11.3 ALIGNING FORECAST VERIFICATION WITH USER-SPECIFIC NEEDS - AN EXAMPLE FOR AVIATION

Michael P. Kay*[1,2]
Sean Madine[3,2]
Jennifer Luppens Mahoney[2]
Joan E. Hart[1,2]
[1]Cooperative Institute for Research in the Environmental Sciences (CIRES), Univ. of Colorado, Boulder, CO
[2]NOAA Earth System Research Laboratory (ESRL), Boulder, CO
[3]Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State Univ., Fort Collins, CO

1. INTRODUCTION

There are many ways to view forecast verification. In this paper, we consider a decomposition in terms of two groups: producers of forecasts and consumers of forecast information. Forecast verification becomes a much more powerful tool to effect change and influence decisions when viewed from the consumer's, or user's, point of view.

This paper attempts to define the role of consumer-oriented verification, and its relevance to the increased information needs of NextGen, using several examples from a recent verification effort performed for a set of convective forecasts used as input to a planning process that controls the large-scale configuration of, and movement of air traffic through, the National Airspace System (NAS) of the United States. Section 2 describes the philosophy behind consumer-oriented verification. Several examples of user-specific verification information are presented in Section 3 and a summary is presented in Section 4.

2. MOTIVATION

Forecast verification serves the role of identifying the accuracy of forecasts, with the goal of improving future predictions. Historically, verification of meteorological forecasts is performed using the the strict definition of the forecasts with little concern for how the forecasts are actually being used. Such an approach is perfectly suited to providing feedback to the forecast producers but is not necessarily in line with the information needs of forecast consumers.

In contrast to the producer-centric view, consumer-oriented, or user-specific, verification focuses on a particular user, or group of users, whose needs and use of forecast information are similar. Consumers of weather forecasts, in contrast to the producers, are often much more difficult to identify for all situations. Further, each possible consumer will have different information needs from the forecasts. It is infeasible for all particular consumers to have tailored forecasts produced exactly for all of their needs. Users often do not take forecasts at face value and modify or interpret the forecasts in ways that suit their needs. This is not to say that the consumers use the forecasts in incorrect ways or violate the spirit of the forecasts, however surely some do such things. In order for these groups to gain the most utility from the forecasts it is important to try to verify the forecasts in a way that represents how the data are actually being used by these groups.

The success of any consumer-oriented verification effort is dependent on several key points: the identification of the consumer, an understanding of how they are using the forecast data, defining the purpose of the verification, and the communication of the verification results to the

**Corresponding author address:* Michael P. Kay, NOAA R/GSD5, 325 Broadway, Boulder, CO 80305; e-mail: mike.kay@noaa.gov.

consumer as it relates to their use of the forecasts. The identification of the particular user (hereafter user and consumer are used interchangeably) that verification is being performed for is critical. Once the user has been identified, the next step is to identify how the forecast(s) that are to be verified are being utilized. This knowledge is then used to define the purpose of the verification exercise. It is important to realize that this is a crucial step in the verification process (Panofsky and Brier, 1958). Without a clear purpose, the potential benefits of the verification will fail to be realized. Further, a lack of focus may allow for confusion on the part of the community as to why certain choices were made, or why one approach was taken instead of another, which will distract attention away from the results and towards questions where there are potentially no right answers. Finally, the verification results need to be communicated, wherever possible, in a way that the user can readily understand and apply them to their particular problem domain.

3. Examples

This section provides several examples of verification information taken from a recent convective forecast intercomparison. Five forecasts were evaluated for use at the Air Traffic Control System Command Center (ATCSCC) for the period 11 June 2007   31 August 2007. The ATCSCC is responsible for providing the large-scale operating environment for all commercial air traffic within the United States and associated coastal waters. The ATCSCC uses convective forecast information to adjust their overall plan for how aircraft operations may be impacted throughout the NAS. Air traffic delays due to convection have increased in recent years and with air traffic volumes projected to increase dramatically in the future, it is important that meteorological predictions of hazardous convection be as accurate as possible. This study looks at the verification of one particular type of forecast employed by the ATCSCC: predictions of significant convection (40 dBZ or VIP level 3) for the afternoon period when convection has the greatest potential to cause substantial disruptions and problems for traffic flow management of the NAS.

The forecasts in the experiment were chosen because they are either currently being used operationally or are being considered for future use. The Final and Preliminary versions of the Collaborative Convective Forecast Product represent the current state-of-the-art human-generated strategic aviation forecasts for convection. The Rapid Update Cycle (RUC) Convective Probability Forecast (RCPF) is an experimental probabilistic convective forecast that is produced on a 20-km grid and represents the probability of convection at each grid point at each valid time (Weygandt and Benjamin, 2004). Two forecasts of simulated composite radar reflectivity were also assessed: an experimental version from the RUC model running at 13 km and an operational product from the North American Mesoscale (NAM) model running at 12 km. All forecasts were verified against National Convective Weather Diagnostic (NCWD) data which has an approximate 4-km resolution.

