NOAA Technical Memorandum OAR GSD-35

2007 CONVECTIVE FORECAST SCIENTIFIC EVALUATION

M.P. Kay
S. Madine
J. Luppens Mahoney
J.E. Hart

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

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

NOAA Technical Memorandum OAR GSD-35

2007 CONVECTIVE FORECAST SCIENTIFIC EVALUATION

Michael P. Kay
Sean Madine
Jennifer Luppens Mahoney
Joan E. Hart

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

UNITED STATES
DEPARTMENT OF COMMERCE

Gary Locke
Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

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

Office of Oceanic and
Atmospheric Research

Dr. Richard Spinrad
Assistant Administrator

Contents

Section | Page

2007 Convective Forecast Scientific Evaluation

Quality Assessment Product Development Team

21 December 2007

Executive Summary

The goal of this scientific evaluation is to assess the performance of several convective forecasts with respect to their application to operational air traffic flow planning. Five forecasts, including Collaborative Convective Forecast Product (CCFP) Preliminary and Final, Rapid Update Cycle (RUC) Convective Probability Forecast (RCPF), RUC Reflectivity, and the North American Mesoscale model (NAM) Reflectivity were evaluated using the National Convective Weather Detection (NCWD) product from 11 June – 31 August 2007. This Executive Summary highlights the main aspects of the scientific evaluation, but further detail and analyses are summarized in a comprehensive report.

The main verification approach applied in the study was used to intercompare the forecast quality of the five products at strategic flight planning time periods and within impacted sectors. This unique measure of forecast quality was linked directly to the application of the convective forecasts to the operational flight planning process.

Relevant results from the study indicated:

- Nearly identical forecast performance from the CCFP Preliminary and the CCFP Final
- The CCFP and RCPF performed similarly for nearly all time periods, except at the 2-h outlook period where CCFP performed slightly better.
- At early valid times and for shorter outlook periods, CCFP performed slightly better than RCPF.
- At later valid times and longer output periods (i.e., when convective weather has the potential to severely impact air traffic), the RCPF performed as well as the CCFP.
- Analysis of the top ten high-impact air traffic days indicated that the performance of the RCPF at the 8-h outlook period for the afternoon shows some promise for planning purposes.
- On high-coverage, high-impact days neither CCFP nor RCPF performed significantly different from the other.

- The CCFP, for every valid time of interest, better identified the sectors that were impacted by convection than did the RCPF.
- The reflectivity products (NAM and RUC) showed virtually no skill at forecasting hazardous convection, but did provide some guidance at long lead periods for areas of concern.
- The probabilistic aspects of CCFP and the RCPF indicated low reliability.

Recommendations that were identified from the results include:

- Meteorologically, the forecast skill of the CCFP Final and Preliminary perform similarly, thus we recommend that the meteorological collaboration and its relationship to the planning process be further evaluated before possible elimination.
- In the near term, RCPF should be used as input to the CCFP generation process.
- In the longer term, we recommend adoption of a gridded probabilistic forecast as meteorological input to the traffic planning process.
- Since the radar reflectivity products have in some cases alerted planners to course areas of hazardous weather 8-24 h in future, we recommend further resources be devoted to the development of these products to improve their ability to better forecast convective intensity and structure. This may also benefit other automated convective forecasts.

1. Introduction

The planning process instituted by the FAA Air Traffic Control System Command Center (ATCSCC) uses specific weather information to develop a strategic air traffic flow plan. The plan is often implemented to reroute traffic when hazardous convective weather occurs within the National Air Space (NAS). In order to better understand the application of convective weather forecasts into the ATCSCC planning process, five convective forecast products were objectively evaluated from 11 June – 31 August 2007 at key strategic decision points throughout the day when the impact of convective weather on the air traffic flow is often the greatest. This report is a supplement to a subjective evaluation performed by AvMet Applications (Phaneuf and Simenauer 2007).

This report summarizes the results from the objective evaluation and is organized as follows: Section 2 describes the data used in the evaluation. Section 3 details the methodology of the intercomparison of the products. The results of the intercomparison and the stratification by air traffic impact appear in Section 4. Finally, conclusions are presented in Section 5.

2. Data

2.1. Forecasts

A broad cross section of operational and research forecasts were studied (Table 1). The CCFP Final, CCFP Prelim, and NAM simulated composite reflectivity products are all operational forecasts that are produced and disseminated every day by operational centers. The RCPF and RUC simulated composite reflectivity products are experimental.

Table 1. Forecast products evaluated.

Short Name	Long Name	Spatial Resolution (km)	Temporal Resolution (h)	Issuance Frequency (h)	Latency (h)
CCFP Final	Collaborative Convective Forecast Product Final	polygons	2	2	0
CCFP Prelim	Collaborative Convective Forecast Product Preliminary	polygons	2	2	0
RUCSR	Rapid Update Cycle (RUC) Model Simulated Reflectivity	13	1	1	2
RCPF	RUC Convective Probability Forecast	20	1	1	2
NAMSR	North American Mesoscale Model Simulated Reflectivity	12	3	6	2

2.2. Observations

The National Convective Weather Detection (NCWD) was used as the observation field for convection. The NCWD represents the best CONUS-scale operational depiction of convection available. The 40-dBZ value is considered to be the threshold for convection that is hazardous to aviation. The NCWD is defined in terms of video integrator processor (VIP) levels instead of reflectivity values. The actual threshold value used to indicate hazardous convection was VIP level 3, which is approximately equal to 40 dBZ.

3. Methodology

This section introduces the verification methodology and terminology developed for the intercomparison. In order to tie the verification analysis to the ATCSCC decision-making process, several temporal forecast attributes are defined. Before introducing the new terms, it is necessary to enumerate the basic temporal attributes that form the fundamental basis for the intercomparison**.**

- **Forecast initial time** is the time the model "cycle" begins (time of initialization of the model) or when the forecast is created (in the case of a human-produced forecast like CCFP Final).
- **Forecast available time** is the actual clock time at which a given set of model outputs (for several **valid** times, but linked to a single **initial** time) is available to an end user of the product.
- **Forecast latency** is the elapsed clock time from the **initial** time until the **available** time
- **Forecast lead period** is the elapsed clock time from **available** time until the **valid** time (Note: not the traditional "lead time"). The lead period takes into account the real-world problem of only being able to use what forecast data is available when decisions are made.

The following terms are used to provide a common terminology for the evaluation of the forecasts in relation to ATCSCC strategic planning activities.

- **Telecon times** are the series of strategic planning teleconferences (telecons) conducted by the ATCSCC every two hours during the day to adjust air traffic routing over the CONUS.
- **Outlook periods** are key decision points at 2-, 4-, 6-, and 8-hours that occurred during each telecon time.
- **Outlook time** is the actual time of day and is computed from the telecon time + outlook period.

To summarize, each forecast gains a set of additional temporal attributes that allow the forecast to be specified relative to important strategic planning times at the ATCSCC. Product latency becomes an important part of the analysis because it dictates the currency of the information available to be used for planning purposes. Forecasts can then be compared in an operational context, in a way that mimics the use of what information is available to the user on a routine basis.

Three strategic planning telecon times and the corresponding key decision points were the focus of this study (1115 UTC, 1315 UTC, and 1515 UTC). They represent operationally important strategic planning periods during the afternoon when convective coverage typically has the greatest traffic impact. The complete set of forecast initial times and outlook periods are shown in the tables of Appendix 1. The tables highlight the fact that at each telecon time, the ATCSCC planners use the latest forecast information that is available, thus adjusting for the forecast latency.

3.1. Verification Approaches

The intercomparison was done using two techniques: a traditional approach, which use the gridded forecasts at a grid box by grid box granularity and a sector-based approach, which is used to evaluate the forecasts with respect to impacted air traffic sectors. Within the traditional approach and where appropriate, forecasts were analyzed both deterministically and probabilistically. The forecast products from Table 1 used for these analyses represent a diverse spectrum of prediction types: categorical/dichotomous (CCFP Final, CCFP Prelim), probabilistic (RCPF), and continuous (RUCSR, NAMSR). To intercompare the forecasts in a meaningful way, some of the forecasts needed to be transformed. The transformations, shown in Table 2, attempt to preserve the way these forecasts are perceived and used operationally.

Table 2. Transformations applied to forecasts to facilitate intercomparisons. The dichotomous transformations are also used for the sector-based approach.

Forecast	Deterministic Transformation	Probabilistic Transformation
CCFP Final	Existence of polygon means forecast of event	Mean polygon forecast probability becomes forecast probability
CCFP Prelim	Existence of polygon means forecast of event	Mean polygon forecast probability becomes forecast probability
RCPF	Threshold at 40%	
RUCSR	Threshold at 40 dBZ	Not used
NAMSR	Threshold at 40 dBZ	Not used

3.1.1. Traditional

In this study, the traditional grid-to-grid verification approach is used as the primary verification technique for evaluating the quality of the convective forecasts for both the dichotomous and the probabilistic forecast types. The forecasts are bilinearly interpolated to a common 4-km grid and compared to a grid of observations that are closest in time to the forecast valid time (typical time difference between a valid forecast and an observation is less than five minutes).

3.1.2. Sector-based

An alternate sector-based verification approach, which may be more appropriate for evaluating the operational application of the forecasts to ATCSCC operations was developed for this study. The fundamental unit of measure in this approach becomes the air traffic sector rather than the 4-km grid box. Air traffic sectors are useful because they represent the volumes that are used for strategic air traffic planning. Using the definition from this approach, a sector is defined as impacted if either the areal coverage of the forecast or observed convection exceeds five percent. This areal coverage threshold was chosen by studying the observed areal coverage value of all sectors during the month of June 2007 (not shown). For this study, the “super high” sectors were used for the evaluation, since they represent the set of sectors that are specific for en-route air traffic. In contrast to the traditional approach, the sector-based approach uses observations that are gathered over a one-hour time smear centered on the forecast valid time. This temporal smear was done to better capture the idea that sectors are impacted over time and not instantaneously.

An example graphic highlighting a CCFP Final forecast with sectors colored by how the forecast verified is shown in Fig. 1. A squall line moving into the Tennessee Valley dominates the weather situation. The CCFP forecast correctly predicts the sector impacts through much of the extent of the forecast polygons. However, the northern polygons were considered false alarms – areas where events were forecast, but did not occur. Convection occurring over the southeastern U.S., ahead of the squall line, was not captured by the CCFP, and the sectors were considered missed events.

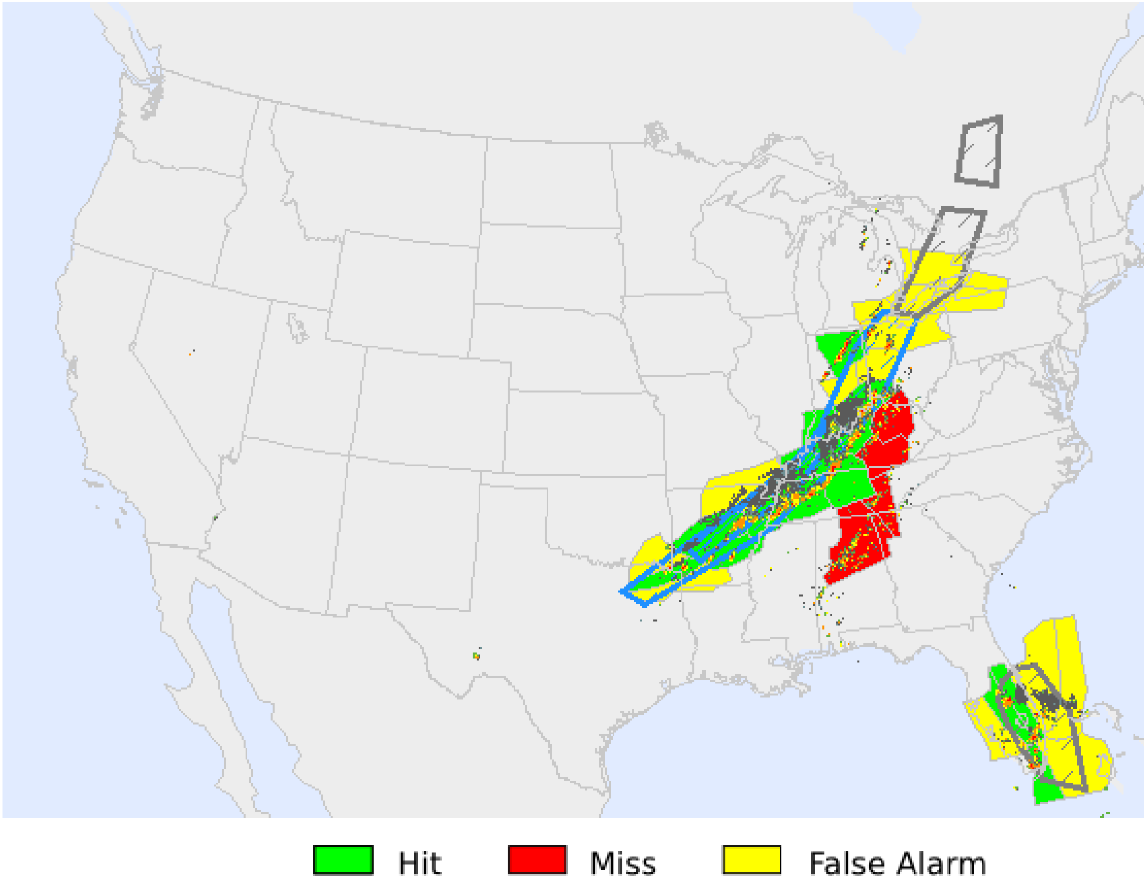


Figure 1. Sector-based verification of the 2-h CCFP Final forecast from 8 June 2007 issued at 1500 UTC. NCWD observations are shown as well. Impacted sectors are color-coded to depict the verification results.

