operational CCFP to provide additional, beneficial information beyond that which is available from the CCFP alone. Historically, the dominant combination of coverage and confidence attributes for CCFP is sparse coverage/low confidence. In any given year, between 60 and 70 percent of all areas included in CCFP forecasts are of this type. Traffic flow managers often dismiss these areas when managing the NAS. They issue TMIs only when high confidence areas are present. While many sparse coverage/low confidence areas are linked with low-impact weather events, not all of these areas should be discounted. Because of the frequent issuance and under-utilization of CCFP sparse coverage/low confidence areas, a separate analysis was

carried out to determine if there is a supplemental relationship between CCFP and CoSPA for these particular forecasts for improving their use for traffic flow planning.

The clustering technique introduced in Section 4.2.5 is utilized for this analysis, and is applied to 6-h forecasts at the strategic issuance times of 1100, 1300, and 1500 UTC. Regions of significant coverage for both CoSPA forecasts and CIWS observations are derived within each sparse coverage/low confidence area. CoSPA can be viewed as a valuable supplement to CCFP sparse coverage/low confidence polygons by increasing the situational awareness of the potential hazards in these often-dismissed polygons. Such a benefit is realized when at least one area of medium or greater (yellow regions in Figure 4.7) coverage from CoSPA is found within a CCFP area that also contains one or more CIWS observations of medium or greater coverage. Likewise, identifying CCFP areas devoid of such dense CoSPA or CIWS observations can increase confidence that the area can reasonably be ignored for strategic TMI issuances.

Table 5.3 provides counts of the occurrence of (1) CCFP sparse coverage/low confidence polygons; (2) medium coverage (dense) observations, and; (3) forecast objects, followed by the summary skill statistics. It is important to note that the frequency of dense CIWS observations within CCFP sparse coverage/low confidence regions rose from 15% in 2010 to 36% in 2011. In other words, in 2011, CCFP sparse coverage/low confidence areas were more than twice as likely to contain convection that could disrupt air traffic than was evident in 2010, even though the frequency of issuance of these areas was nearly the same for each year.

Despite changes in weather patterns and in subsequent skill of the CCFP sparse coverage/low confidence areas from 2010 to 2011, the skill of CoSPA to supplement these frequently-dismissed CCFP areas remains quite good. The high value of PODn indicates that a CoSPA forecast of sparse or no convection located within a CCFP sparse coverage/low confidence polygon can increase user confidence that this region is likely to result in minimal traffic disruption. The PODy value of 0.53 in conjunction with the low FAR of 0.36 (a significant decrease from the 2010 value of 0.62) indicates that when CoSPA forecasts a region of dense convection within a CCFP sparse coverage/low confidence polygon, confidence that this area will contain impactful convection increases and the area should be reconsidered in the planning process.

Table 5.3. Summary of the number of identified objects at the medium and above coverage threshold in the CIWS VIL analysis field and the CoSPA VIL forecast that are coincident in a CCFP sparse coverage, low confidence area for strategic issuance times at the 6-h lead time from 1 June-30 September.

	2011	2010
# of CCFP Sparse/Low	1170	1101
# of Dense Observations	416	165
# of Dense Forecasts	342	282
PODy	0.526	0.649
PODn	0.837	0.813
FAR	0.360	0.621

5.3 Performance of CoSPA During Winter

Although CoSPA is primarily a summer-time convective forecast, when it becomes operational to support FAA decisions, CoSPA is expected to run continuously, providing forecasts for all seasons. Thus, it is necessary to investigate the skill of CoSPA during the winter. CoSPA was evaluated primarily over the southern U.S. during the period 20 December 2010 to 28 February 2011. The climatology of CoSPA is examined to gain insight into the precipitation phase field, echo tops, and VIL thresholds relevant for winter precipitation. Following this climatological analysis, a skill assessment of CoSPA performance over the southern United States will be presented.

5.3.1 Winter Climatology

Precipitation Phase

Although it is not our goal to evaluate the quality of the precipitation phase field, it is used in this study to stratify the CoSPA VIL field into regions of convective and winter weather and for understanding the distribution of frozen phase precipitation within different CoSPA VIP severity thresholds. As an aside, a preliminary investigation of the METAR observations and the precipitation phase field indicated a significant level of consistency between the observations and the forecast.

Figure 5.27 shows the occurrence of warm, mixed, and cool phase precipitation corresponding to VIP-level 3 (left) and VIP-level 2 (right) for CIWS, CoSPA, and the HRRR. At the VIP-level 3 threshold, 95% of the CIWS observations are contained in the warm precipitation phase with little mixed precipitation. The CoSPA forecast indicates a similar pattern, but with the addition of a few mixed and frozen

precipitation events. The HRRR has a large bias, representing general overforecasting for all precipitation phases and events. Interestingly, CoSPA is able to calibrate this large HRRR bias as it is integrated into CoSPA. When the VIP-level 2 threshold is considered, more mixed and frozen events are introduced into the observations and into the forecast. Since our goal is to evaluate the ‘convective’ features of CoSPA during the winter months, all results presented in this section will be computed using the VIP-level 3 threshold.