Complete details of the methodology employed for the experiment and the complete set of results may be found in Kay et al. (2008).

3.1 Telecon times and additional temporal constraints

The ATCSCC is responsible for initiating and coordinating a series of teleconferences (hereafter telecons) every two hours to update and adjust their plan for managing the NAS on a given day. At each of those telecons, decisions are made for points in time 2 h, 4 h, 6 h, and 8 h in the future.

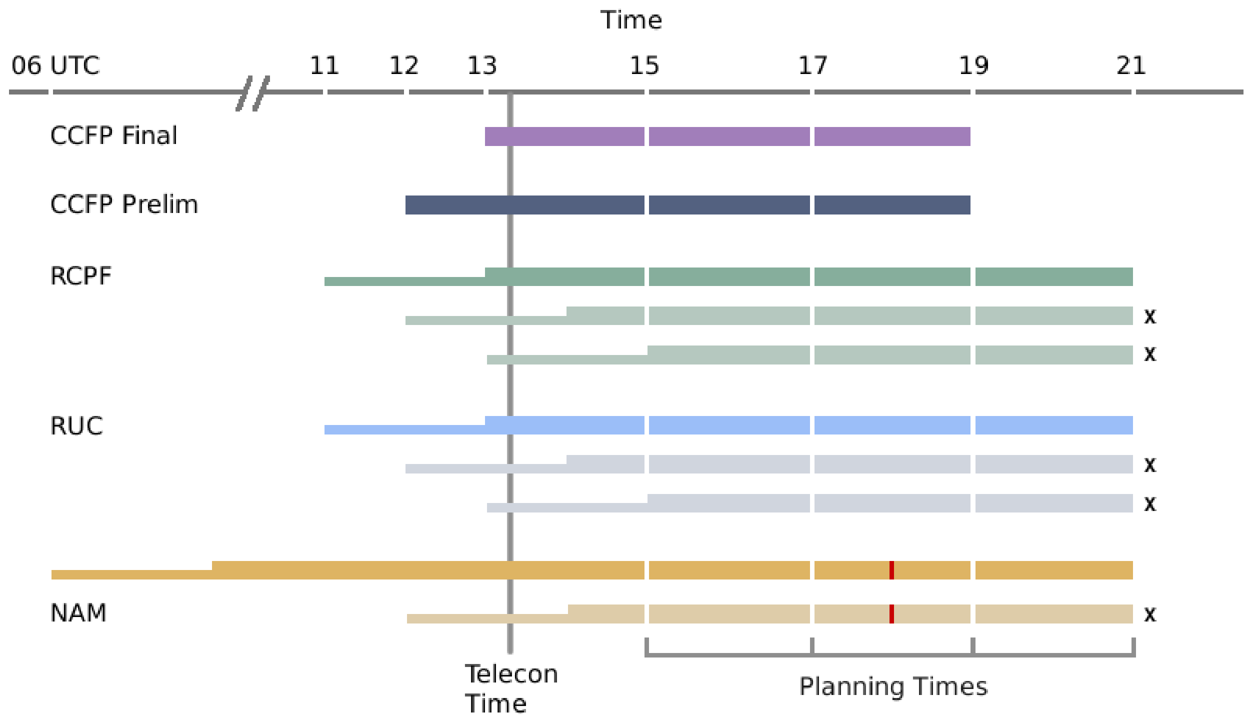

Fig. 1: Diagram highlighting the concept of product availability for the 1315 UTC telecon. Each bar indicates a particular issuance of a forecast. The left-hand side of each bar indicates the issuance time. Thin portions of bars indicate latency due to computation and/or product delivery. X's to the right of bars indicate products that are unavailable for use for the 1315 UTC telecon. Red lines in the NAM bars indicate that there is a valid time of 1800 UTC and that this does not align with the planning times of 1700 and 1900 UTC.

For this study, the 1115, 1315, and 1515 UTC telecons were the primary telecons of interest because they represent the key telecons that initiate the plans for the strategic afternoon period. For the ATCSCC, this period (1300 UTC 2300 UTC) is the part of the day that truly matters.