3.2. Verification Statistics

In this study a limited subset of statistics were used to describe the important aspects of the forecast quality. These statistical measures are presented in Table 3.

The four possible event scenarios that comprise the statistical measures in a dichotomous setting are:

- Forecast Yes, Observed Yes (denoted YY)
- Forecast Yes, Observed No (YN)
- Forecast No, Observed Yes (NY)
- Forecast No, Observed No (NN)

The ATCSCC strategic planners are concerned with the amount of convection captured by the forecasts, forecast biases, and overall forecast performance. These forecast characteristics are summarized by the statistics chosen for this analysis and are presented in Table 3.

Table 3. Statistical measures used in this study.

Score	Long Name	Definition	Interpretation
POD	Probability of Detection	$POD=\frac{YY}{YY+NY}$	The amount of convection occurring within the forecast areas. Values range from zero to one.
BIAS	Bias	$BIAS=\frac{YY+YN}{YY+NY}$	The amount of over- or underforecasting. A bias of 1 means the forecast occurs as much as the observation but says nothing about location. Values range from zero to infinity.
CSI	Critical Success Index	$CSI=\frac{YY}{YY+YN+NY}$	The ratio of the amount of correctly predicted convection to the amount of convection that was either forecast or observed. Values range from zero to one.

4. Results

In this section, the results from all of the intercomparisons performed are presented. Note that the transformations necessary to perform the verification limit the extent to which the results, and therefore conclusions, should be viewed as absolute measures. For example, the change to a forecast threshold could significantly impact the overall statistical performance of the forecast (e.g., if the RCPF data was thresholded at 25% instead of 40%, the statistical results could be quite different). Similarly, for the sector-based analysis, a change in the sector coverage threshold could also lead to differing results. The three primary thresholds that affect the verification results are the forecast thresholds for the RCPF and simulated radar reflectivity fields, the observation thresholds, and the sector coverage threshold. The RCPF threshold of 40% was chosen from past research (Weygandt and Benjamin, 2004 and personal communication with S. Weygandt, 2007). The NCWD, NAMSR, and RUCSR threshold of 40 dBZ was chosen due to past communication with end users and strategic planners. The sector coverage threshold was derived from a climatological analysis of observed sector coverage values during June 2007.

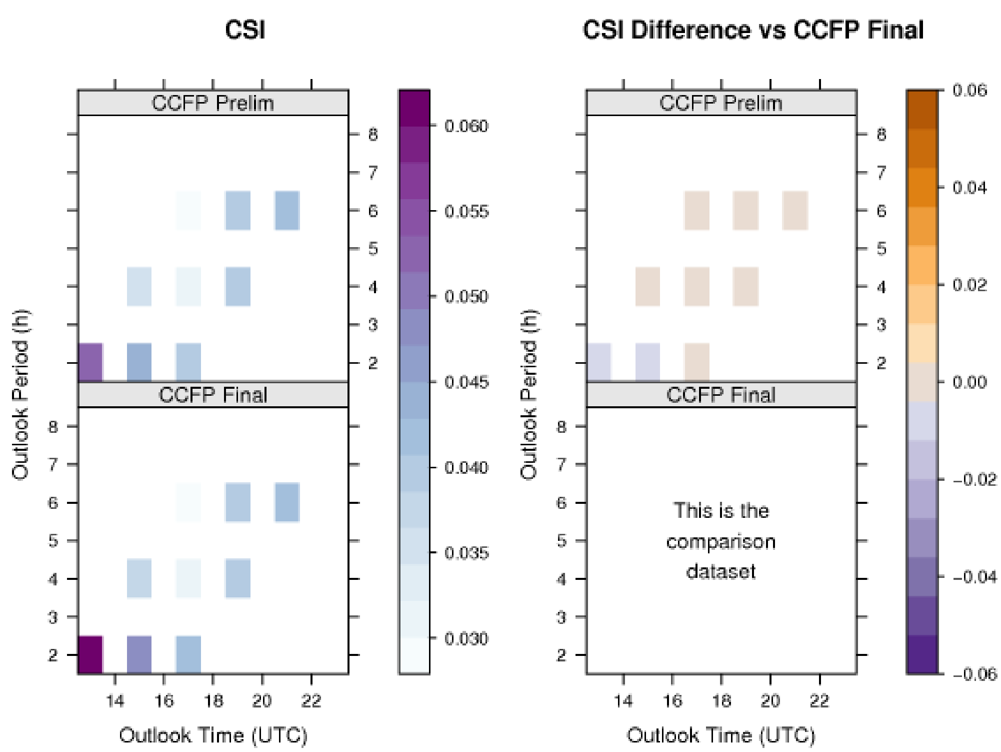
4.1. CCFP Final and CCFP Prelim

This section summarizes the differences in performance of the CCFP Prelim and

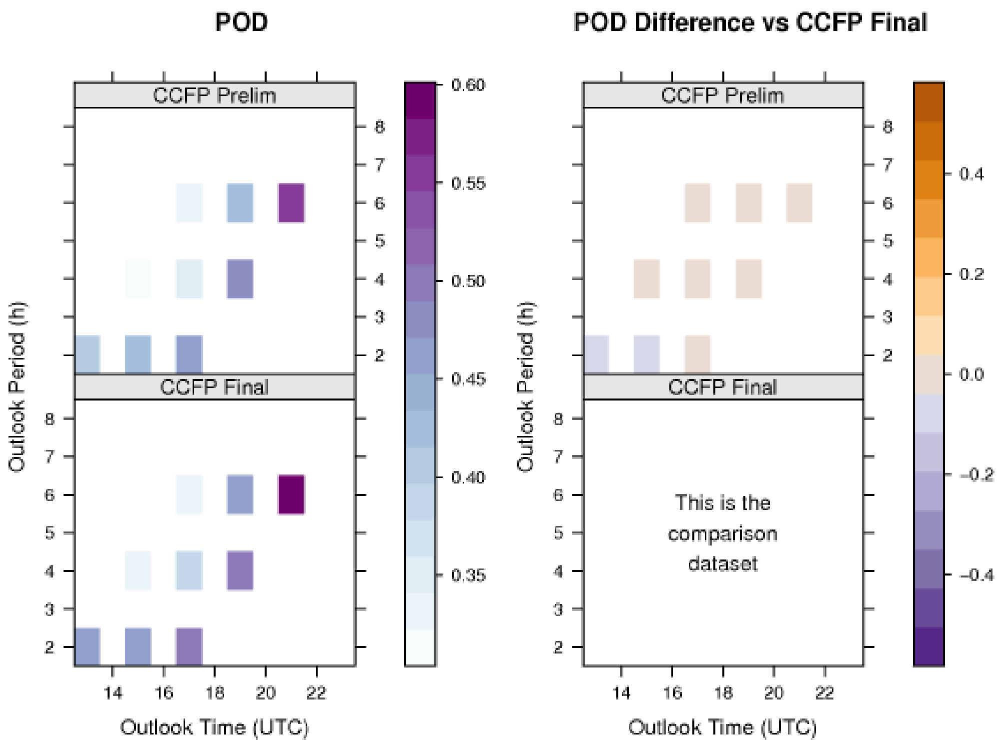
CCFP Final forecasts. Operationally, a significant amount of time and effort is placed on coordinating and creating the CCFP Final forecast from the initial forecast put forth by the CCFP lead forecaster at the Aviation Weather Center. It is therefore prudent to understand if the CCFP Final forecast shows any noticeable improvements from its preliminary counterpart.

The results, shown in Fig. 2, reveal that, in terms of the verification of CCFP, the two forecasts perform nearly identically for all of the performance measures used in this study. Values for POD, BIAS, and CSI are within a few percent of each other for all outlook times and outlook periods studied. The largest differences in forecast performance are noted for the 2-h outlook period and for outlook times that occur during morning hours (i.e., 1300 and 1500 UTC) when convective impact on air traffic is at a minimum. For longer outlook periods (4- and 6-h) and outlook times during the day when convective impact is large, the CCFP Prelim and CCFP Final perform nearly identically. These results may suggest that the initial polygon placement as represented by the CCFP Prelim forecast primarily governs the placement of the Final forecast, but that the differences between the two forecasts occur for other attributes, such as the confidence attribute, which was not directly evaluated in this study. Future work should be done to identify the extent to which the CCFP Final and Prelim forecasts differ on a polygon-by-polygon basis.

a)



b)



POD
POD Difference vs CCFP Final
CCFP Prelim
CCFP Final
Outlook Period (h)
Outlook Time (UTC)
This is the comparison dataset
0.60
0.55
0.50
0.45
0.40
0.35
0.4
0.2
0.0
−0.2
−0.4

c)

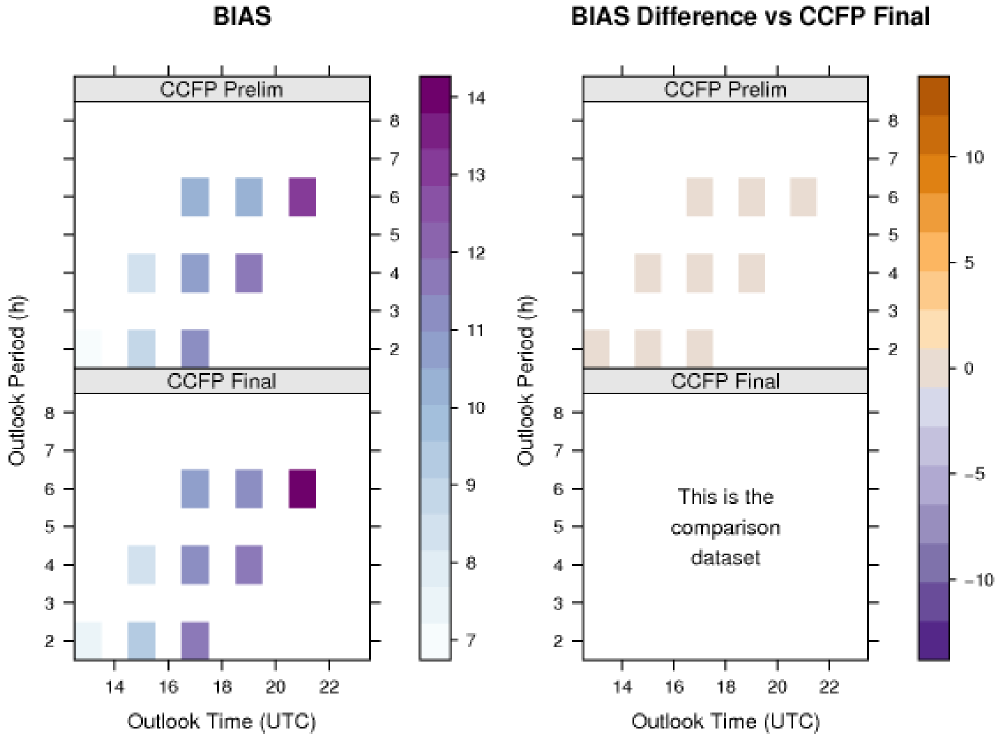


Figure 2. Level plots of a) CSI, b) POD, and c) BIAS for each outlook period and outlook time in the telecon-constrained dataset. Colored squares on left-hand panels represent the value of the given statistic for each forecast time. The right-hand panels depict the score differences relative to those for CCFP Final; warm colors are where the CCFP Prelim value exceeds the value for CCFP Final and cool colors are where CCFP Prelim values are smaller.

4.2. CCFP Final and RCPF

The CCFP Final represents the current operational convective forecast used for strategic air traffic management planning. The RCPF represents a fully-automated prediction product that provides similar convective forecast information as the CCFP forecasts (Weygandt and Benjamin, 2004). In a time of increasingly automated forecasting systems, it is important to understand the state of the art for both human- and computer-generated predictions for aviation forecasting. Note: Readers should keep in mind that when assessing these results for the RCPF and the CCFP, that in some instances, the RCPF may have been used as input to the generation of the CCFP, thus impacting the independent evaluation of the two products.