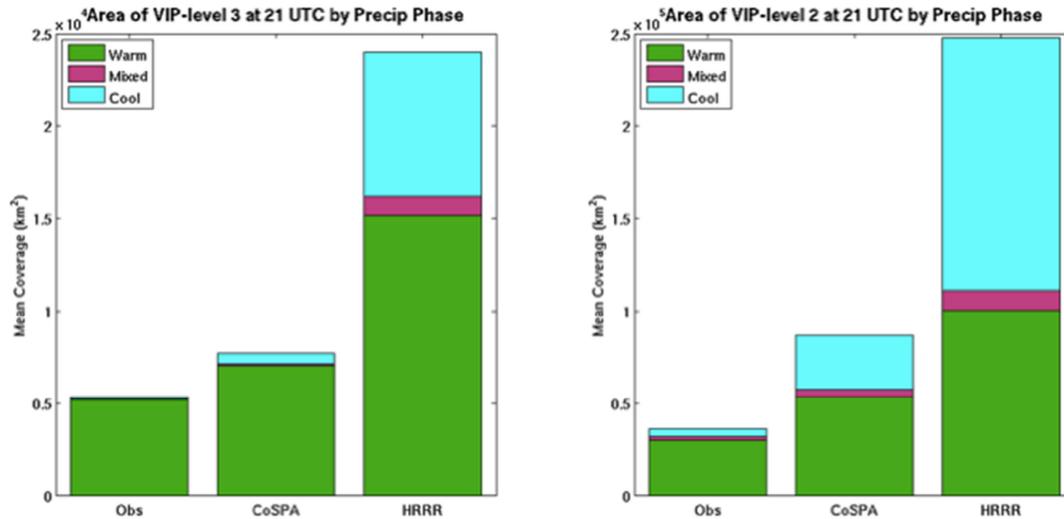


Figure 5.27. Distribution of the area of the phase of precipitation for CIWS analysis, CoSPA VIL forecasts, and HRRR VIL forecasts for VIP-level 3 (left) and VIP-level 2 (right). Green is warm phase, magenta is mixed phase and cyan is cool phase (frozen).

Figure 5.28 illustrates VIP-level 3 convection observed by CIWS (left) and the 6-h convective forecast from CoSPA (right), both for 2100 UTC, a time that is frequently impactful to air traffic. It is evident from this analysis that the most significant convection during winter months is located in the SE U.S. During winter, this convection is typically due to large-scale synoptic forcing, rather than the smaller-scale airmass thunderstorms that are typical in the summer. Specifically, intense convection was located over the panhandle of Florida, with areas of active convection identified along the coastal states. Similar areas of convection were forecast by CoSPA, but the amount of forecast convection exceeded that which was observed in Mississippi, Alabama, Kentucky, West Virginia, and the southern part of New England. Note that the intense area of weather present in CoSPA over southern New England is most likely due to frozen or mixed-phase precipitation that appears as intense convection on the CoSPA display.

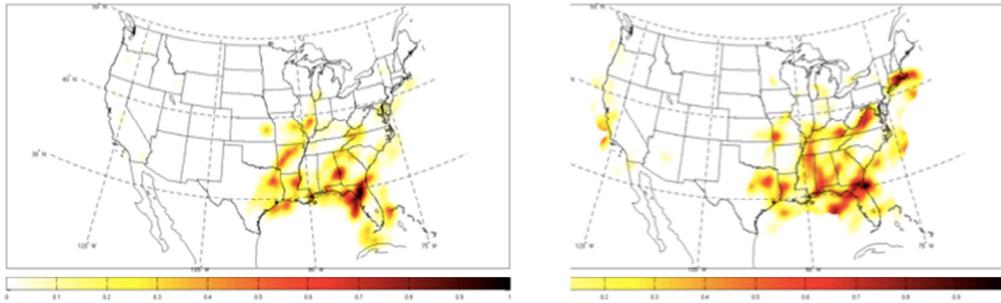


Figure 5.28. A comparison of the CIWS analysis with a VIP-level 3 threshold applied to the winter 2010-2011 study (left) valid at 2100 UTC, and the CoSPA forecast issued at 1500 UTC which is valid at 2011 UTC (right).

Echo Tops

Figure 5.29 shows the distribution of echo tops in 5-kft increments from 5 kft to 50 kft (x-axis) for CIWS observations, CoSPA, and HRRR. As expected, the characterization of the echo tops provided by the forecasts is quite different during the winter period as compared to the summer, due to differences in the tropopause height. Observed echo top heights are typically lower than 25 kft in the winter, while during the summer, echo tops often exceed 30 kft, impacting en route, commercial air travel. Although CoSPA and HRRR exhibit distributions that are similar to the observations, both nearly double the area of echo top coverage in the winter.