Because the ATCSCC makes decisions relative to the telecons, it is imperative to adopt a telecon-centric temporal point of view. In this view, only data that are available to the planners *when they need to make decisions* can be compared. More specifically, this means that product latency, defined as the amount of time it takes for a product to be produced and disseminated, must be explicitly accounted for in the verification. For example, a 1200 UTC run of the NAM model takes approximately two hours to integrate on the supercomputer, be post-processed, and be disseminated to users. Thus, the 1200 UTC NAM model data, even for a 1-h forecast valid at 1300 UTC, isn't available operationally until 1400 UTC. In contrast to the computer-based forecasts, the CCFP forecasts, which are human-generated, are typically available within a few minutes of their scheduled issuance times. If one were performing a producer-centric verification of the forecasts, latency would not be of concern, and the forecast initial times, lead times (the amount of time, in hours, in the future that the forecast is being made for), and valid times would all be matched accordingly. The consumer-oriented reference frame highlights that latency cannot be ignored and must be explicitly accounted for in the

Table 1: Telecon data table showing the forecasts that are available for planning purposes by the ATCSCC for the 2-h, 4-h, 6-h, and 8-h outlook times for the 1515 UTC telecon. Additionally, the lead periods are shown in parentheses below each forecast valid time indicating which forecast is being used for that outlook time. If a forecast valid time does not align with an outlook time it is highlighted in a bold blue font.

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	1200	1400	**1800** (4)	**1800** (4)	2100 (7)	**0000** (10)

verification. A term, the lead period, is introduced to represent the lead time of each forecast minus the latency. The latency for all of the computer-generated forecasts was approximately two hours while there was no latency for the CCFP products. These concepts are illustrated diagrammatically in Fig. 1 for the 1315 UTC telecon time. The complications arising from real-time decision-making must be accounted for in the verification process where possible if the results are to be relevant for the targeted consumer.

A secondary aspect of the ATCSCC's needs is the desire to have forecasts verified at particular planning times at fixed points in the future relative to each telecon. The temporal terminology is defined as follows: for each telecon, there are a set of outlook periods 2-h, 4-h, 6-h, and 8-h in the future where decisions are made. The time of day (UTC) that is represented by the telecon time plus the appropriate outlook period is the outlook time. This could equivalently be called an effective valid time as well. In the event that the raw meteorological forecast information is not available for a particular outlook time a mapping must be made from the nearest available valid time for that product to this particular outlook time. The NAM model, which is produced every six hours, and has three-hourly lead times, is the only dataset affected by this issue. Again, this transformation occurs because of the introduction of a known user who uses the available information and makes decisions with them on a schedule that does not align with the schedule of the meteorological data. A verification effort aimed at the producers of the forecasts would not be concerned with such issues. Table 1 depicts the set of available forecasts for the 1515 UTC telecon planning process and highlights the changes necessary to account for product latency and production schedule differences.

Verification results, presented in the form of several dichotomous statistics as a function of outlook time, are shown in Fig. 2. The ATCSCC cares about how much convection is captured by the forecasts (POD), how much overforecasting occurs (BIAS), and desires an overall measure of skill for the forecasts (CSI). The reader should consult a text such as Wilks (1995) for a description of these scores as well as several others. There are a number of interesting results are apparent in Fig. 2. First, the CCFP Preliminary and Final forecasts perform nearly identically. Future work should focus on differences in those forecasts that do not appear in the objective verification. The RCPF performs better in the afternoon than in the morning and outperforms the CCFP Final in the afternoon. The differences in the morning between these forecasts appear to be related to convective initiation. The CCFP forecasters are able to do a better job in the morning, when initiation is much less of an issue and extrapolation on ongoing convection is an accurate forecasting approach. Overforecasting, as indicated by the BIAS panel of Fig. 2 indicates that all forecasts, except for the simulated radar reflectivity products, significantly overforecast the amount of hazardous convection. Additional analysis in Kay et al. (2008) suggests that the simulated radar reflectivity products contain significant intensity and location errors and therefore should be used with caution.