4.2.1. Traditional Approach

Figure 3 shows forecast performance as a function of outlook period for the telecon-constrained data. At the 2-h outlook period, the CCFP Final forecasts have larger CSI and POD values when compared to RCPF. In an absolute sense, the CSI values are remarkably low for both products resulting from the high forecast biases. The CCFP Final captures approximately 48% of all observed hazardous convection, compared to 28% for the RCPF.

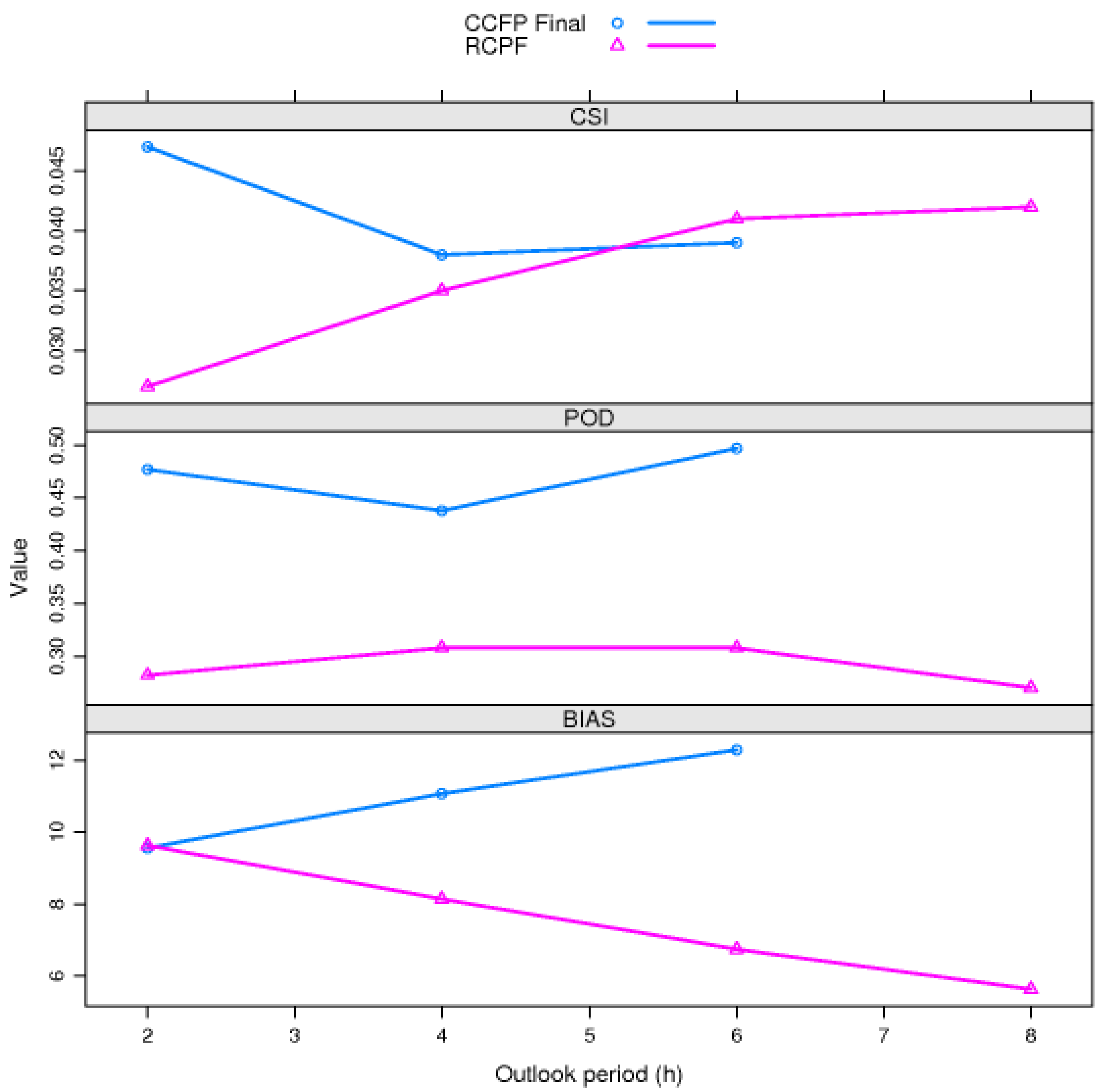


Figure 3. CSI, POD, and BIAS values for the CCFP Final and RCPF forecasts as a function of outlook period for the all telecon-constrained forecasts.

At outlook periods beyond two hours (Fig. 3), the overall skill, as illustrated by the CSI, becomes comparable between the two products. For the longer outlook periods, the RCPF bias decreases significantly, while the CCFP Final bias increases, suggesting less overforecasting by the RCPF at longer outlook periods than for CCFP. Overall, POD values remain relatively constant for each product. The RCPF forecast skill is maintained for the 8-h outlook period.

Figure 4 illustrates the forecast skill by outlook time. Outlook times between 1900 and 2300 UTC are typically key decision points for traffic flow planning. During this time period, the RCPF and CCFP Final forecasts show distinctly different trends in POD and BIAS. The CCFP forecasts show best skill (CSI) in the morning with scores reaching a minimum at 1700 UTC and remaining approximately constant through 2300 UTC. The RCPF scores quite poorly until 1700 UTC, but then improves quickly, reaching a maximum CSI of 0.047 at 2100 UTC. During this period the CCFP Final BIAS values increase from 7 to nearly 14, while the RCPF BIAS decreases from 8 to 6.

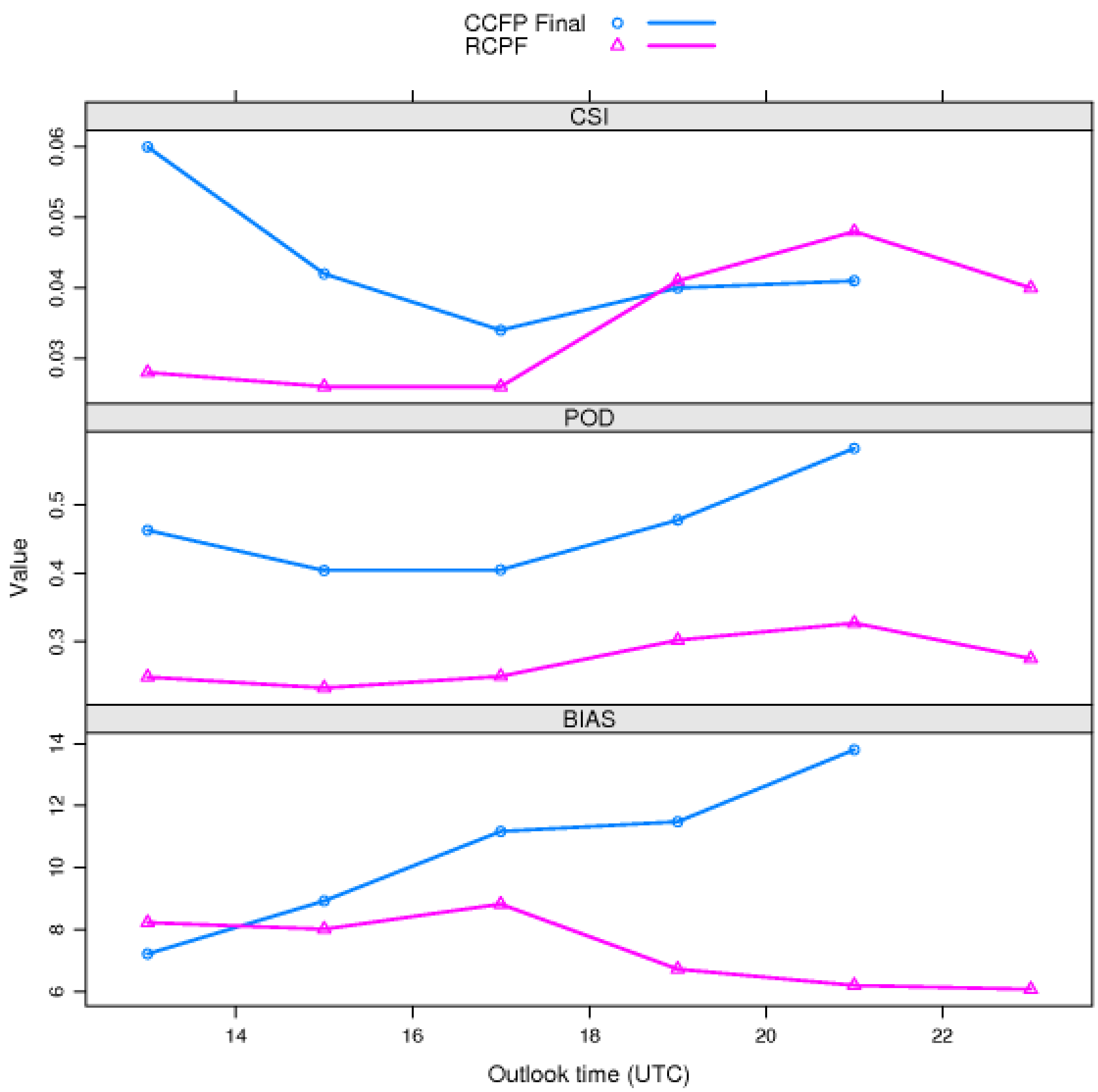


Figure 4. CSI, POD, and BIAS values for the CCFP Final and RCPF forecasts as a function of outlook time for the all telecon-constrained forecasts.

The lower CSI for RCPF in the morning may be due to convection occurring in the model in the wrong location. The morning times, when initiation is typically at a minimum during the warm season, appears to be dominated by ongoing convection. In these circumstances, CCFP forecasters are able to capture the location and movement of convection quite well. In contrast, as the day progresses, the CCFP forecasters appear to struggle with convective initiation and therefore issue very large areas as a result. A complete view of the diurnal cycle of the verification statistics covering the entire daily period is shown in Appendix 2. Additionally, detailed diagnostic plots covering all outlook times and outlook periods for all five forecasts studied in this project are presented in Appendix 3.

4.2.2. Sector-based Approach

The trends in the sector-based verification results are similar to those from the traditional approach with some important differences. Overall, for both forecasts, the POD and CSI values increased while the BIAS values decreased when compared to the results from the traditional approach. CCFP Final CSI values exceed RCPF for all outlook periods, which indicates that the CCFP Final does a better job of forecasting impacted sectors than the RCPF. At longer outlook periods, the CCFP has higher CSI and POD values that correlate well with its BIAS values and are indicative of minor overforecasting (Fig. 5). CCFP Final biases increased in the afternoon to levels indicating minor overforecasting and RCPF biases decreased to levels indicative of minor underforecasting. The overall trends in the BIAS curves are similar to those shown in Fig. 3 for the traditional approach. Both forecasts severely overforecast using the traditional approach and are approximately unbiased at forecasting impacted sectors. When the results are viewed by outlook time (Fig. 6), the CCFP Final performs better than RCPF at 1900 UTC and 2100 UTC compared to its performance in the traditional verification (Fig. 4) where it performed worse than RCPF. Overall, the sector-based results indicate that both forecasts do a significantly better job at forecasting impacted sectors rather than convection occurring at the grid-box scale. Detailed diagnostic plots that provide additional information for the sector-based verification may be found in Appendix 4.