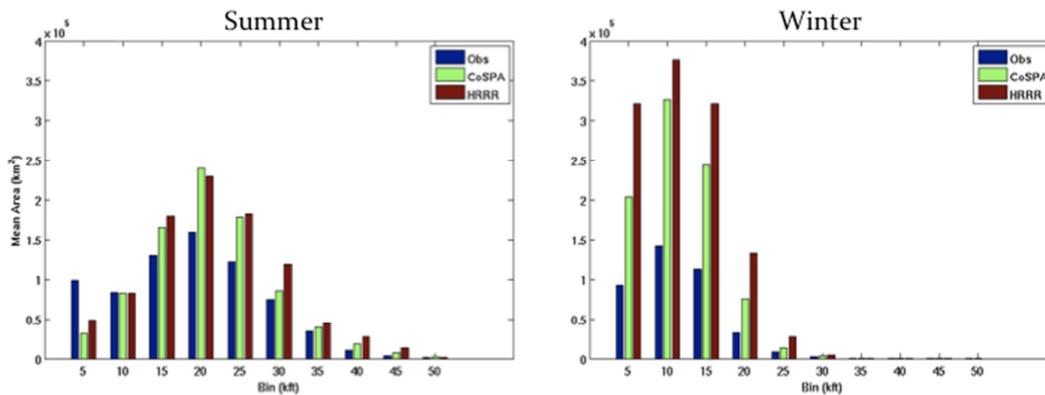


Figure 5.29. The distribution of echo tops for summer 2010 (left) and winter 2010-2011 (right) for CIWS (blue), CoSPA (green), and HRRR (red).

5.3.2 Performance Analysis

This section presents a performance analysis of CoSPA over the winter period. Note that performance metrics for the summer period are included to provide for relative comparison of forecast quality in the winter. Because convective events are less frequent in the winter, the data are stratified by 'like days' to facilitate a comparison with summer results. 'Like days' are defined as the set of days during winter that meet or exceed the average convective coverage in the summer. Note that this filter decreases the number of datasets included in the results from 70 days for the full winter period to 15 days for the 'like days' set. Some results without the 'like days' stratification are also provided in this section.

Figure 5.30 illustrates the performance, represented by the Fractions Skill Score (FSS) for CoSPA and HRRR over the SE U.S., for the 1500 UTC issuance/6-h lead time for the summer 2010 period (top) and the full winter 2010-2011 period (bottom). The overall forecast skill is reduced during the winter months as represented by lower FSS values. This is expected, and is due to the rarity of convective events as compared to summer. Interestingly, it appears that CoSPA better-integrated HRRR data via the 2010 blending algorithm during the winter time period than it did during the summer 2010 time period over the SE, as indicated by higher FSS values for CoSPA than for HRRR in the wintertime.