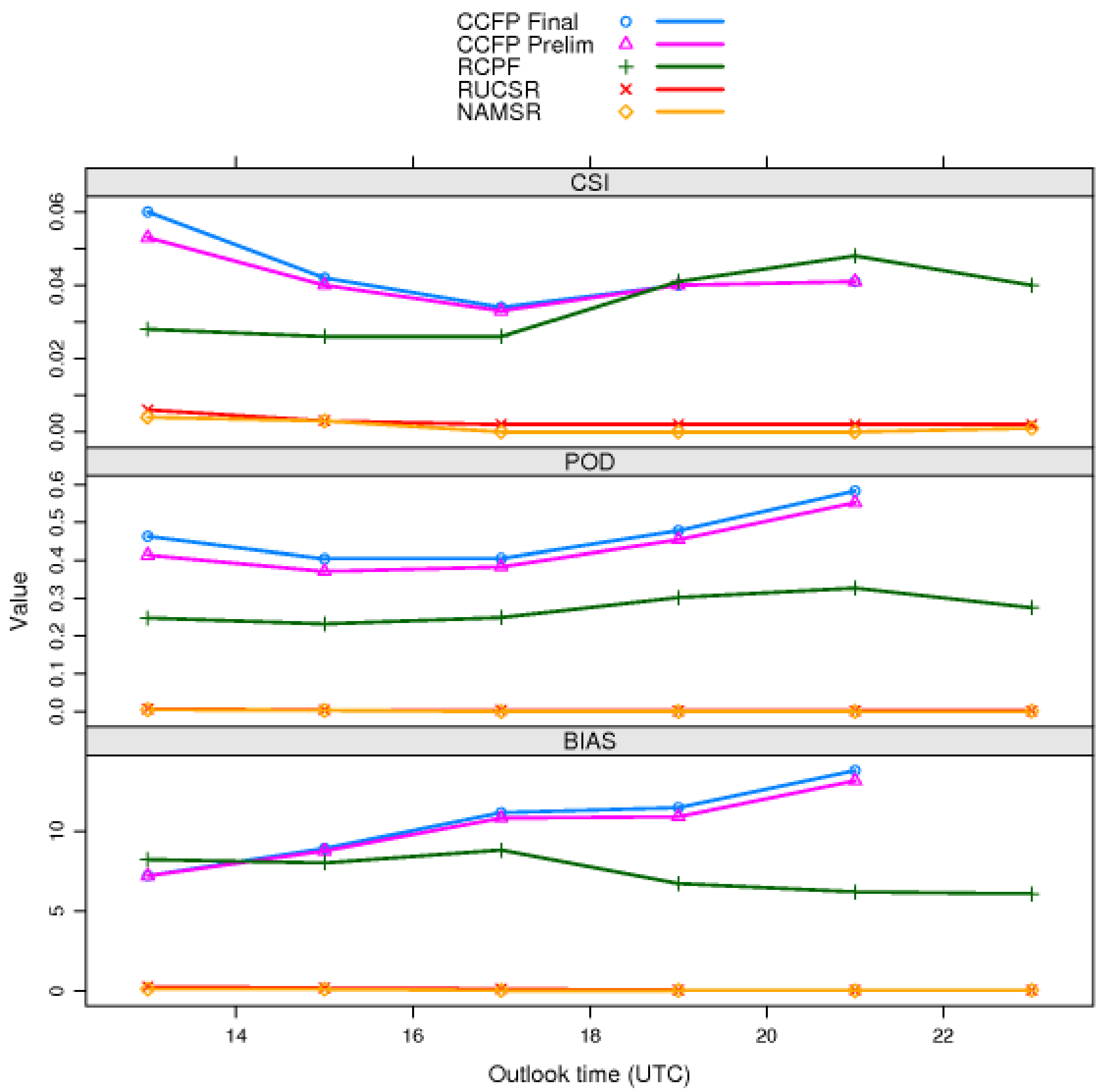

Fig. 2. CSI, POD, and BIAS values for all forecasts in the telecon-constrained portion of the day versus outlook time. Note the RUCSR and NAMSR values which barely exceed zero at all outlook times. Also note the differing behavior of the RCPF (green; plus signs) with the CCFP Final forecasts (blue; circles).

3.2 Sector-based verification

To this point, verification can be performed for all forecasts in a temporal frame of reference that is relative to key decision-making times for the ATCSCC. To further account for the needs of this consumer, it is noted that the ATCSCC is largely unconcerned with isolated thunderstorms or individual grid boxes; the scales of concern are much larger than the 4 km resolution of NCWD. A more relevant spatial scale for this user is that of an air traffic control sector. There are approximately 200 sectors covering the CONUS and adjacent coastal waters. The sectors vary in size and orientation; smaller sectors are typically used near airports and where there is a high density of air traffic. The ATCSCC planning