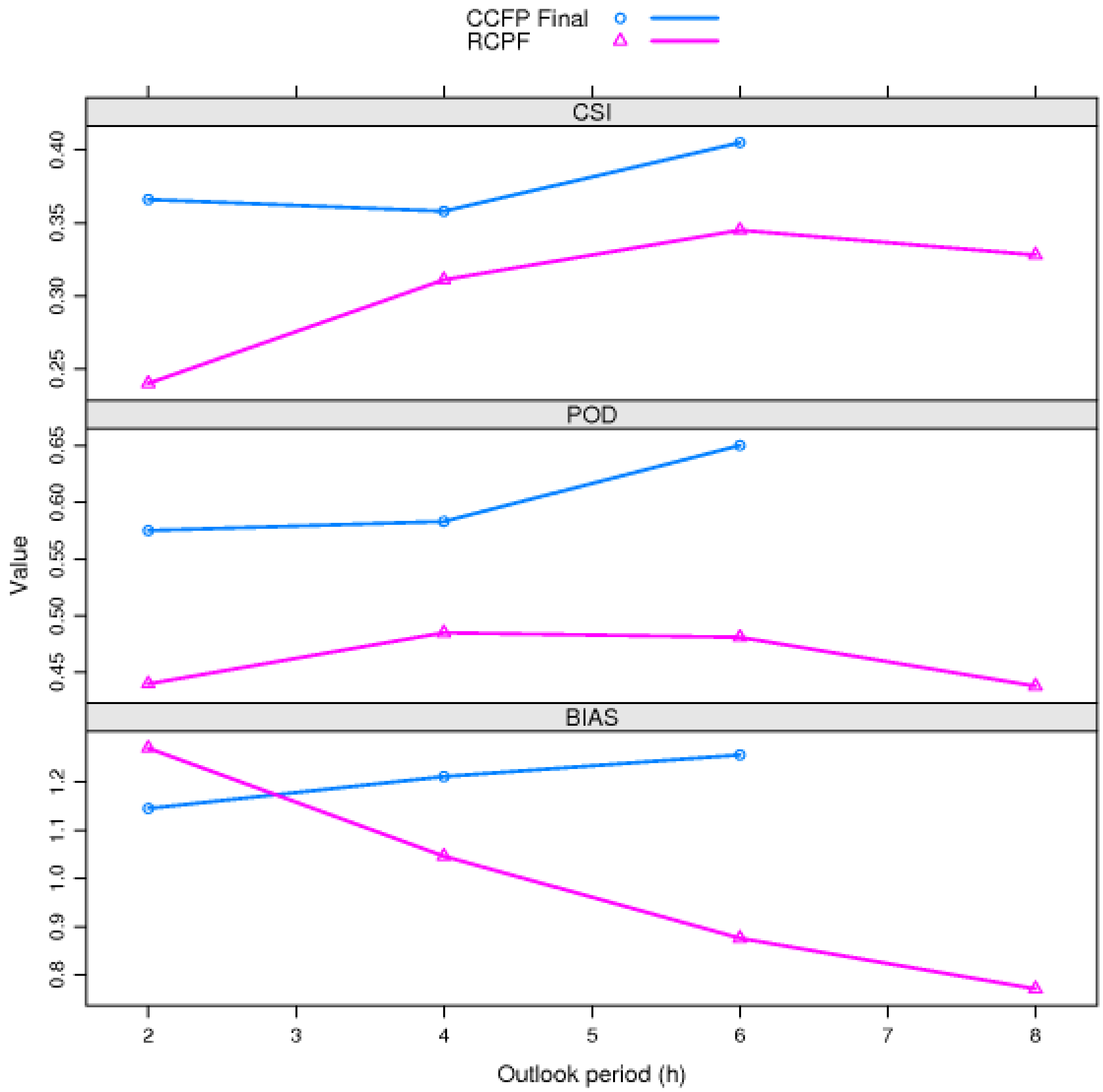


Figure 5. As in Figure 3 except for the sector-based approach.

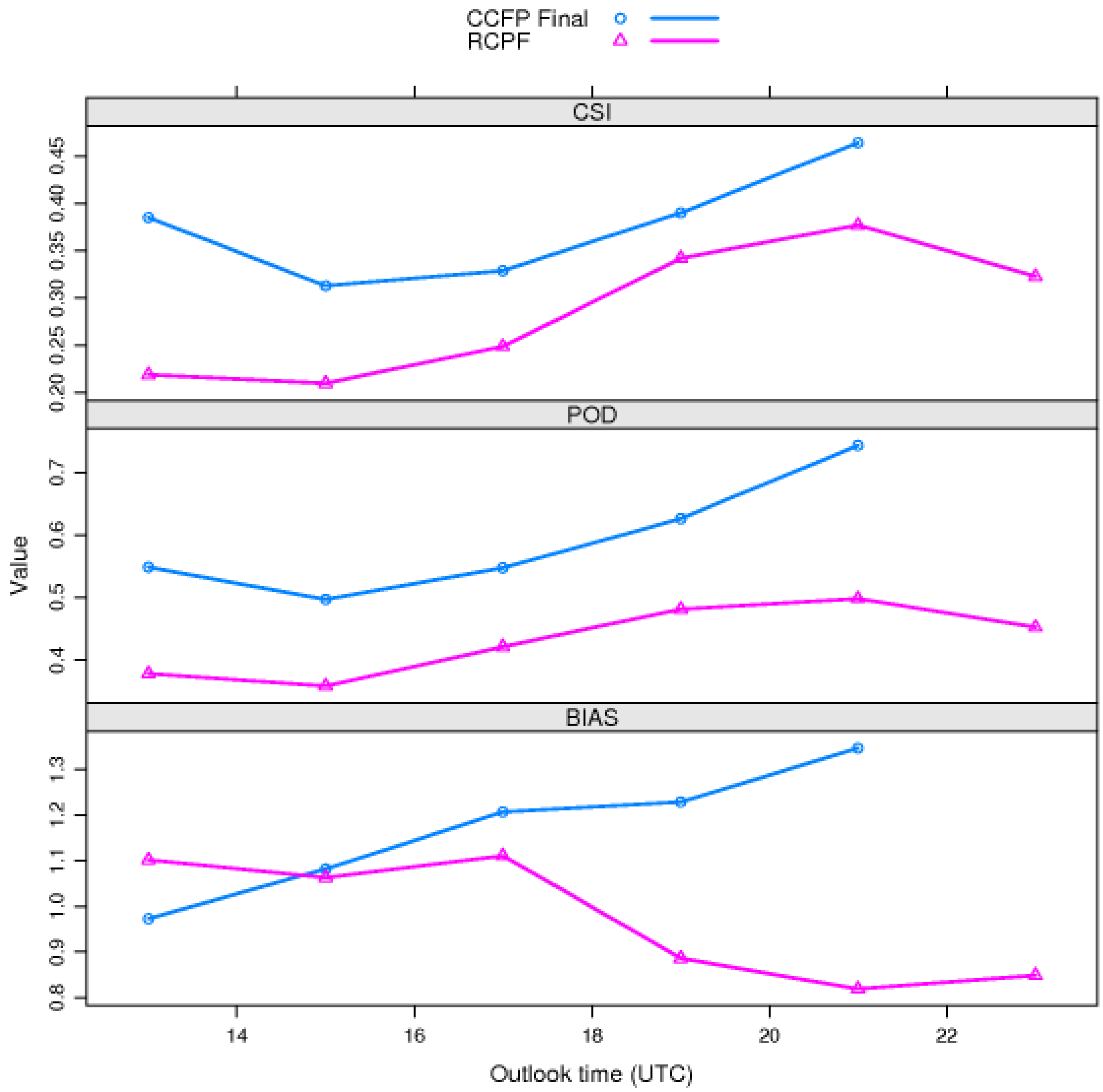


Figure 6. As in Figure 4 except for the sector-based approach.

4.3. Forecast Performance as a Function of Air Traffic Impact and Convective Coverage

The forecast quality of CCFP Final and RCPF was also analyzed with respect to NAS weather impact. This analysis was linked to operational impact via the introduction of two additional variables: the amount of convection over CONUS on a given day and the amount of traffic impact that resulted from the convection. These variables are used to define regimes of varying operational relevance to strategic traffic flow planning (Fig. 7).

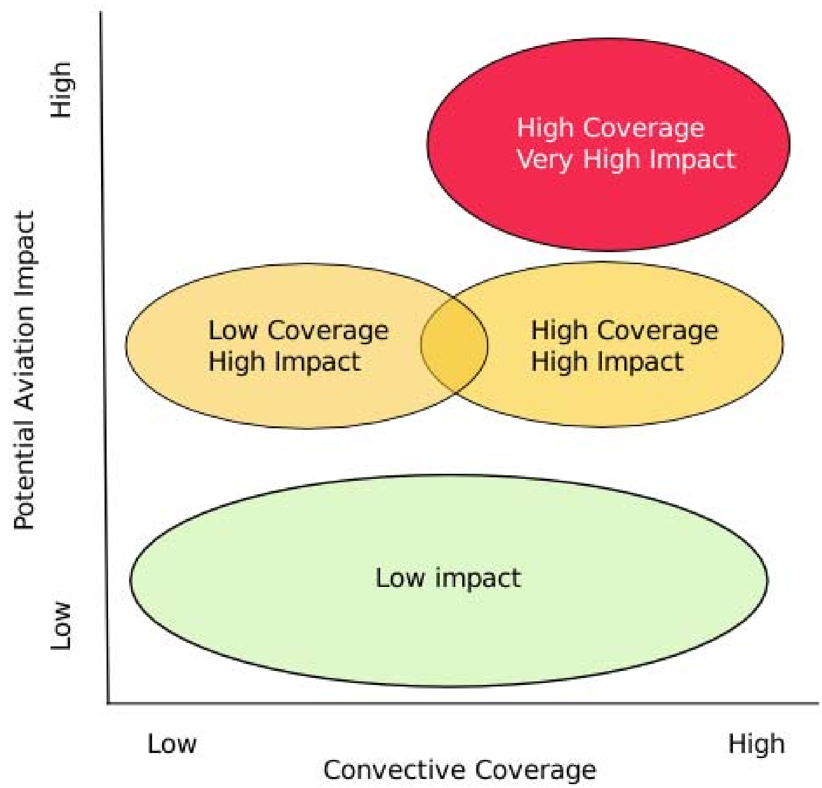


Figure 7. Schematic depiction of regimes that are of differing importance to aviation planning as a function of location of convection (as measured by potential aviation impact) and overall convective coverage over CONUS.

For each day in the study period, the average 2-h, 4-h, and 6-h outlook period CSI values for the three telecon times is computed for the RCPF and CCFP Final forecasts. The convective coverage is the normalized maximum hourly convective coverage over CONUS for that day, while the potential aviation impact is a measure similar to the Weather Impacted Traffic Index (WITI; Callaham et al. 2001). Values for the coverage and impact variables were normalized by the maximum values achieved for each variable, respectively, during the study period. The results, depicted as the difference between the two daily CSI values as defined by RCPF-CCFP Final, are shown in Fig. 8. Perhaps the most prominent result is that there is no systematic behavior observable within each outlook period. Neither forecast systematically outperforms the other for any of the regimes described in Fig. 7. This was confirmed by subjectively analyzing a large number of forecasts representing a wide variety of convective coverage/potential aviation impact scenarios. The general trend in CSI differences across outlook periods mirrors the aggregate results presented in Fig. 3. The largest difference between CCFP Final and RCPF CSI values occurred at the 2-h outlook period. At the 6-h outlook period the RCPF and CCFP Final CSI values were very similar. On high-coverage, high-impact days, the CSI (shown in Fig. 8) indicates that neither forecast system performed significantly different from the other.

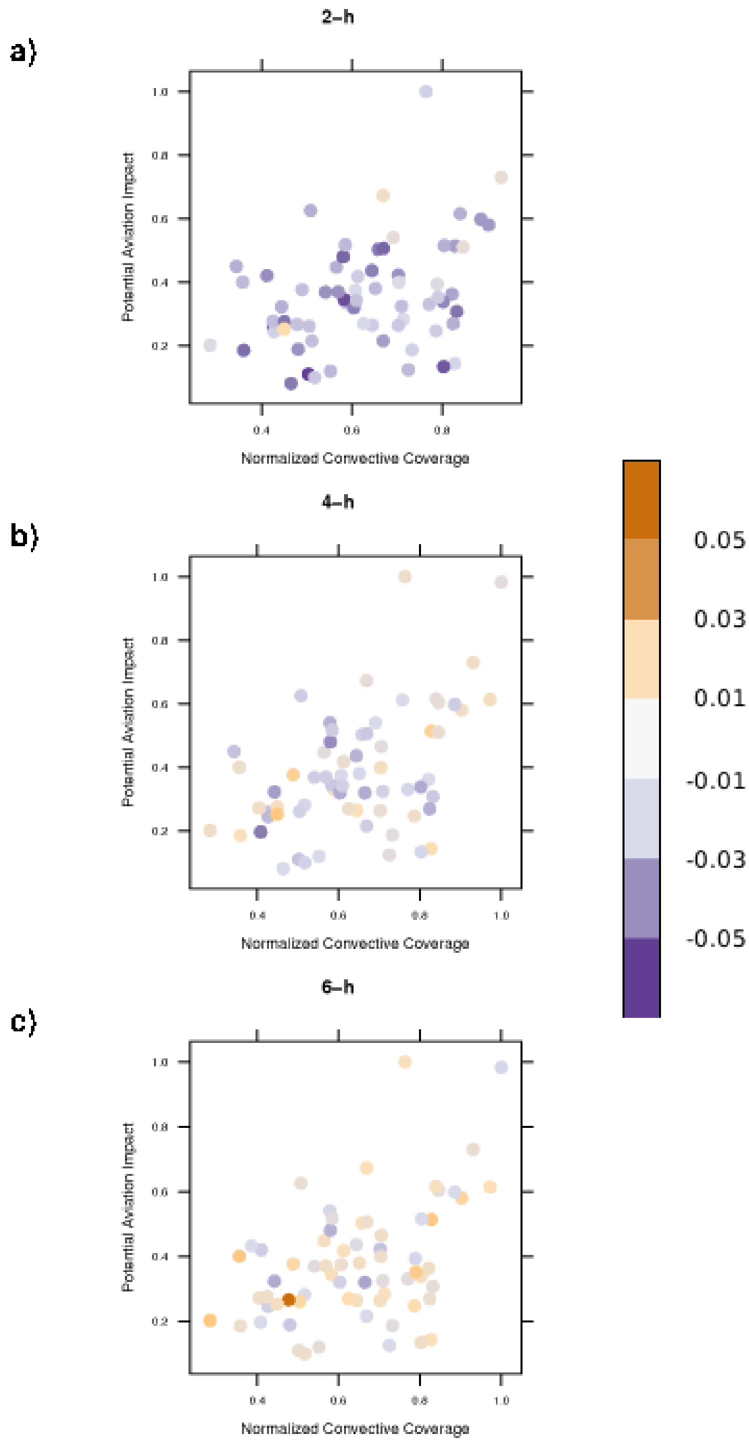


Figure 8. Difference between average CSI values (RCPF – CCFP Final) for a) 2-h, b) 4-h and c) 6-h outlook period forecasts for the telecon-constrained traditional verification approach for each day in the study period as a function of the normalized convective coverage and potential aviation impact. Cool colors show where RCPF values are less than CCFP Final; warm colors represent situations where RCPF values exceed CCFP Final values.