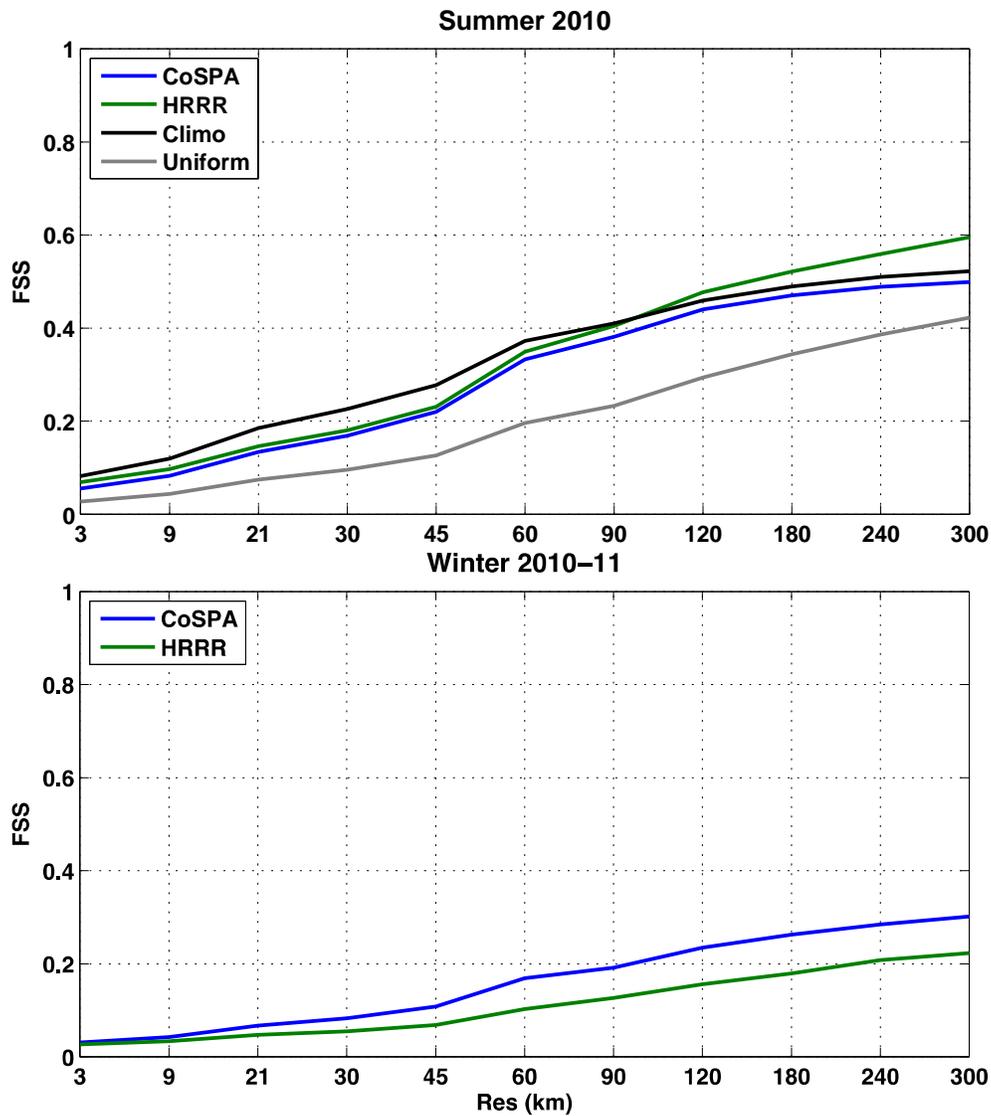


Figure 5.30. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green). Climatology (black) and uniform (gray) are included for summer 2010 (top) and not included for the winter months (bottom).

When days during the winter are stratified into ‘like days’, as shown in Figure 5.31, CoSPA performance is slightly greater than what was measured during the summer period. The wintertime improvement, for resolutions greater than 30 km, is due to larger-scale meteorological convective regimes in the SE. CoSPA also outperforms the HRRR to a greater degree in the winter than in the summer period.