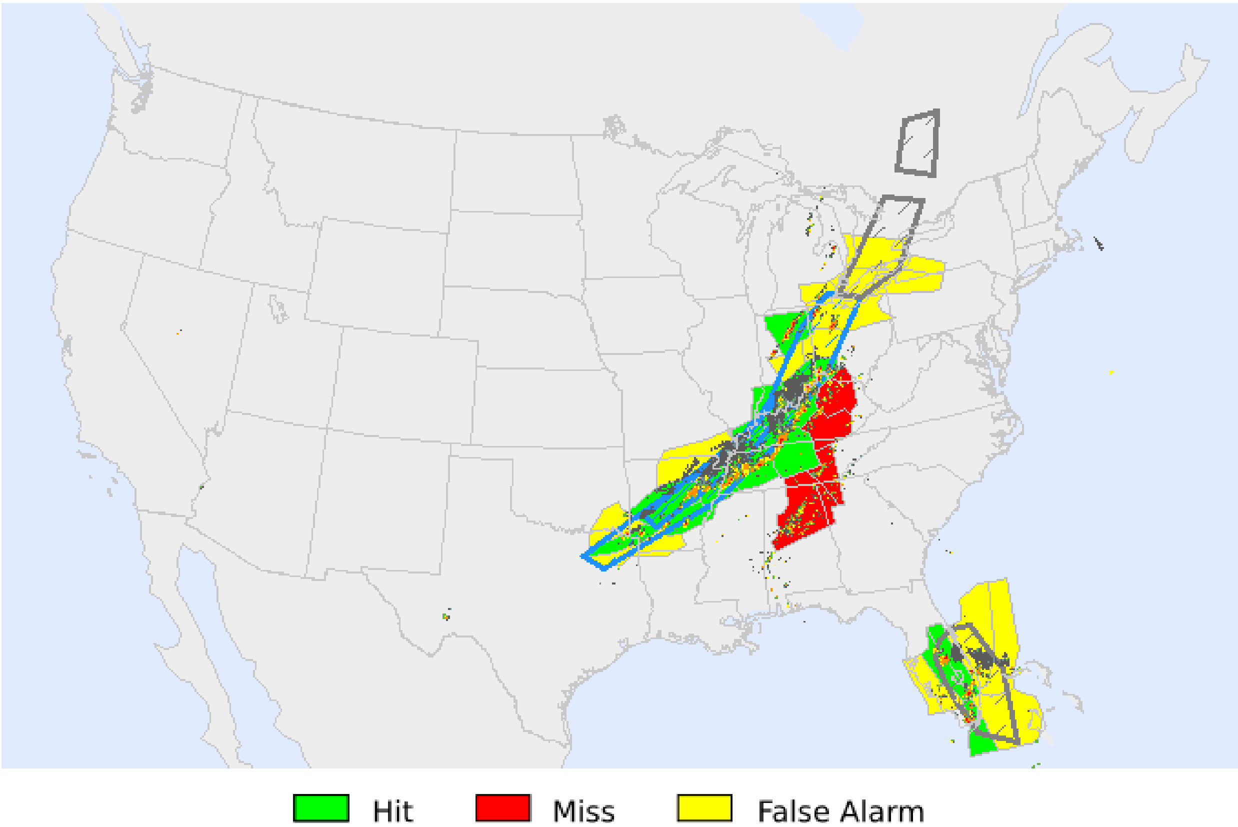

Fig. 3. Sector-based verification of the 2-h CCFP Final forecast from 8 June 2007 issued at 1500 UTC. NCWD observations shown as well. Impacted sectors are color-coded to depict the verification results.

process focuses on the impact of convection on sectors, and groups of sectors, and how to route traffic when these sectors become impacted. It is therefore relevant to view the forecast information in terms of predictions of impacted sectors rather than impacted grid boxes. For this study, the super high sectors were used since they represent the set of sectors that are specific for en-route air traffic. A climatology of observed sector coverages (the sum of the amount of significant convection in the sector divided by the size of the sector) was developed for each sector. From this climatology, a threshold value of 5% was chosen to represent an impacted sector. Therefore, if 5% or more of a sector was covered by a forecast of significant convection, then that is considered a forecast of an impacted sector. If the observed coverage was 5% or greater, then this would be an impacted sector, otherwise it would be considered a non-impacted sector.

An example graphic highlighting a CCFP Final forecast with sectors colored by how the forecast verified is shown in Fig. 3. 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. In contrast to the traditional verification approach, which uses the closest observation time to the relevant valid time, the sector-based approach uses observations that are gathered over a 1-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.

Overall skill, as measured by CSI, for the sector-based verification is shown in Fig. 4. Compared to the CSI panel of Fig. 2, the CSI values for the sector-based approach have become quite large, with peak values near 0.45. This indicates that, when treating the forecasts as prediction of impacted sectors, the forecasts (CCFP Final, CCFP Preliminary, and RCPF) are quite successful. For reasons noted previously, the simulated radar reflectivity forecasts perform poorly at identifying impacted sectors. The CCFP Final forecast outperformed the RCPF for all outlook times and outlook periods except for the 4-h outlook period valid at 1900 UTC. The biases (not shown) were reduced substantially for the sector-based approach compared to the traditional gridded approach with values near one common for the CCFP and RCPF forecasts. The transformation of the forecasts from the view of predictions at specific grid boxes of hazardous convection to predictions of impacted sectors appears to be quite valuable for providing useful for the ATCSCC's needs.