The ten days with the largest aviation impact, according to the potential aviation impact data, were analyzed in detail. On these days, the long-range 6-h and 8-h outlook periods are particularly important. However, with the absence of an 8-h CCFP forecast, the 6-h forecasts were chosen for further investigation. Detailed depictions of the individual forecast statistics are presented in Fig. 9. The depictions show the difference between the performance on high-impact days and the general results discussed at the beginning of this section and shown graphically in Appendix 3. Compared to the overall values, both forecasts perform better in the afternoon and worse in the mornings. CCFP Final in particular has lower CSI values in the mornings. The lower CCFP Final CSI values are related to the lower POD values and higher BIAS values and may be indicative of misplaced forecast areas. The improved RCPF values at the 8-h outlook periods for the afternoon times is encouraging for planning purposes on high impact days.

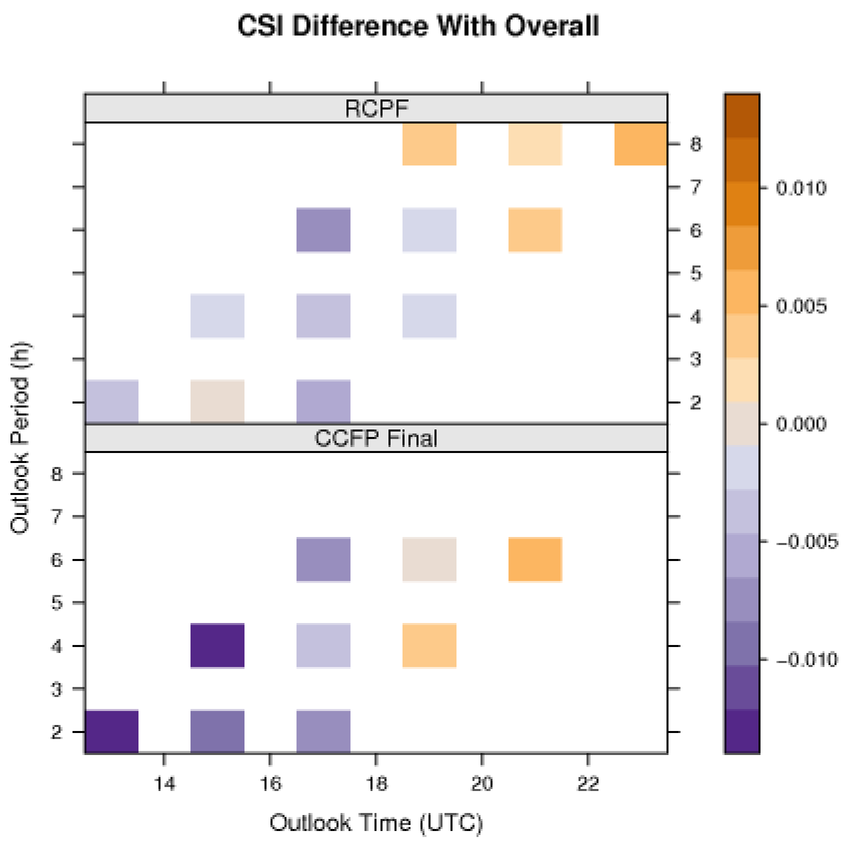


Figure 9a. See complete caption on Fig. 9c.

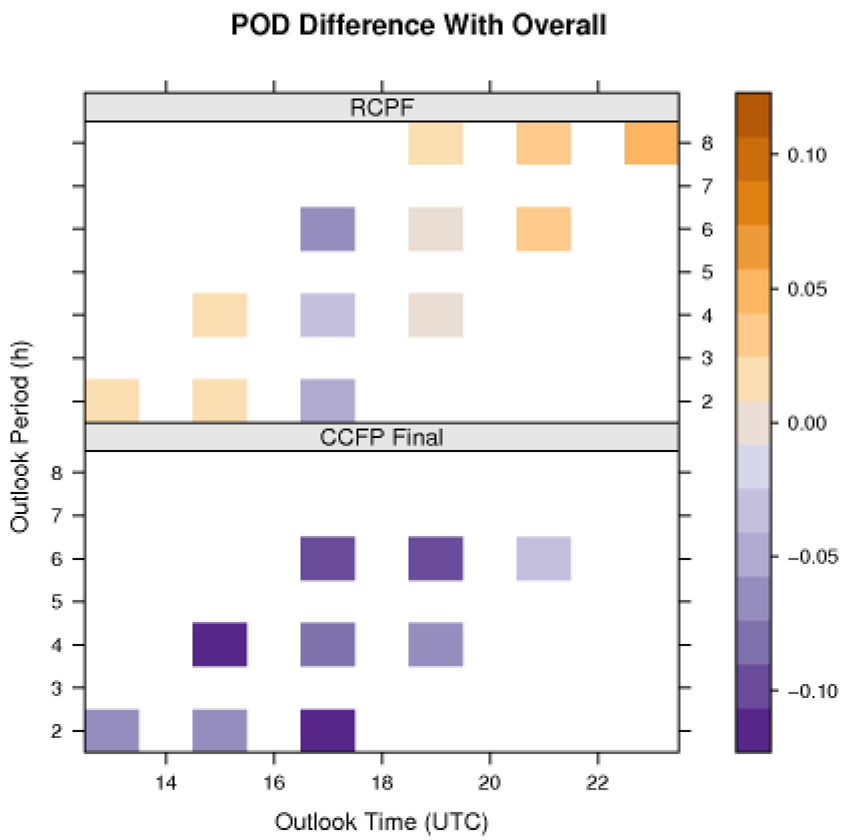


Figure 9b. See complete caption on Fig. 9c.

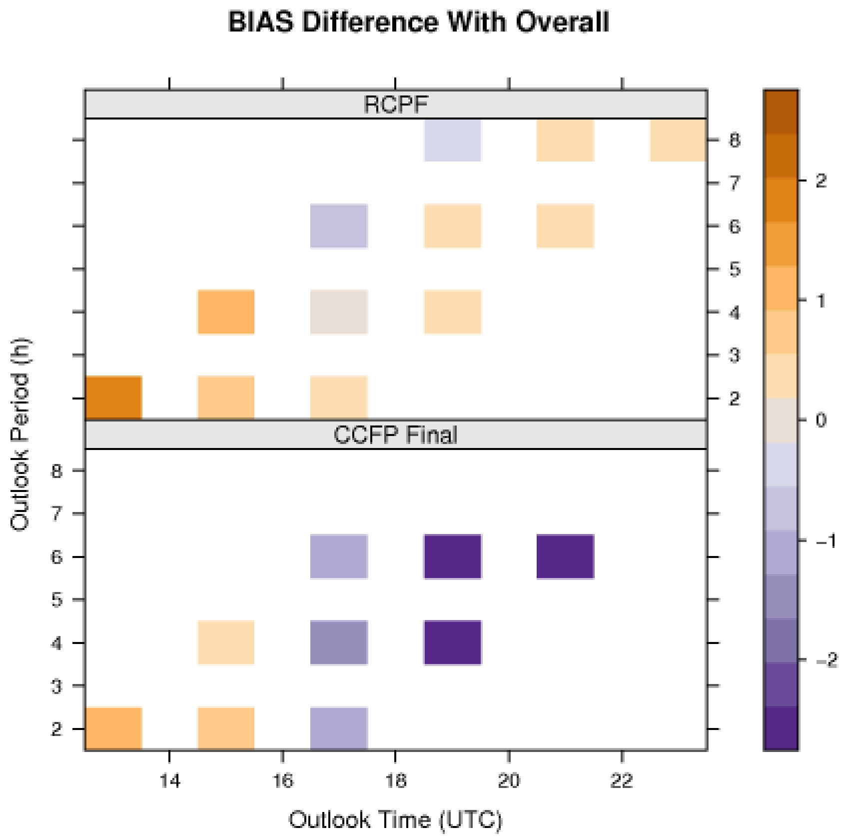


Figure 9c. Level plots of a) CSI differences, b) POD differences, and c) BIAS differences for the 6-h outlook period forecasts in the high-impact day subset compared to the overall 6-h forecast values for each forecast product. Warm colors indicate where the statistic is larger for each product for the high-impact subset compared to all days in the study period for the same product.

4.3.1. Probabilistic Comparison

The RCPF and CCFP Final forecasts were also assessed probabilistically. The RCPF represents a probabilistic forecast, while the CCFP Final was transformed into a probabilistic forecast by assigning the expected areal coverage for each polygon to each grid box contained within the CCFP polygon. The probability bins used for this comparison for CCFP Final are 25-49%, 50-74%, and 75-100%. The CCFP bins were derived from the forecast coverage attribute that is a part of every CCFP forecast polygon. A 10% probability bin width was used for RCPF. A key aspect of a good probabilistic forecast is that the forecast is reliable. Reliability refers to the agreement between the forecast probabilities and the observed relative frequencies for each of those probabilities. The probabilistic performance of CCFP Final and RCPF is shown on a reliability diagram in Fig.10. Both forecasts are poorly calibrated. For example, the largest observed relative frequency is associated with the RCPF 90-100% probabilities and only represented an observed relative frequency, or verification rate, of 10%. The poor calibration for both forecasts can be linked to the large biases seen in the dichotomous analysis.

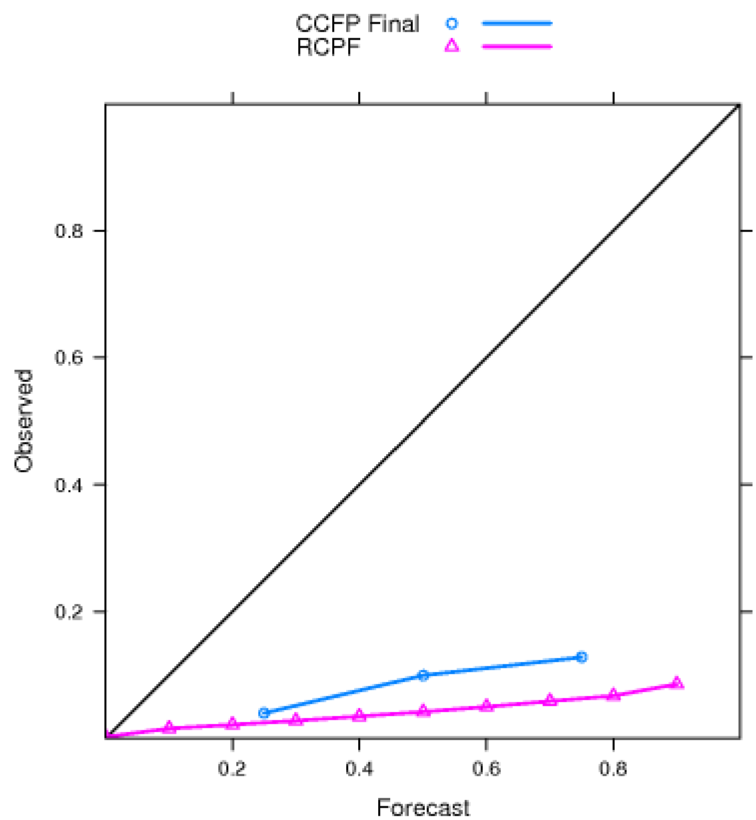


Figure 10. Reliability diagram for the CCFP Final and RCPF forecasts. Points are plotted at the beginning of each probability range.

4.4. NAM and RUC Simulated Reflectivity

Recent efforts have been undertaken to provide simulated radar reflectivity fields from numerical models. These fields have great appeal because of their intuitive mapping to radar observations that are typically used to depict convective hazards. The simulated reflectivity fields have the advantage of being instantaneous measures at a particular time as opposed to other measures of convection such as accumulated convective precipitation which requires an aggregation period (typically one to three hours) and therefore gives a more 'smeared' view of convective location. In this section the simulated reflectivity output from the operational NAM model is compared to simulated reflectivity from an experimental version of the RUC model.

The goal of this analysis was to study the ability of the two algorithms to correctly predict hazardous convection (defined as VIP level 3; or 40 dBZ) owing to its importance in disrupting air traffic operations at such intensities. The reflectivity forecast verification results, presented in the context of the other forecasts studied, are shown in Fig. 11. The products show no skill at predicting hazardous convection, as measured by the verification techniques used for this study. In particular, the BIAS values near 0 indicate that both models typically do not ever forecast convective intensities greater than 40 dBZ. The model reflectivity values represent a grid-box average at their native resolutions (NAM – 12 km, RUC – 13 km). For this analysis, no attempt was made to create a smoothed observation field, which may have provided correspondence between the forecast and the observations, since the location of the hazardous convective must be preserved.

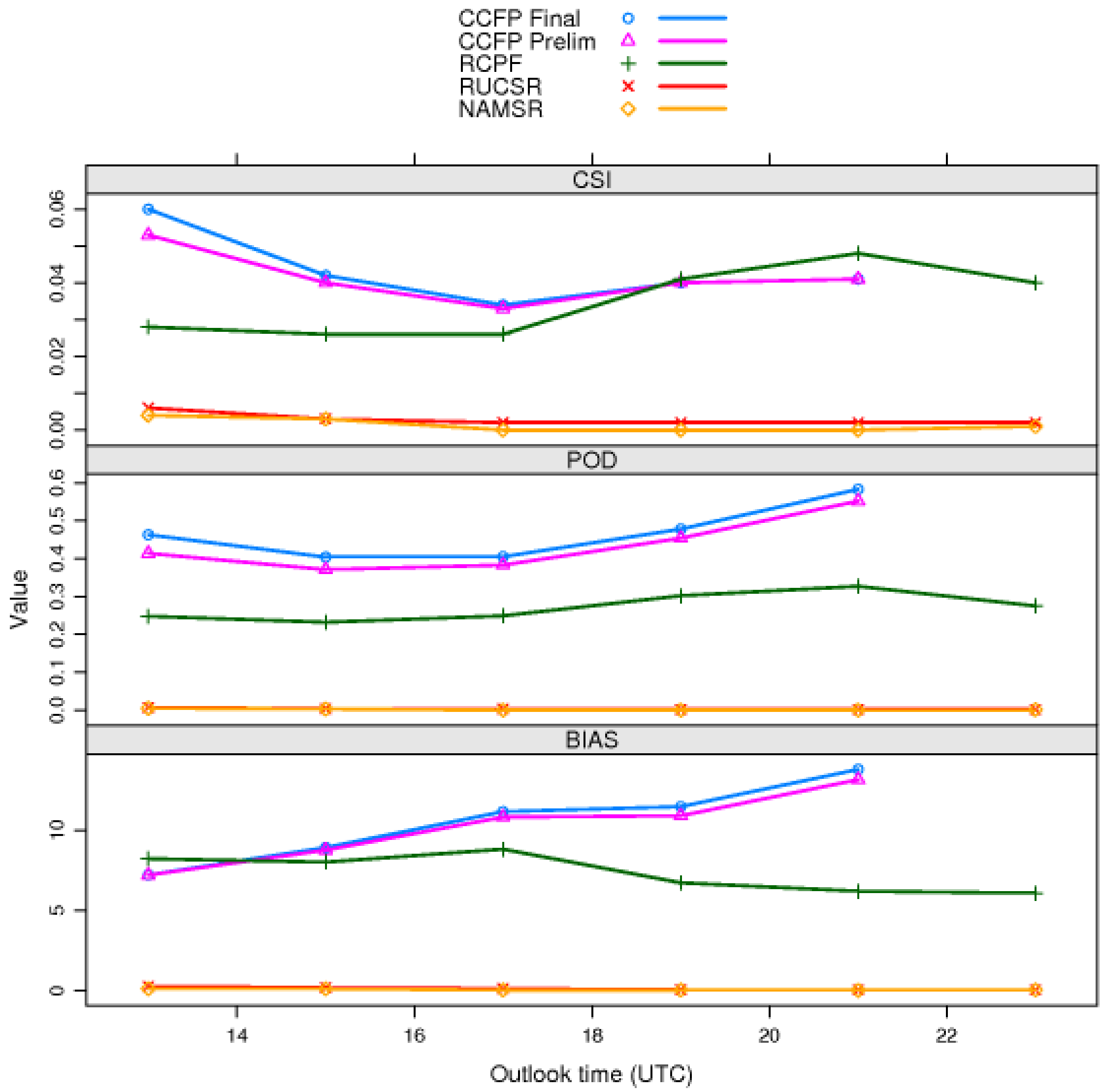


Figure 11. CSI, POD, and BIAS values for all forecasts in the telecon-constrained portion of the day. Note the RUCSR and NAMSR values which barely exceed zero at all outlook times.

In order to gain additional insight into the performance of the algorithms, in particular the limited amount of significant convection captured by the forecasts, an analysis of the joint distribution of forecasts and observations was created. The conditional probabilities of a given forecast intensity were derived for each observed intensity. The data were derived from all available telecon-constrained simulated reflectivity forecasts and the corresponding observation grids. If the models performed perfectly, 100% of the VIP 3 observed grid boxes would be associated with forecasts of VIP level 3. Similar behavior is expected for all other VIP levels. If the probability densities deviate from perfect agreement, this would indicate issues such as a systematic bias in the forecast. Figure 12 shows the conditional probability of forecast VIP levels given observed VIP levels for both the NAMSR and RUCSR forecast products. The most striking result is that the forecast VIP levels are typically 0 wherever there is any observed convection (VIP level 1 or greater). For both models, some structure is seen for VIP levels 1 and 2, with the NAM in particular showing slightly higher forecast VIP levels. These results might indicate that the problems are not merely a result of a miscalibration of the forecasts. The radar products appear to quite often have convection in the wrong location, regardless of intensity.

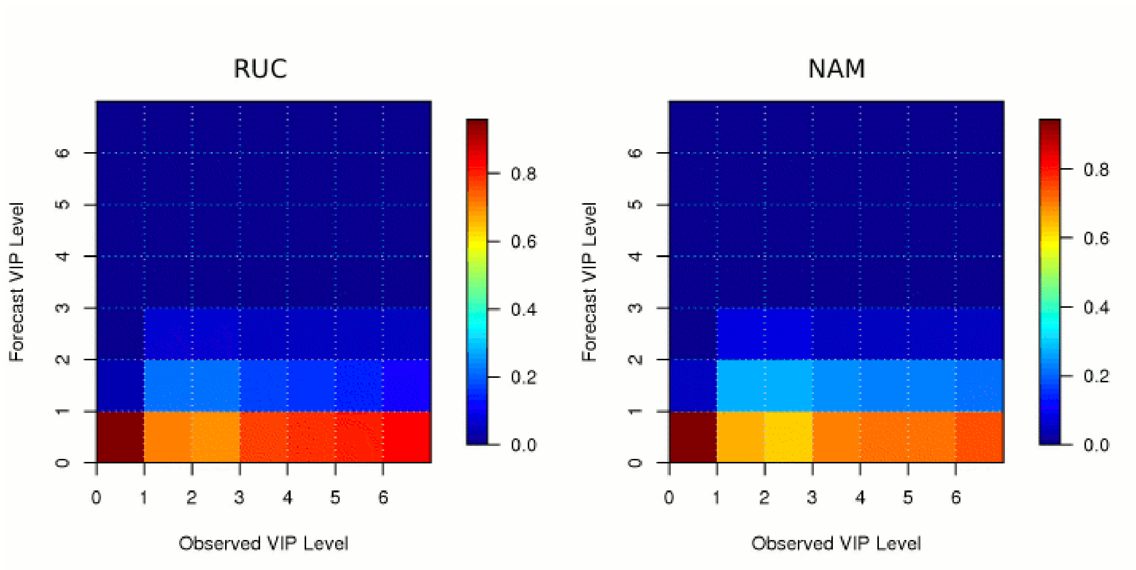


Figure 12. Conditional probability of forecasting a particular VIP level given an observation of that VIP level for the RUCSR and NAMSR products. Ideally all values would lie along the diagonals from the lower left to the upper right of the diagrams.

5. Conclusions

This study evaluated the quality of five convective forecasts with respect to operational strategic air traffic flow planning. The results indicated that:

- Nearly identical forecast performance from the CCFP Preliminary and the CCFP Final was determined.
- The CCFP and RCPF performed similarly for nearly all time periods, except at the 2-h outlook period where CCFP performed slightly better.
- At early outlook times and for shorter outlook periods, CCFP performed slightly better than RCPF.
- At later outlook times and longer output periods (i.e., when convective weather has the potential to severely impact air traffic), the RCPF reformed as well as the

CCFP.

- Analysis of the top ten high impact air traffic days indicated that the performance of the RCPF at the 8-h outlook period for the afternoon shows some promise for planning purposes.
- On high-coverage, high-impact days neither CCFP nor RCPF performed significantly different from the other.
- The CCFP, for every valid time of interest, better identified the sectors that were impacted by convection than did the RCPF.
- The reflectivity products (NAM and RUC) showed virtually no skill at forecasting hazardous convection, but did provide some guidance at long lead periods for areas of hazardous concern.
- The convective probability aspects of CCFP and the RCPF indicated low reliability.

Acknowledgments

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

References

Callaham, M., DeArmon, J., Cooper, A., Goodfriend, J., Moch-Mooney, D., and Solomos, G., 2001: Assessing NAS Performance: Normalizing for the effects of weather. 4th USA/Europe Air Traffic Management R&D Symposium, Santa Fe, NM.

Phaneuf, M., and D. Simenauer, 2007: 2007 Convective forecast subjective evaluation. Available from M. Phaneuf at mphaneuf@avmet.com.

Weygandt, S.S., and S.G. Benjamin, 2004: RUC model-based convective probability forecasts. *Preprints 11th Conf. Aviation, Range, and Aerospace. Meteor.*, Hyannis, MA, Amer. Met. Soc.

Appendix 1. Teleconference Data Tables

Tables indicate the forecast data available for the three strategic teleconferences. In the following tables all absolute times are UTC and all relative times are in hours.

1115 UTC

Forecast	Initial Time	Issue/ Available Time	Valid Time (Lead Period)			
			1300	1500	1700	1900
CCFP Prelim	1000	1000	1300 (3)	1500 (5)	1700 (7)	
CCFP Final	1100	1100	1300 (2)	1500 (4)	1700 (6)	
RCPF	0900	1100	1300 (2)	1500 (4)	1700 (6)	1900 (8)
RUCSR	0900	1100	1300 (2)	1500 (4)	1700 (6)	1900 (8)
NAMSR	0600	0800	1200 (4)	1500 (7)	1800 (10)	1800 (10)

1315 UTC

Forecast	Initial Time	Issue/ Available Time	Valid Time (Lead Period)			
			1500	1700	1900	2100
CCFP Prelim	1200	1200	1500 (3)	1700 (5)	1900 (7)	
CCFP Final	1300	1300	1500 (2)	1700 (4)	1900 (6)	
RCPF	1100	1300	1500 (2)	1700 (4)	1900 (6)	2100 (8)
RUCSR	1100	1300	1500 (2)	1700 (4)	1900 (6)	2100 (8)
NAMSR	0600	0800	1500 (7)	1800 (10)	1800 (10)	2100 (13)

1515 UTC

Forecast	Initial Time	Issue/ Available Time	Valid Time (Lead Period)			
			1700	1900	2100	2300
CCFP Prelim	1200	1200	1700 (3)	1900 (5)	2100 (7)	
CCFP Final	1300	1300	1700 (2)	1700 (4)	1900 (6)	
RCPF	1100	1300	1700 (2)	1700 (4)	1900 (6)	2100 (8)
RUCSR	1100	1300	1700 (2)	1700 (4)	1900 (6)	2100 (8)
NAMSR	0600	0800	1800 (4)	1800 (4)	2100 (7)	0000 (10)

Appendix 2. Summary Measures vs Valid Time for All Data Sources

Time series of CSI, POD and BIAS measures for the traditional approach stratified by valid time.

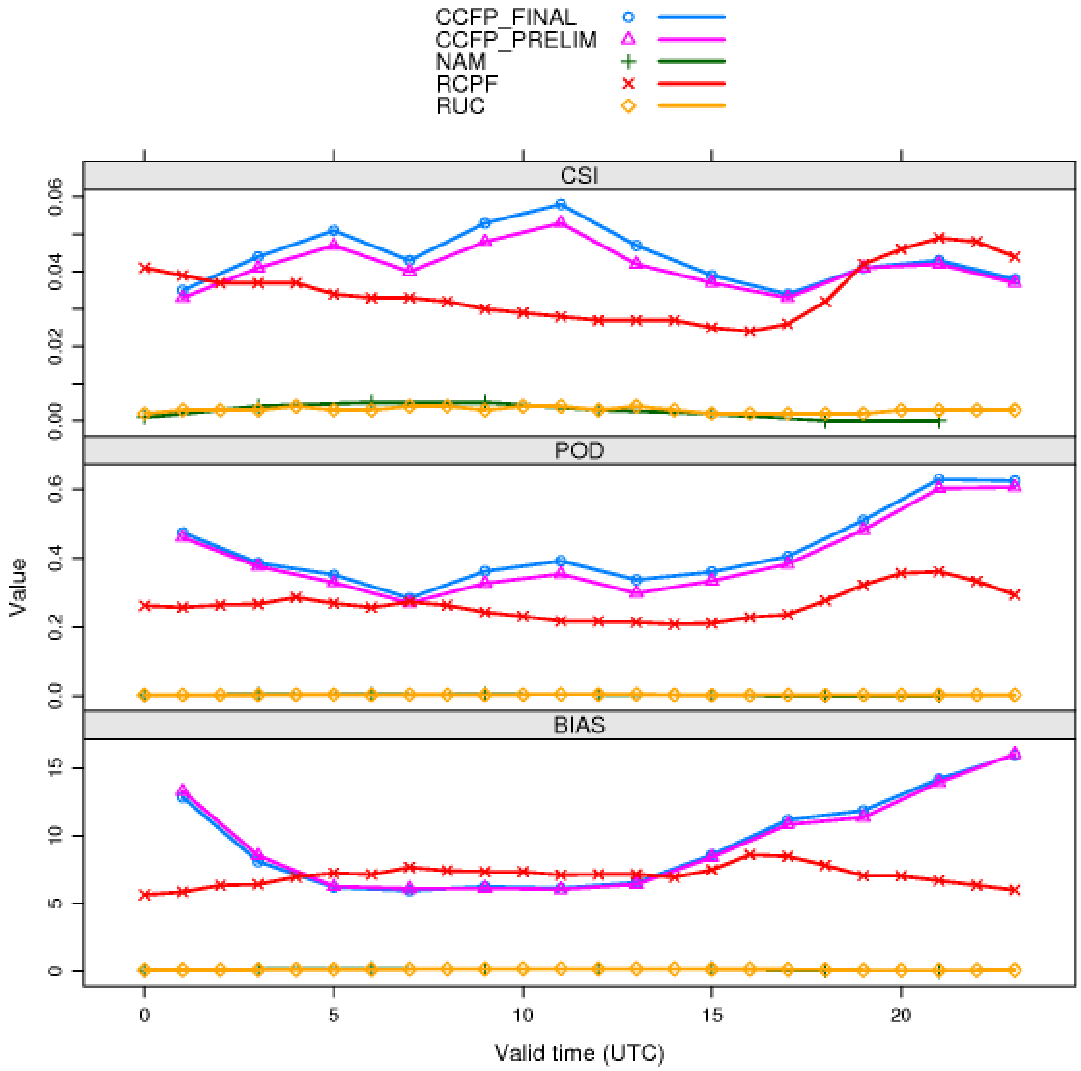


Figure A2-1. Time series of CSI, POD and BIAS values for forecasts aggregated by valid time.

Appendix 3. Diagnostic Plots for Traditional Approach

The following plots augment the information presented in Section 4.2.1.

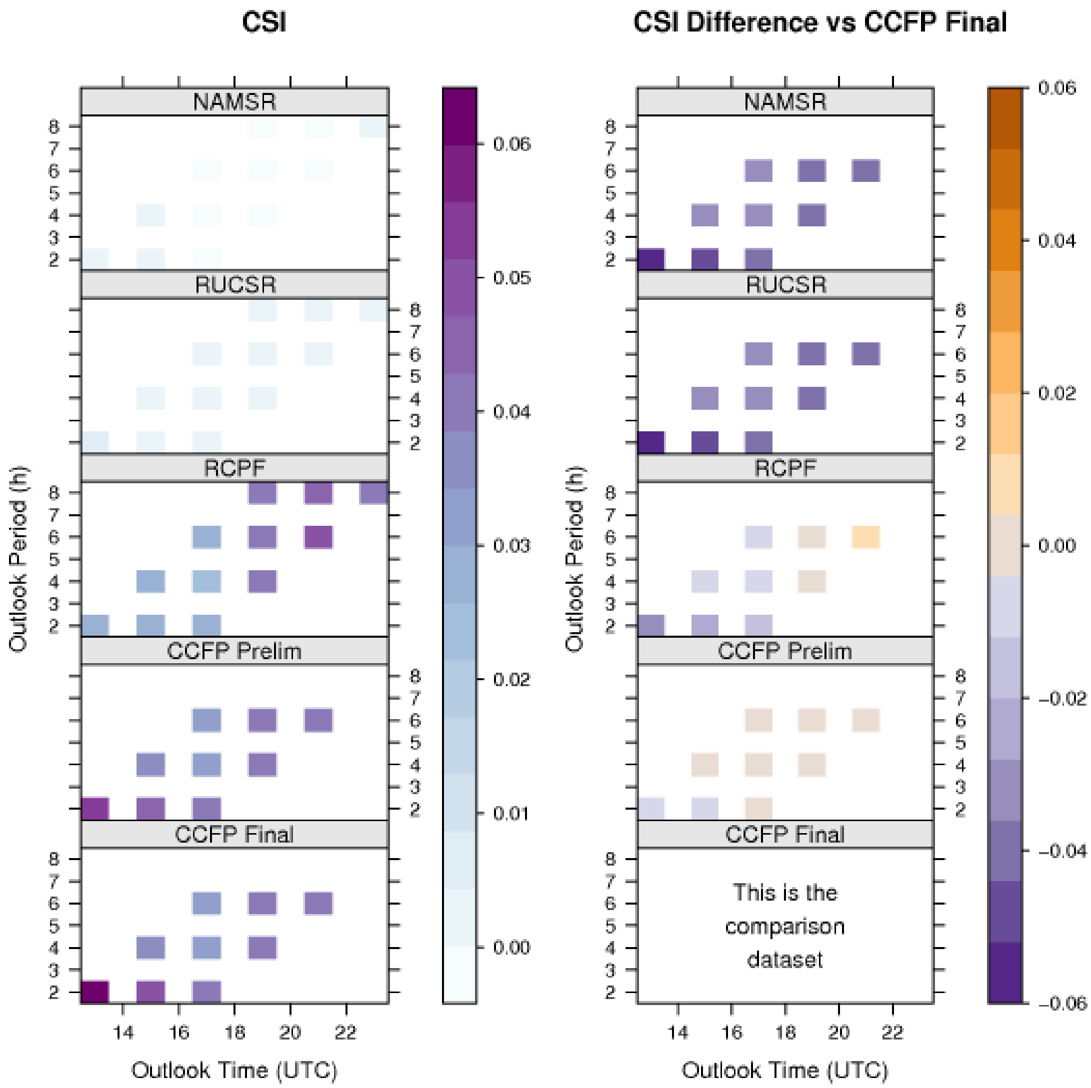


Figure A3-1. Level plots of CSI as a function of outlook period and outlook time (left) and CSI difference (defined as CSI of each forecast minus the CSI for CCFP Final) as a function of outlook period and outlook time (right) for the traditional approach.

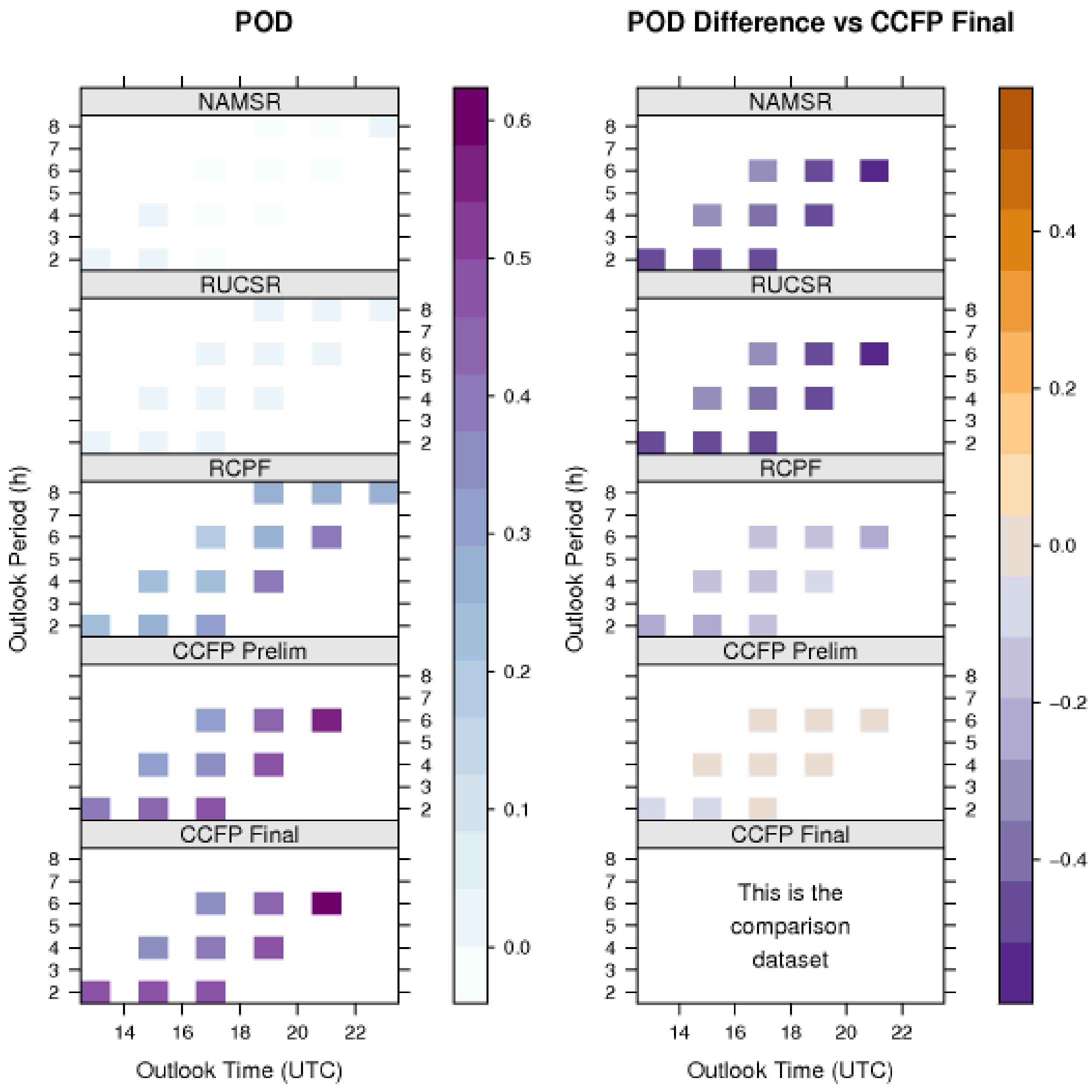


Figure A3-2. As in Fig. A3-1 except for POD.

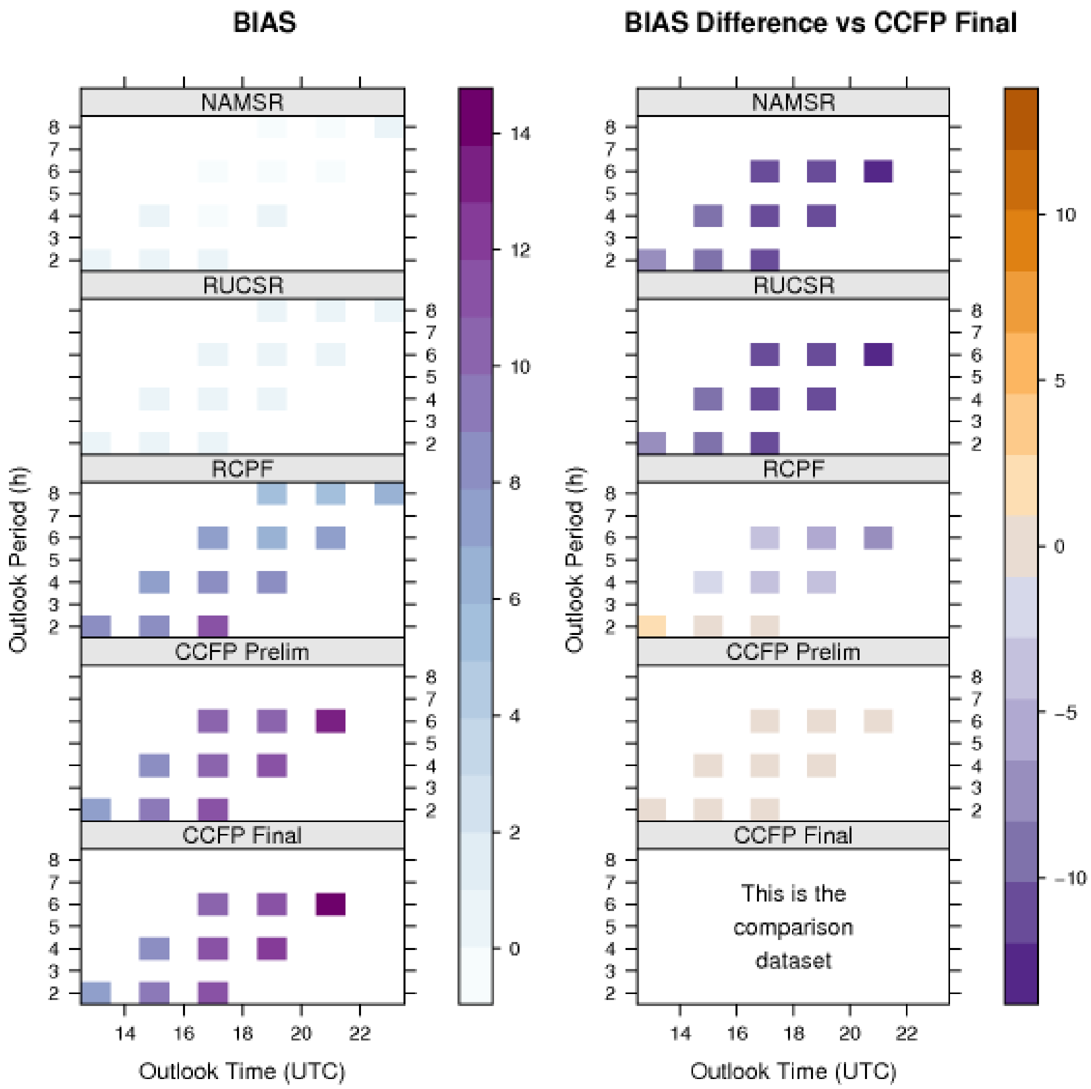


Figure A3-3. As in Fig. A3-1 except for BIAS.

Appendix 4. Diagnostic Plots for Sector Approach

As in Appendix 3 except for the sector approach.

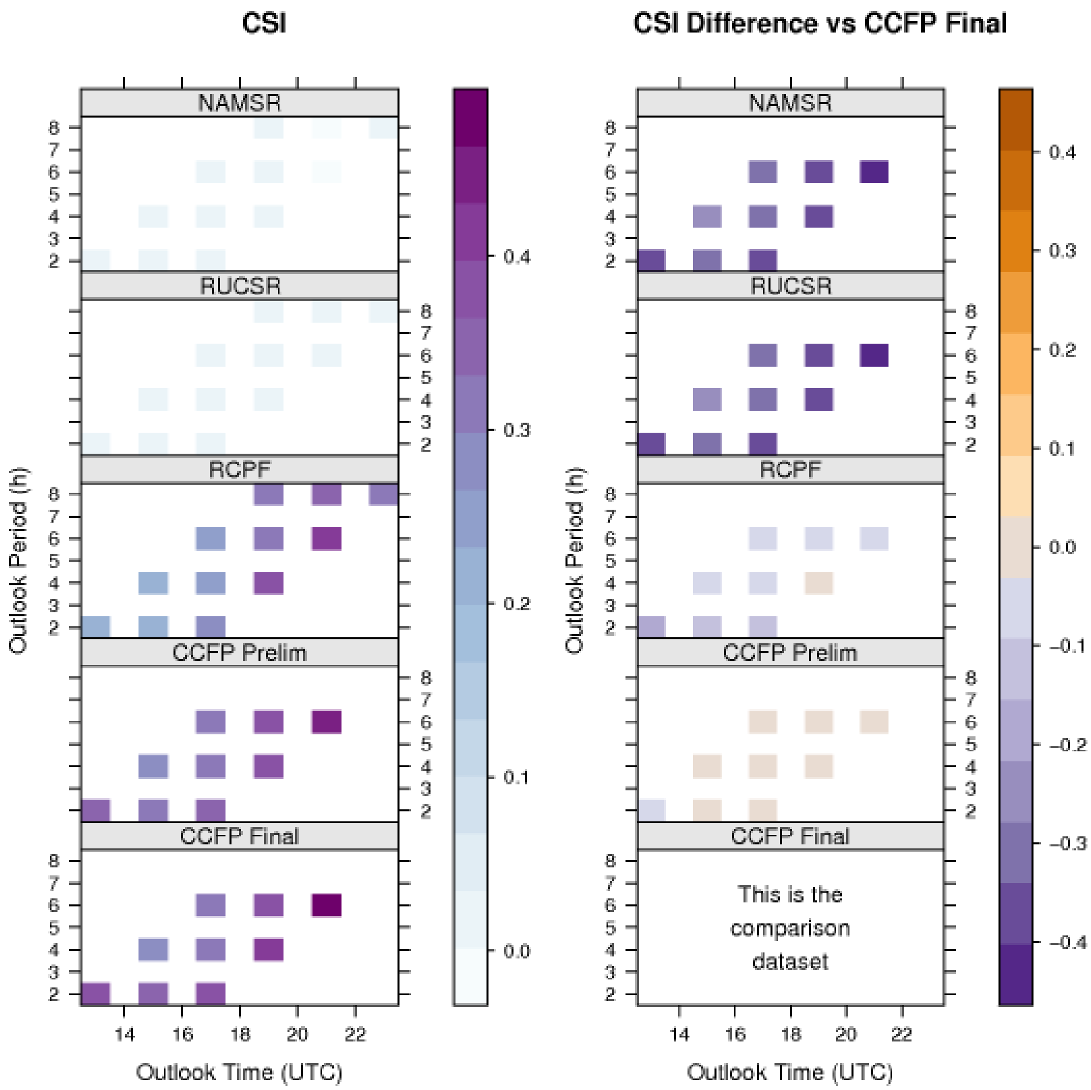


Figure A4-1. Level plots of CSI as a function of outlook period and outlook time (left) and CSI difference (defined as CSI of each forecast minus the CSI for CCFP Final) as a function of outlook period and outlook time (right) for the sector approach.

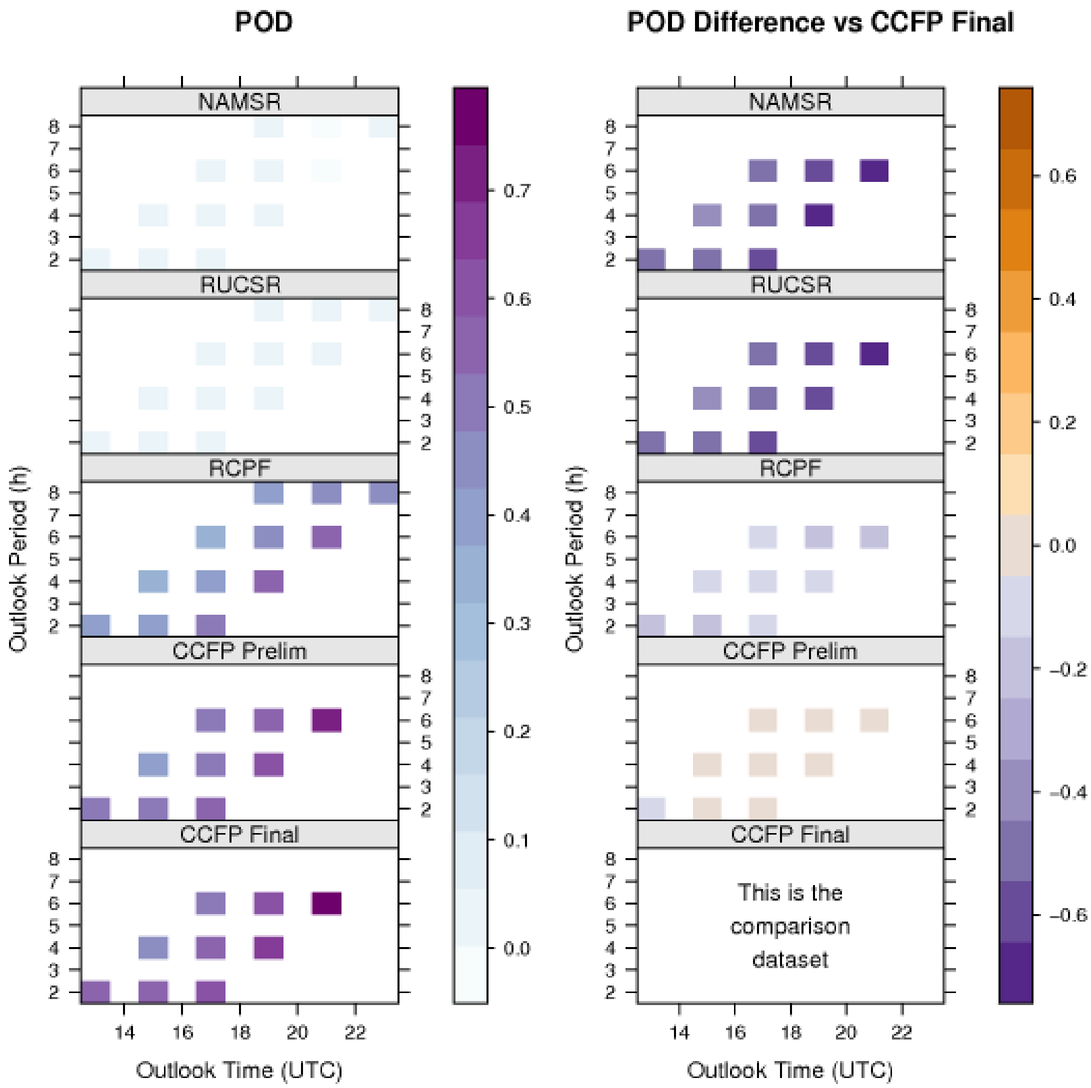


Figure A4-2. As in Fig. A4-1 except for POD.

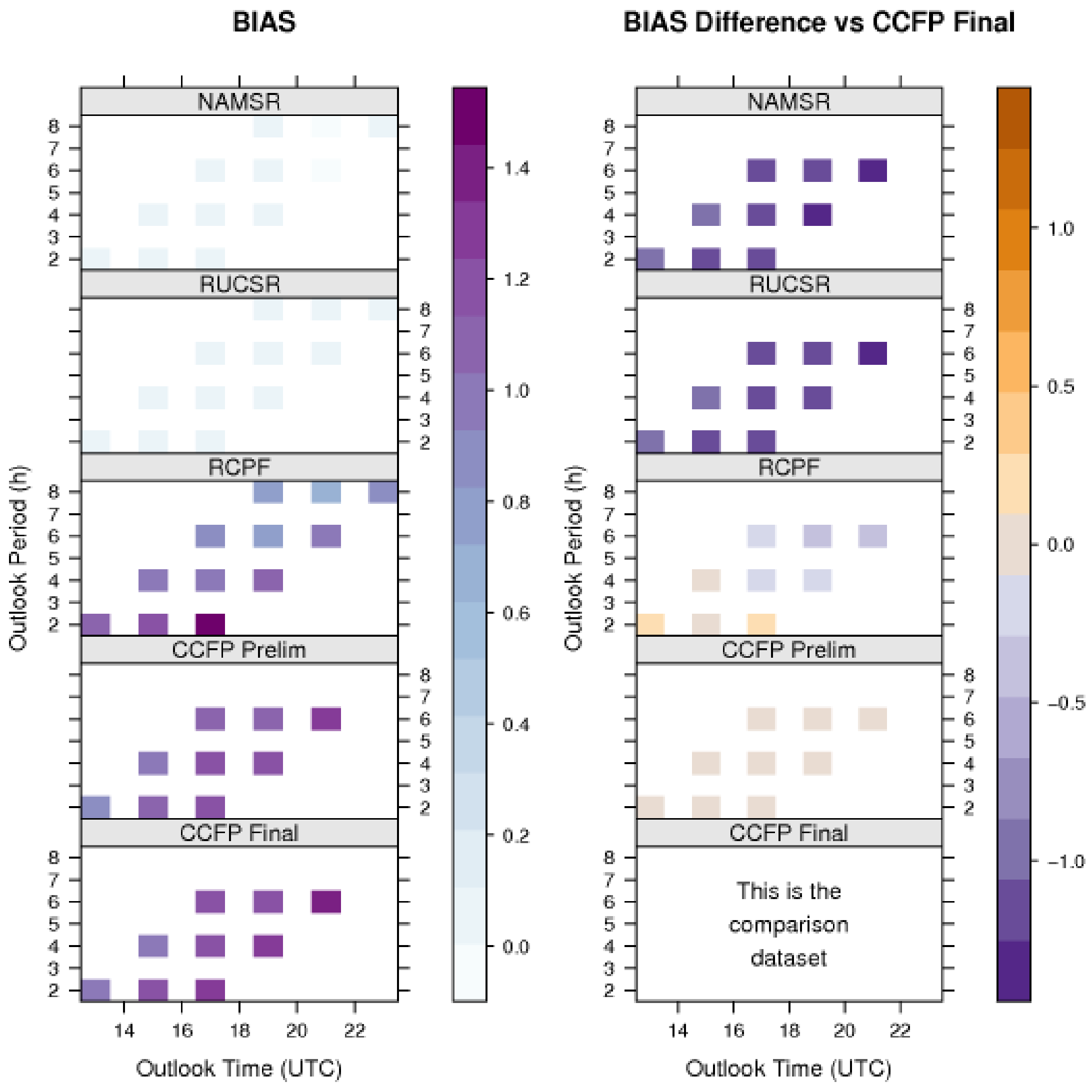


Figure A4-3. As in Fig. A4-1 except for BIAS.