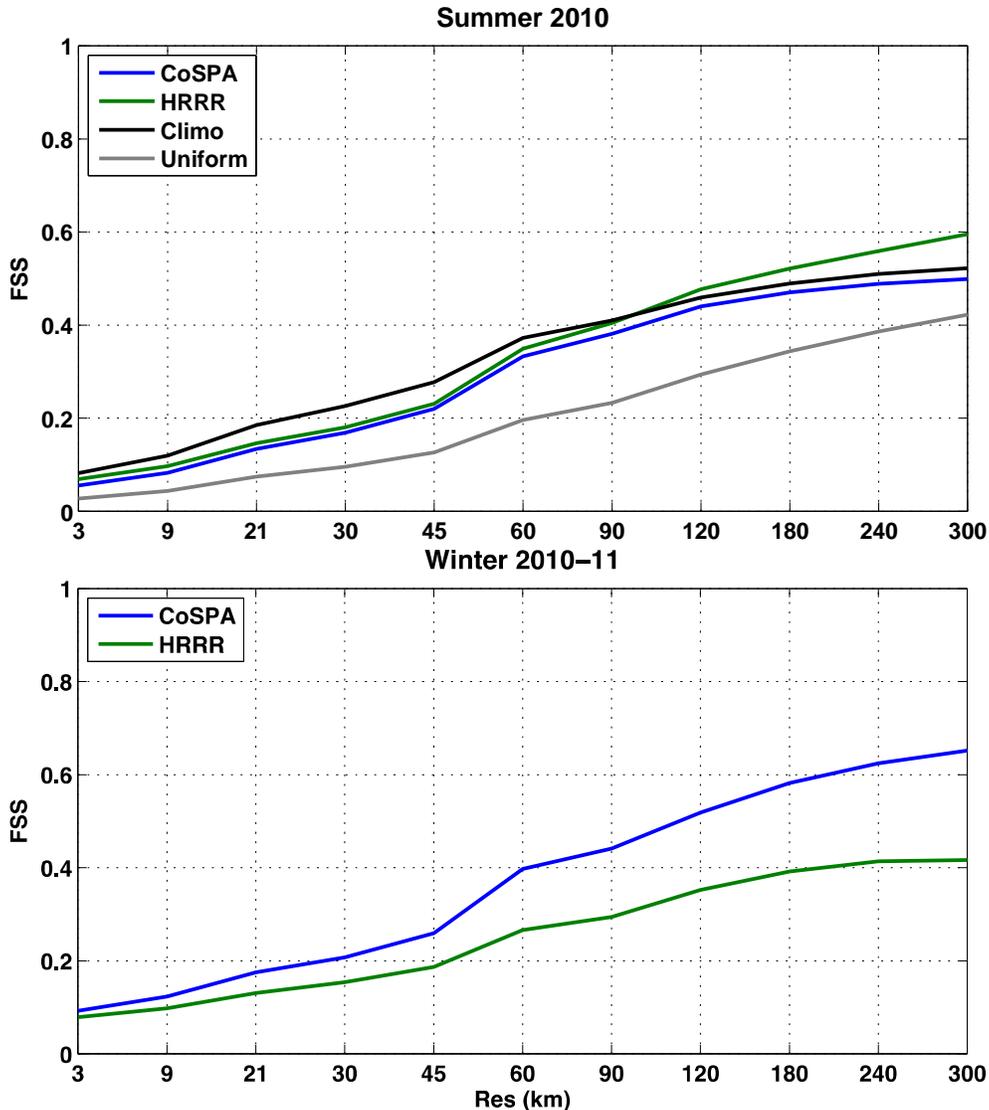


Figure 5.31. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green). Climatology (black) and uniform (gray) are included for summer 2010 (top) and not included for the 'like days' during winter months (bottom).

When examining the skill of echo tops as shown in Figure 5.32, it is apparent from the similarity of the FSS results that CoSPA gets much of its echo top information from the HRRR. The performance of CoSPA for the 25k-ft echo top threshold is nearly twice that of the skill for the 30-kft echo top threshold due to the rarity of echo tops in the winter that are greater than 30 kft. In fact, the echo top area of 25 kft and above is nearly the same as VIP-level 3 and above in the winter. Note that during the summer months, VIP-level 3 and above has nearly the same coverage as 35-kft and above echo tops.

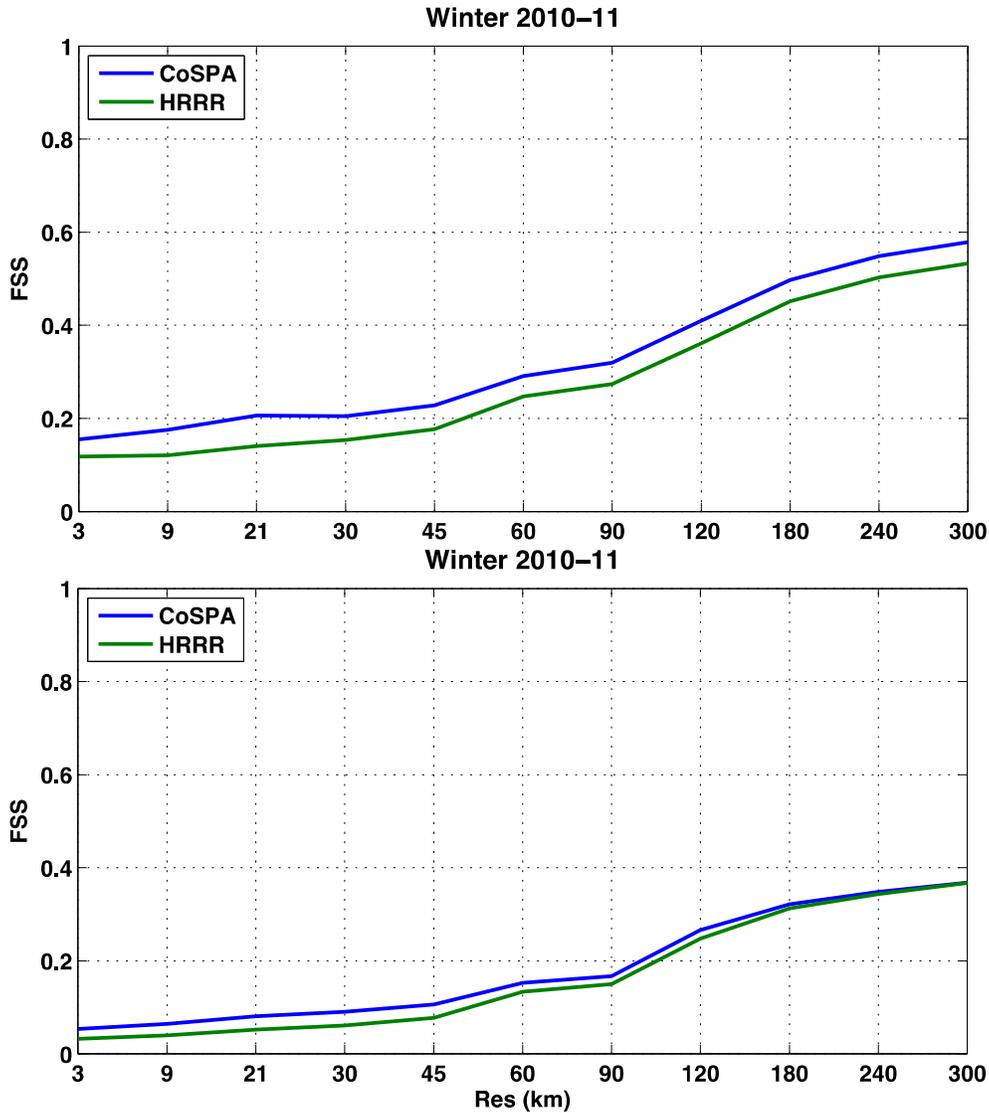


Figure 5.32. Mean FSS as a function of resolution for CoSPA (blue) and the HRRR (green) during 'like days' in winter 2010-2011 for echo tops greater than or equal to 30 kft (top) and 25 kft (bottom). HRRR latency is accounted for.

Figure 5.33 presents CSI as a function of Flow Constraint Index (FCI) threshold for the summer 2010 and the entire winter 2010-2011 periods. Overall, the skill of the winter forecast remains nearly as high as the skill of the summer forecast for convective events that have a significant impact on the NAS (events to the right of the dotted vertical yellow line). Interestingly, the skill of the winter forecast for low impact events (shown to the left of the dotted vertical yellow line) is considerably less than for summer events, most likely due to the rarity of convection in winter compared with the frequent diurnal convection found in summer. Little difference in skill between CoSPA and HRRR is noted in the winter period, as indicated by the overlap in confidence intervals.

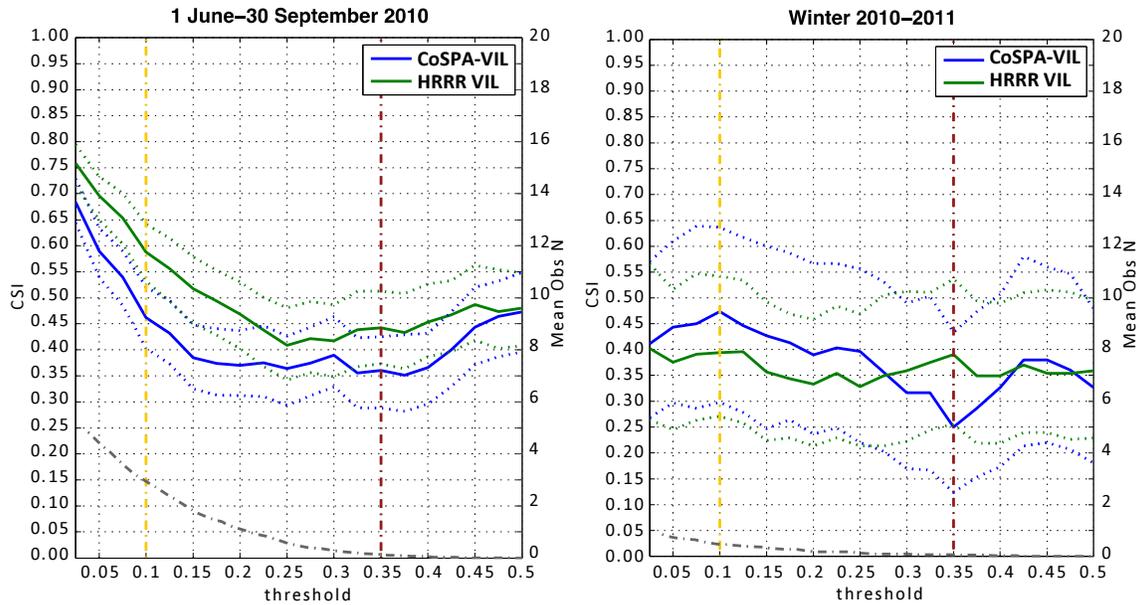


Figure 5.33: CSI as a function of FCI threshold (impact) for the summer 2010 (left) and winter 2010-2011 (right) for the SE U.S. at ARTCC scale for forecasts issued at 17 UTC and a 6-h lead time. CoSPA VIL is shown in blue and HRRR VIL is in green. The gray-dashed line referring to the right y-axis gives the average number of ARTCC hexagons impacted at the given threshold. The yellow-dashed vertical line is a medium impact threshold and the maroon-dashed line is a high impact threshold.

Figure 5.34 and Figure 5.35 show the CSI and bias for CoSPA during ‘like days’ in the winter period for the 1500 UTC issuance as a function of lead time for three spatial resolutions. Skill during the winter period for ‘like days’ is consistent with skill computed for the summer period, but with greater variability in the winter, which may be due to the smaller sample size. The increase in skill may result from overforecasting that was observed in the bias statistic for 3- to 5-h lead times. The underforecasting problem found in CoSPA during the summer months was due to the poor blending during this period in the SE. Also, similar to the summer, performance of CoSPA remains high for the 6-h to 8-h lead times.

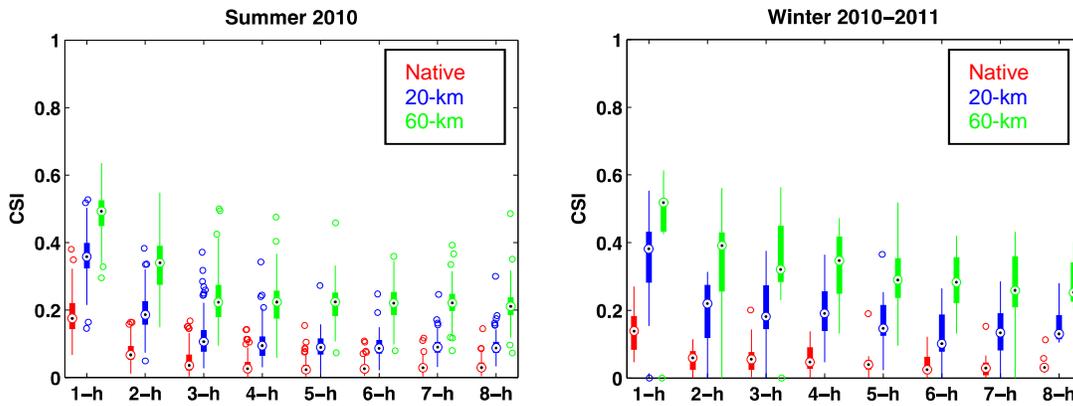


Figure 5.34: CSI as a function of lead time and resolution for summer 2010 (left) and significant days in winter 2010-11 (right) for CoSPA at VIP-level 3 at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.

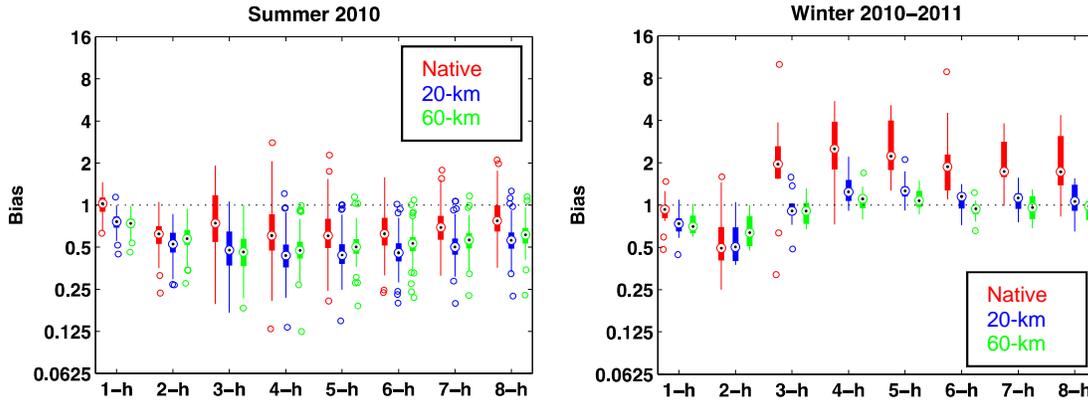


Figure 5.35: Bias as a function of lead time and resolution for summer 2010 (left) and significant days in winter 2010-11 (right) for CoSPA at VIP-level 3 at 1500 UTC. Native resolution is shown in red, 20-km in blue, and 60-km in green.

6 Summary and Conclusions

The goals of this evaluation are to assess: 1) the quality of CoSPA for use in traffic flow planning with a particular focus on the quality of the modifications introduced into the 2011 version of the CoSPA algorithm; 2) the quality of CoSPA as a supplement to the CCFP, since TFM planners are currently instructed to use CoSPA in conjunction with CCFP for developing traffic flow plans, and; 3) the quality of CoSPA during winter months for capturing convective weather over the southeast U.S., as CoSPA may be provided to users for decision-making year round.

Since a true performance baseline is costly to achieve and was not available from 2010 to 2011, the findings presented in this report should be interpreted as relative measures of forecast quality and indicators of forecast improvement. Note: a full

baseline study is needed in order to measure the absolute change in forecast quality of any future CoSPA algorithmic modifications.

There were a few findings in the 2011 study that were similar to the 2010 study. CoSPA outperformed all other products by a significant margin at the 2-h lead time. It should be noted that forecasts at the 2-h lead time are provided by the CIWS extrapolation product. The discontinuity from the 2:00-h lead time to the 2:15-h lead time continues to exist, and will be reflected in the CoSPA display. Additionally, CoSPA skill does not degrade much from the 6- to 8-h lead times, indicating that it may be useful at longer lead times which are not available from CCFP. Finally, the operational reliability of the product, with and without scheduled maintenance, was determined to be nearly the same from 2010 to 2011, with slightly less reliability measured in 2011. HRRR, as used by CoSPA, maintained approximately a 3-h latency during 2011.

During the summer period, a substantial change to the blending algorithm occurred as dynamic weights were applied over the CONUS to more effectively utilize information from the HRRR. This enhancement to the blending scheme vastly improved CoSPA performance in the SE at the 6-h lead time and beyond. However, the results indicate that there is an underforecasting problem in both VIL and echo tops in the 3- to 5-h lead-time frame, more notable in the echo top field. This is most likely due to enhancements noted in the blending scheme.

When stratifying results at strategic issuance and lead times greater than or equal to 6 h, CoSPA shows increased skill over other products during weakly forced days that cause ground delay programs (GDPs) in the NE. This is most likely due to the convection-resolving HRRR model. This was shown in both 2010 and 2011. CoSPA and CCFP (re-categorized) possessed similar skill during AFP issuance days, which tend to be strongly forced, structured convective events that are often easier to forecast than weakly forced convective events.

When applying a cluster methodology to the forecasts to create CCFP-like objects, CoSPA performs about the same in predicting sparse coverage, low confidence polygons in both the strategic time frame and in post-initiation issuance times. CoSPA also performs similarly to CCFP and LAMP. There is a slight decrease in skill in all products at the late lead times for the post-initiation issuances, highlighting the difficulty in forecasting convective cessation. At thresholds of sparse coverage/high confidence where TFM may need to take action, CoSPA retains higher skill than the other products. When CoSPA is used as a supplement to CCFP sparse coverage/low confidence polygon forecasts, confidence in whether to use the polygons for traffic flow planning increases.

The quality of CoSPA during the winter months was assessed as though the 2010 version of the CoSPA summer algorithm would remain available to the users all year long. Therefore, the algorithm was only assessed in areas where convection was present, mainly over the SE U.S. The analysis indicates that CoSPA performed well

over the SE during the winter months, with comparable skill measured for the summer months. In addition, CoSPA outperformed the HRRR in the SE, possibly due to the strongly-forced, well-defined convection that is present in the SE during winter, whereas the summer convection is diurnally driven, often ill-defined, and harder to forecast. Since echo tops are typically found below 30 kft in winter and echo tops below 30 kft are not provided to the user through the CoSPA display, decision makers may find it difficult to use this product for winter-time strategic planning.

7 References

Dupree, W., J. Pinto, M. Wolfson, S. Benjamin, M. Steiner, J. Williams, D. Morse, X. Tao, D. Ahijevych, H. Iskenderian, C. Reiche, J. Pelagatti, and M. Matthews, 2009: The Advanced Storm Prediction for Aviation Forecast Demonstration, WMO Symposium on Nowcasting, Whistler, B.C., Canada.

FAA, 2011: CoSPA users' guide v2.0 for 2011 summer season, NextGen and Operations Planning, Research & Technology Development, Aviation Weather, New Weather Capabilities, AJP-682.

Ford, L.R., Jr., and D.R. Fulkerson, 1956: Maximal flow through a network. *Canadian Journal of Mathematics*, **8**, 399-404.

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 Conference on Numerical Weather Prediction, Washington, D.C.

Iskenderian, H, J. Pinto, M. Wolfson, S. G. Benjamin, M. Steiner, S. S. Weygandt, C. Alexander, W. J. Dupree, J. K. Williams, J. R. Mecikalski, W. F. Feltz, K. Bedka, D. Morse, X. Tao, D. A. Ahijevych, C. Reiche, T. Langlois, K. L. Haas, L. J. Bickmeier, P. M. Lamey, J. M. Pelagatti, and D. D. Moradi, 2011: Update on CoSPA Storm Forecasts, 15th Conference on Aviation, Range, and Aerospace Meteorology (ARAM). Los Angeles, CA.

Krozel, J., S. Penny, J. Prete, and J. S. B. Mitchell, 2004: Comparison of Algorithms for Synthesizing Weather Avoidance Routes in Transition Airspace, *AIAA Guidance, Navigation, and Control Conf., Providence, RI*.

Lack, Steven A., G.J. Layne, S. Madine, and J.L. Mahoney, 2011: Quality Assessment of CoSPA, NOAA Technical Memorandum OAR GSD-38.

Lack, S.A., G.J. Layne, M.P. Kay, S. Madine, and J. Mahoney, 2010a: Evaluating supplemental relationships between convective forecast products for aviation. 14th Conference on Aviation, Range, and Aerospace Meteorology (ARAM). Atlanta, GA.

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

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

Layne G.J., M.S. Wandishin, M.P. Kay, S.A. Lack, M.A. Petty, and J.L. Mahoney 2012: Relating a convective translation metric to convective impact, Third ARAM Special Symposium on Weather-Air Traffic Management Integration, New Orleans, LA.

Roberts, N.M. and H.W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, 136, 78–97.

Stensrud, D. J. and M.S. Wandishin, 2000: The correspondence ratio in forecast evaluation. *Wea. Forecasting*, 15, 593–602.

Weygandt, S., T. Smirnova, C. Alexander, S. Benjamin, K. Brundage, B. Jamison, and S. Sahn, 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.

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

Acknowledgements

This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

We would like to thank Matt Fronzak from MITRE for stratifying delay data from the FAA's Operations and Performance databases.

Appendix

2011 Uptime Statistics- (HRRR typically has 3-h latency)

Uptime for 6/1-9/29: **95.6%** Uptime with scheduled maintenance: **96.4%**

2010 Uptime Statistics- (HRRR typically has 3-h latency)

Uptime for 6/1-9/29: **96.0%** Uptime with scheduled maintenance: **98.0%**