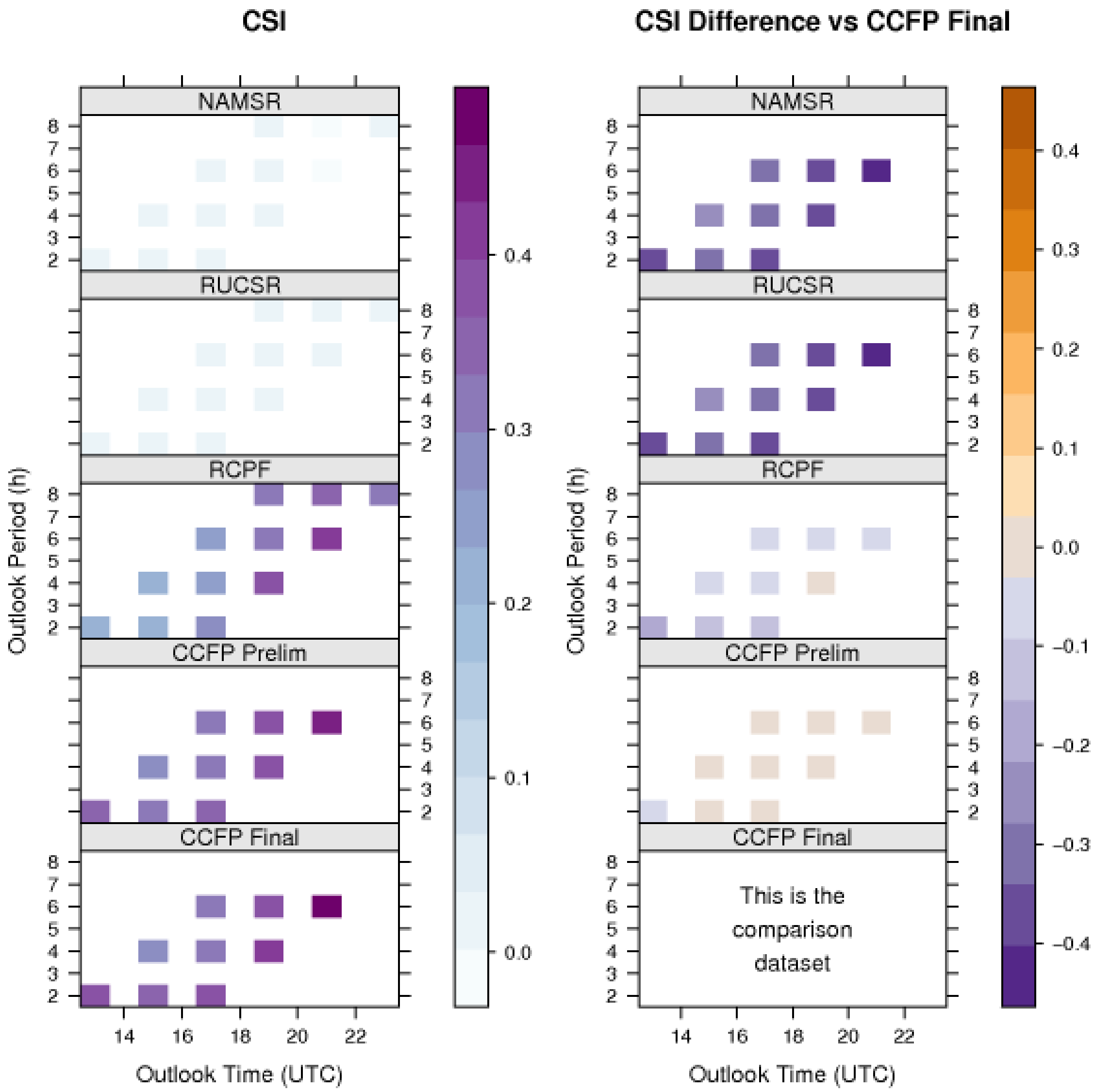

Fig. 4: Level plots of CSI and CSI difference as a function of outlook period and outlook time for all forecasts. Left column shows CSI values for each model. Each square is color-coded by the CSI value for that particular outlook period and outlook time. Right column indicates the difference in CSI values for each dataset from left hand column relative to the CCFP Final CSI values; warm colors indicate where CSI for the forecast exceeds CCFP Final while cool colors indicate where values are less than CCFP Final values.

3.3 *Incorporation of external information*

A very important way to gain additional insights into forecast performance as it relates to a specific user is to incorporate external information into the verification process. Typically, this will be non-meteorological information. The most likely use for this data is to provide a means to categorize and stratify the verification results. For example, the ATCSCC does not treat all weather situations equally. The overall amount of convection over CONUS is important for determining how to control air traffic in the NAS, but where that convection is located is even more important. This view can be summarized in a two-dimensional space as shown in Fig. 5 where one dimension is the amount of convective coverage over CONUS and the other dimension is the potential impact to aviation. This space can be broken down into a series of regimes of varying importance to the ATCSCC decision-making process. If there is a large amount of convection in the NAS, but much of that

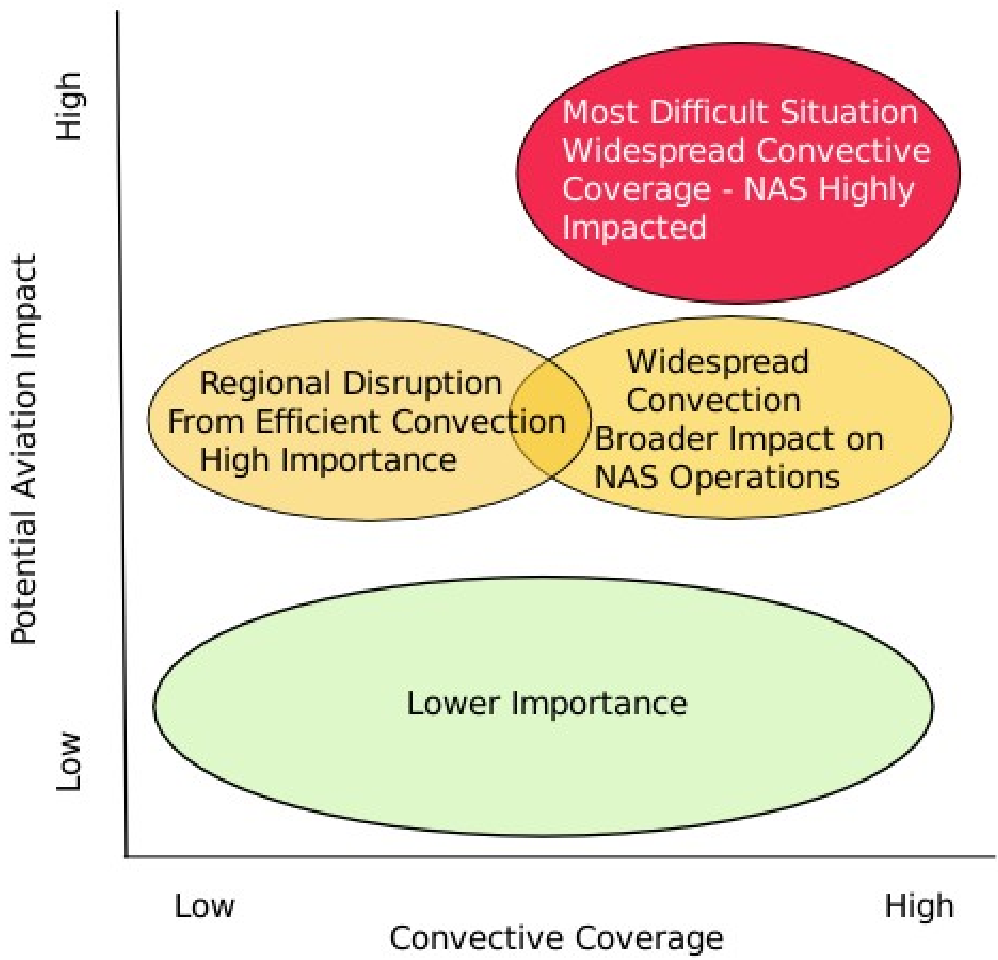

Fig. 5. Schematic depiction of regimes that are of differing importance to aviation planning as a function of location of convection (as measured by potential aviation impact) and overall convective coverage over CONUS.

convection occurs in regions where there is a relatively low amount of air traffic and therefore the potential aviation impact is low, this regime can be considered a low-impact regime. Forecast performance, whether good or poor, is potentially not as important to the operational planning process in these situations. As the aviation impact goes up, typically due to more convection occurring in the eastern U.S., the overall need for accurate forecasts grows accordingly. The most important scenario for the ATCSCC is where there is a large amount of convection, and it occurs where it highly impacts aircraft operations.

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 was 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. 6. The most prominent result is that there is no systematic behavior observable within each outlook period for any of the regimes identified in Fig. 5. This non-result is itself an important result.

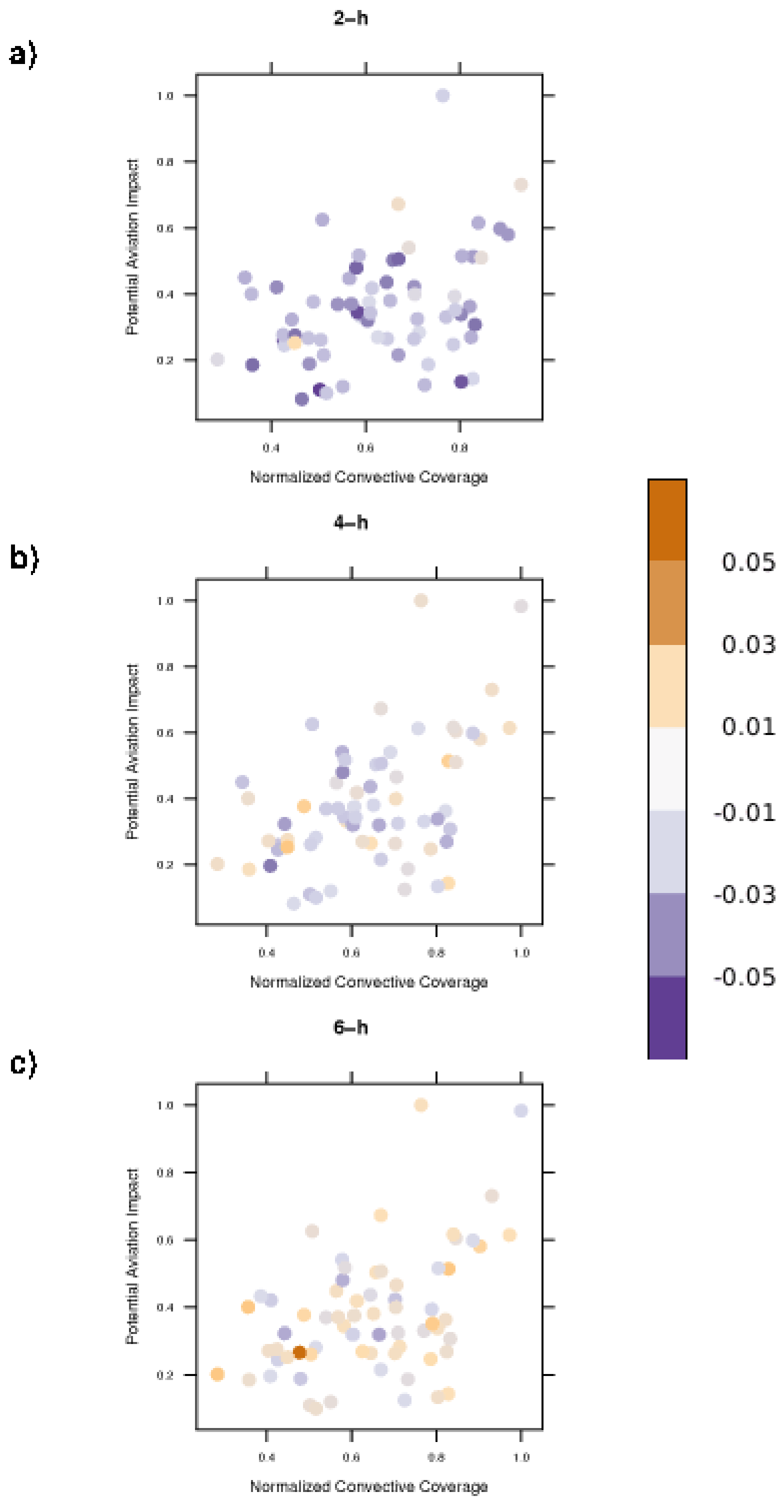

Fig. 6. Difference between average CSI values (RCPF CCFP Final) for a) 2-h, b) 4-h and c) 6-h outlook period forecasts for the telecon-constrained traditional verification approach for each day in the study period as a function of the normalized convective coverage and potential aviation impact. Cool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

4. SUMMARY

This paper has illustrated the importance of consumer-oriented verification using several examples from a recent convective forecast verification exercise. The importance of identifying a user, determining the purpose and goals of the verification, and incorporating the user's needs into the verification results were highlighted. The use of supplemental user-specific information, often in the form of non-meteorological data, was shown to be an important addition to the verification process and served to provide the link between the raw forecast verification and the operational environment. The Network-Enabled Verification Service (NEVS) being developed at the NOAA Earth System Research Laboratory is designed to allow the introduction of user-specific information to flexibly support the operational needs for forecast verification in the NextGen era (Matheson et al. 2008).

While this paper has focused on the use of dichotomous forecast verification, the concepts underlying user-specific forecast verification can be applied to many other types of forecasts. Loughe et al. (2008) discuss the development of a user-specific lead-time metric for forecasts of low ceiling and visibility events. More broadly, the concepts underlying fuzzy-verification techniques, though not stated explicitly in the review paper by Ebert (2007), are directly linked with the goals and needs of consumer-oriented verification.

ACKNOWLEDGMENTS

We would like to thank Danny Sims, FAA System Operations Programs for funding this work.

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.

5. 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.

Ebert, E.E., 2007: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appls.*, in press.

Kay, M.P., S. Madine, J. Luppens Mahoney, and J.E. Hart, 2007: 2007 Convective Forecast Scientific Evaluation. Report prepared for D. Sims, FAA. Available from http://www-ad.fsl.noaa.gov/fvb/publications/, 38 pp.

Loughe, A., S. Madine, J. Mahoney, and M. Graf, 2008: A lead-time metric for assessing skill in forecasting the onset of IFR events. *13th Conf. Aviation, Range, and Aerospace Meteor.*, New Orleans, LA.

Matheson, N.D., M.A. Petty, D.S. Schaffer, S. Madine, and J. Mahoney, 2008: The Network-Enabled Verification Service (NEVS): Integrating data to support user-specific verification. *13th Conf. Aviation, Range, and Aerospace Meteor.*, New Orleans, LA.

Panofsky, H.A., and G.W. Brier, 1958: *Some Applications of Statistics to Meteorology*. Pennsylvania State University, University Park, PA, 224 pp.

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

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp.