THE DESIGN AND EVALUATION OF A MEASURE OF FORECAST CONSISTENCY FOR THE COLLABORATIVE CONVECTIVE FORECAST PRODUCT

Michael P. Kay*
Cooperative Institute for Research in the Environmental Sciences (CIRES)
University of Colorado/NOAA-Research-Forecast Systems Laboratory
Boulder, CO

1. INTRODUCTION

The use of meteorological information for decision-making has never been greater. Forecasts have become ever more sophisticated with the advent of extremely high-resolution numerical models. While numerical models continue to make inroads into operational forecasting, human-generated forecasts remain an integral part of operational weather forecasting today. Modern forecasts produced by humans are created through the integration of a vast array of numerical model output, observational data, as well as other human forecasts. The fact that there are so many data sources available to modern forecasters represents a great challenge to the forecast process itself: namely, how to integrate all of the relevant data available to the forecaster to provide as timely and accurate a forecast possible (Doswell 1986).

This problem is amplified when one considers that non meteorological users are increasingly integrating weather data into their decision-making processes. While meteorologists may know and understand many of the shortcomings and biases of forecasts, end users rarely are knowledgeable of such things. Users may make mental adjustments to forecasts such as attempting to correct for recent model bias (e.g., temperatures are too cold), extrapolation of forecast tendencies, and deriving other forecast products to assist in their own decision-making or forecasting; rarely are forecasts viewed dichotomously as either being wholly correct or incorrect. Recent work by Hamill (2003) on numerical model output found little or no value in adjusting the most recent forecast based upon extrapolation of forecast trends.

We wish to explore one aspect of forecasts related to their use in a decision-making context: forecast consistency. Informally such a topic is referred to among operational forecasters as d(prog)/dt (Hamill 2003). This refers to the change, or lack thereof, in a series of forecasts, each issued at a different time, but all valid at the same time. In particular, such terminology is tied to numerical model evaluation and is often used in a rule-of-thumb sense to make adjustments to model forecasts before making other forecasts or actions based on that data. Such an approach may be perilous for a number of reasons including model updates which change the forecast behavior and perhaps more importantly, that human beings are not good judges of statistical significance especially when small sample sizes are involved (Gilovich 1993).

In the current study, we wish to look at forecast consistency for a human-generated forecast, the Collaborative Convective Forecast Product (CCFP), and the potential uses for such a value. There is a certain intuitive appeal to performing such a study. The consistency of human-generated forecasts has never been actively studied before to the author's knowledge. In addition, the consistency of a series of forecasts is important to

---

*Corresponding author address: Michael P. Kay,NOAA/OAR/FSL, R/FS5, 325 Broadway, Boulder, CO 80305. email: mike.kay@noaa.gov

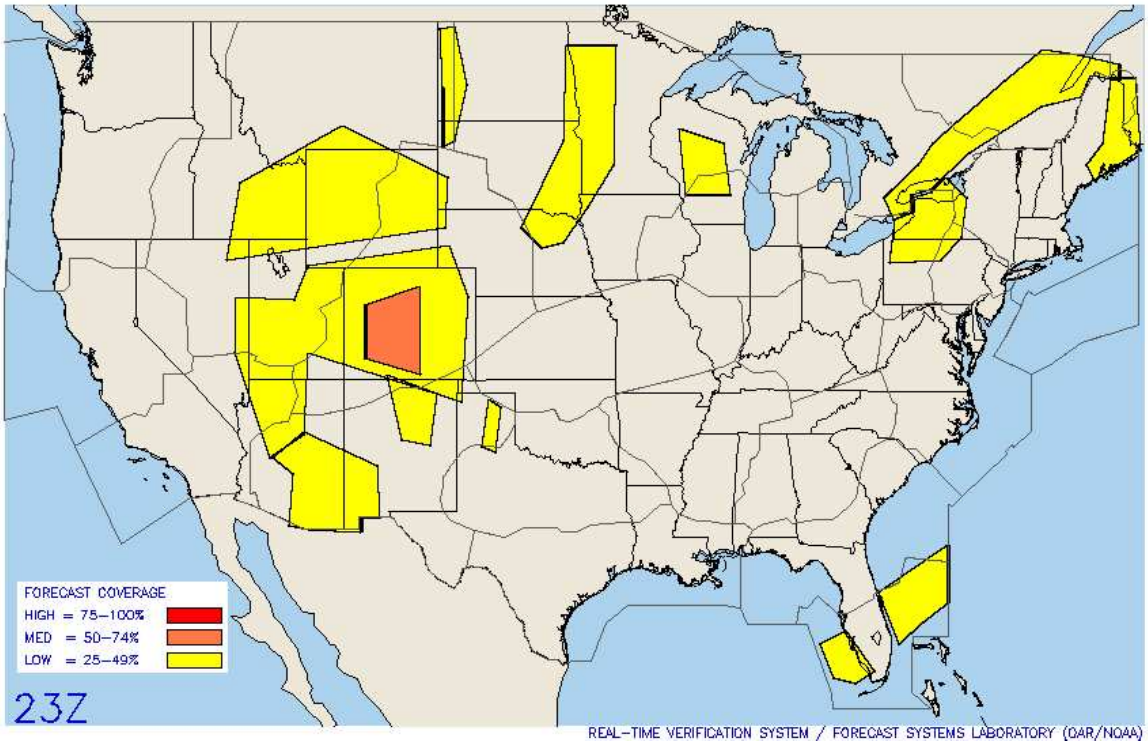

Fig. 1. 2-h CCFP from 20 July 2004 illustrating basic forecast elements of polygons and associated forecast coverage.

users who must make important decisions and take appropriate actions according to the information they receive. For example, the Storm Prediction Center issues a series of areal forecasts, known as the Convective Outlook, several times per day highlighting important areas for severe weather. The forecast includes a description of the relative threat, or risk, for the given forecast period. Many users are highly sensitive to these risk levels and use the information for such actions as altering National Weather Service field office staffing for the day and determining local emergency management activities. If forecasts are highly inconsistent, users are forced to make potentially suboptimal decisions and actions due to the forecast changes. Similarly, strategic planning of air traffic routing requires accurate forecasts in the 2- to 6-h time frame. Conflicting information potentially introduces huge economic consequences when reroutes are unnecessarily introduced. In these situations some amount of consistency is desired by the users. Understanding the consistency of the CCFP may allow its inclusion in the forecast process and provide valuable information to decision-makers.

Section 2 defines consistency and discusses how to present the information to the users. Section 3 illustrates several examples utilizing the CCFP and presents some overall results from the 2003 convective season. Section 3 also includes a discussion about the association of forecast consistency with accuracy as it pertains to the CCFP, something that users do implicitly by considering the forecast consistency to begin with. Finally, section 4 presents the conclusions of this study.

2. DESIGN

The CCFP is a graphical forecast depicting areas over the continental U.S. and adjacent coastal waters where there is a threat of organized convection which may disrupt normal air traffic flow. Forecasts are issued every two hours with 2-, 4-, and 6-h lead-time depictions being produced (Fig. 1). This forecast is the primary tool used for reducing the impact of significant convective weather events on the strategic planning process which occurs in the 2- to 6-h time frame while tactical planning occurs on time scales shorter than 2-h (National Research Council, 2003).

It is instructive to briefly consider how a forecaster would view consistency among a set of areal forecasts such as CCFP. First, in the simple comparison of two forecasts, one would first view whether or not highlighted regions seen on the two are in the same geographic locations. Of most importance to a decision-maker is whether or not a given area is highlighted for certain conditions. Secondary to geographical location is the reason why a certain area is highlighted. In other words, the spatial locations of forecast features between a series of forecasts is perhaps the primary attribute that a forecaster would associate with consistency followed by changes in attributes of those forecast areas.

Based upon this premise, consistency, C, is defined in the following way:

$$C=\frac{cor_{2\_4}+cor_{2\_6}+cor_{4\_6}}{3} \qquad (1)$$

where cor is the Pearson product-moment correlation of the forecast displays. Subscripts refer to the combinations of the 2-, 4-, and 6-h forecasts used for a given valid time.

This definition attempts to mimic how a naïve forecaster would view the consistency of a set of three forecasts. One important modification was made to the correlations used above in (1): entire forecasts (not individual areas) are allowed to move up to 200 km from their original location and the maximum correlation found is used in (1). This is done because a forecaster would see consistency among the areas (shape, orientation, etc.) while allowing for differences in geographic location. The correlations are normalized to the range (0, 1) instead of (-1, 1) for ease of interpretation by the user. For this initial study, equal weighting is used for the correlation terms in (1). It is unclear how to attribute unequal weighting to the correlations otherwise as the strategic planning time window for which the CCFP is used includes the entire 2- to 6-h time range.

While it is possible that forecasters place more emphasis on shorter lead-time forecasts, the correlations between different lead-times in (1) do not allow for a straightforward implementation of a weighting scheme that captures this behavior.

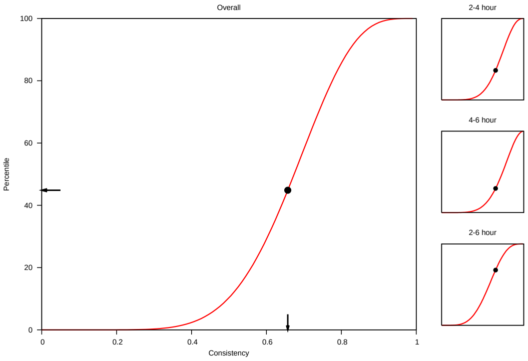

Fig. 2. Display of overall consistency for a CCFP forecast along with correlation coefficients between member forecasts.

Forecast consistency, as defined above, condenses a number of dimensions of the original forecasts themselves into a single number. This value must be displayed in a context where the user can comprehend its significance while at the same time being able to view the components from which it was derived. Such a display must also allow for clear, rapid dissemination of the important aspects of the relevant information. The CCFP consistency is presented to users on a display highlighting the current forecast consistency relative to the distribution of all consistency values for past CCFP forecasts utilizing the cumulative distribution function (CDF). The CDF allows the user to rapidly associate the forecast consistency relative to the consistency of other CCFPs.

To improve the presentation of the empirical CDF, it is modeled by the incomplete Beta function (Wilks 1995) and this is used to generate a smoother display. This information, along with smaller displays showing the correlations between the forecast pairs, is shown in Fig. 2. The user will focus first on the main panel, which highlights the overall forecast consistency, while possibly using the smaller displays, which follow the same format, to gain additional information. Previous work has shown that multipanel plots are very useful once the format of a single panel is ascertained (Tufte 1992).

The consistency distribution graphics such as those shown in Fig. 2 are combined with the the graphical forecast images themselves and these aggregated graphics presented to the user (Fig. 3). Such a display allows the forecaster to see the consistency of the set of CCFPs valid at a given time, the distribution of consistency for all CCFPs, the distribution of correlations between the 2- and 4-hour forecasts, the 4- and 6-hour forecasts, and the 2- and 6-hour forecasts, along with the forecast graphics themselves. This combination of data in one graphic gives decision-makers access to a wealth of information about the CCFP that they may use for making forecasts and other judgments.

The consistency information is available for use shortly after the CCFP has been produced at the Aviation Weather Center and disseminated to users at the Real-Time Verification System (RTVS) (Mahoney et al. 2002) web site[1]. The RTVS is widely used for viewing verification information for a number of aviation-related forecasts. The RTVS is often consulted during the forecast process itself in addition to its use for verification and evaluation purposes at a later time.

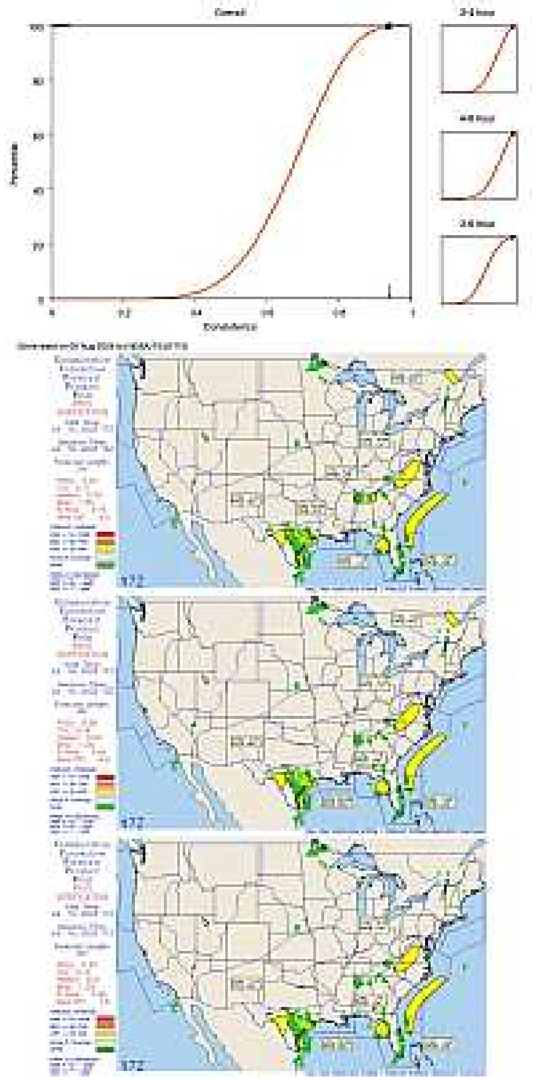

Fig. 3. Consistency display for CCFP from 16 July 2003 valid 2300 UTC illustrating high consistency.

3. RESULTS

Selected results from the 2003 convective season (1 March to 25 October) for the CCFP are presented below. During this period, 1711 combined forecasts were available for analysis. The first result to illustrate is that this measure of consistency does indeed discriminate between forecasts that are highly persistent and those that have large changes. Fig. 4 shows the CCFP forecasts valid at 1700 UTC on 24 June 2003. C for these forecasts is 0.28, which is in the 15th percentile, and indicates a large amount of change between the forecasts. In contrast, the CCFP shown in Fig. 3 has C = 0.94 and

---

[1] http://www-ad.fsl.noaa.gov/fvb/ /rtvs/conv/2001/display/

falls in the 99th percentile.

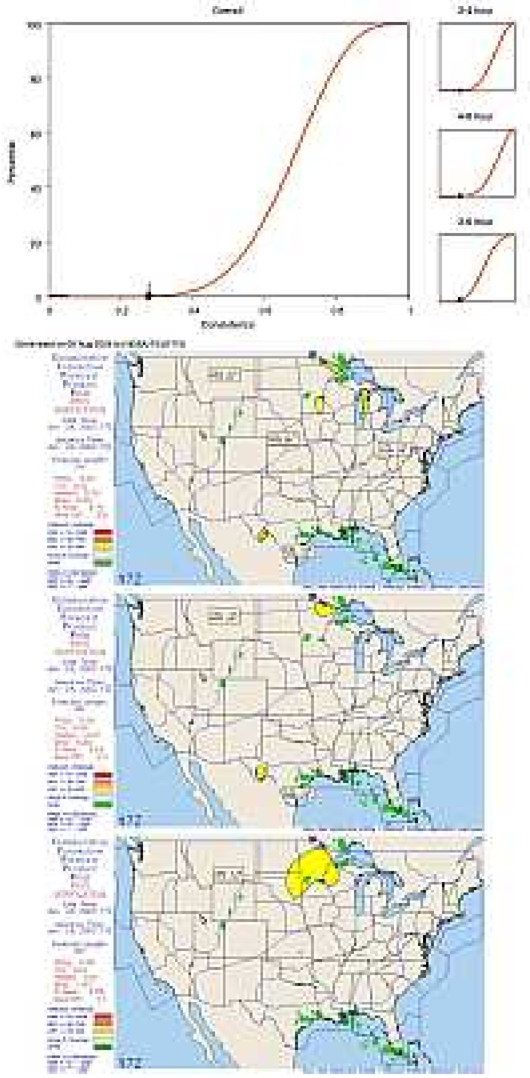

Fig. 4. As in Fig. 3 except for 24 June 2003 valid 2300 UTC. This example shows poor consistency.

The CDFs for the three components of the consistency are shown in Fig. 5. The results highlight the fact that each of the two-hour forecast pairs (2-4 h and 4-6 h) have very similar distributions to each other with comparable mean and variance, while the 4-h correlation (2- and 6-h forecasts) has both a broader spread and a lower consistency (approximately 0.2 less than the 2-hour correlations). This agrees with a logical progression that one would expect better correlation between shorter-range forecasts than at long-ranges due to the diurnal nature of convective evolution where convection typically initiates by early afternoon and dissipates during the evening hours once solar heating of the lower atmosphere has decreased. If the distributions were identical, then one could argue that the forecast process could be refined to producing a single 6-hour forecast instead of the 2-, 4-, and 6-h forecasts.

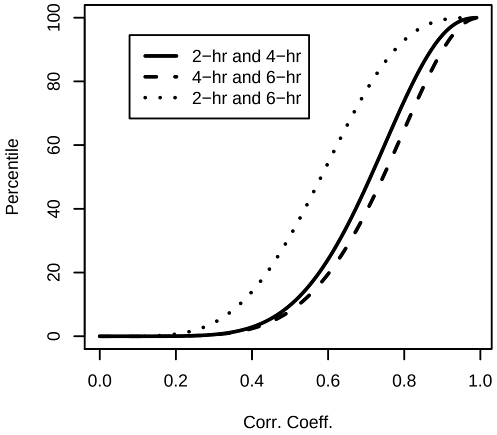

Fig. 5. Cumulative distribution functions for the correlation coefficients for between the 2-, 4-, and 6-hour forecast combinations.

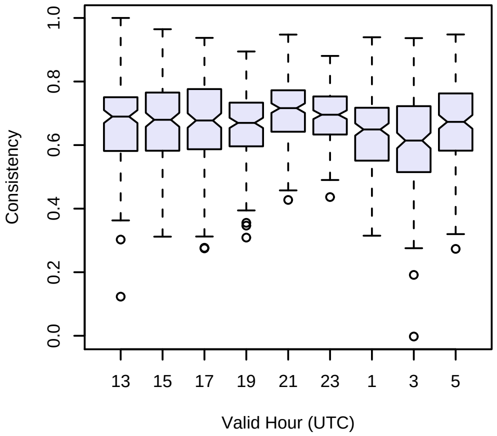

Fig. 6. Box plot of CCFP consistency distributions as a function of time of day (UTC).

The hourly variation in consistency is clearly tied to the diurnal cycle of convection (Fig. 6). The maximum consistency is found at 2100 UTC, while the variation in consistency is smallest for the 2100 and 2300 UTC valid times. Huberdeau and Gentry (2004) found that the CCFP may be used primarily during the earlier parts of the day when strategic planning is most important. Later changes to flight paths are handled with shorter range, tactical planning instead. The mid- to late-afternoon time period, where the CCFP is most consistent, is coincident with a maximum in convective coverage as well as a maximum in forecast areal coverage as well (Fig. 7).

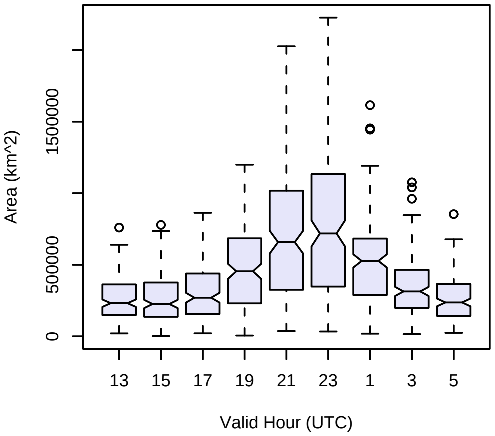

Fig. 7. As in Fig. 6 except for areal coverage of CCFP.

CCFP consistency trends slightly downward during the morning and early afternoon hours indicating that forecasts with differing lead-times that are valid at the same time look less alike with time of day. This coincides well with the arrival of morning upper air observations and later with the arrival of updated numerical model output. CCFP consistency decreases rapidly after 2300 UTC. This could be a result of the scientific limitations of forecasting convective decay on the time scales imposed by the CCFP. This hypothesis appears to be supported by the fact that CCFP accuracy (measured by its critical success index; CSI) reaches its lowest values for the 0300 and 0500 UTC valid times (Fig. 8).

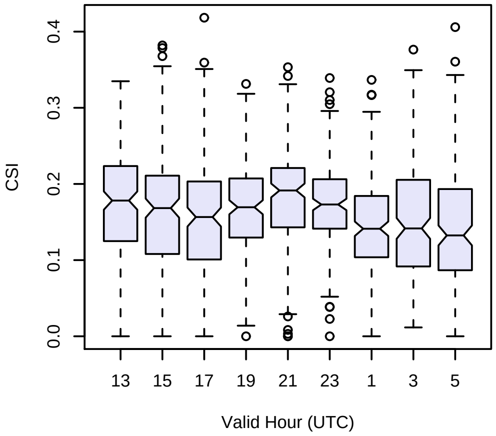

Fig. 8. As in Fig. 6 except for average CSI.

The belief that there is utility in d(prog)/dt as a surrogate for accuracy has driven its informal use as a forecast tool. Hamill (2003) and Dalcher et al. (1988) both studied this use of d(prog)/dt series of lagged average forecasts from a numerical model. Both studies found little to no relationship between d(prog)/dt and forecast skill. In our situation, it is reasonable to think that the human element of the forecasts might lead to a different result for several reasons. Numerical models initialized at different times share information only through initial- and boundary-conditions which may be dependent upon previous runs of the same model. Humans have the benefit of starting with a set of forecasts and, in the simplest case, simply relabeling the previous 4-h forecast as the current 2-h forecast, and the previous 6-h forecast as the current 4-h forecast. Using the previous forecast as an analysis, possibly with no changes, before issuance as the current forecast, allows for potential consistency which is much more difficult to achieve with a numerical model. Additionally, human forecasters are aware of the possible ramifications associated with dramatic changes to their forecasts and thus may seek to minimize such situations, while a numerical model has no such logic involved. Human forecasters also have the advantage of verbal discussion with other forecasters and this type of discussion helps one to minimize inconsistencies and changes to forecasts. While these reasons are purely speculative and may not be applicable to actual situations, they do highlight the potential for a different result than has been previously found for numerical model forecasts.

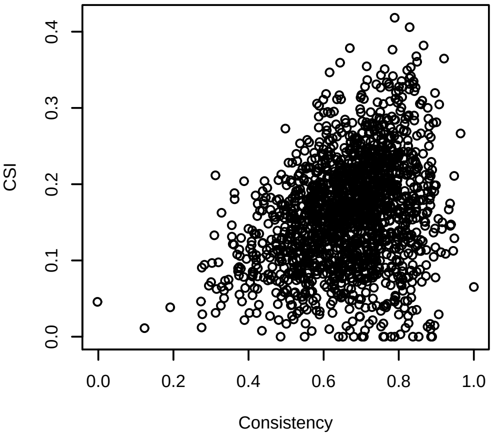

Fig. 9. Scatterplot of CSI vs. consistency for CCFP.

Figure 9 presents a scatterplot of average CSI for the forecast triplets as a function of consistency. As Hamill (2003) notes, many users themselves assume that consistency and accuracy are often used interchangeably. If such a relationship exists, it should become apparent with large sample sizes. The results of this comparison do not confirm that accuracy is related to consistency in any meaningful, linear way; the correlation between the two is 0.31. This data was further segmented into three regimes based upon areal coverage trend: forecasts trending smaller (as lead-time decreases), forecasts trending larger, and those with no apparent trend. This was done to explore whether or not CCFP accuracy is related to trends in the areal coverage of the forecasts. The results (not shown), do not show any meaningful evidence that as forecasts trend smaller as they approach the valid time that the accuracy increases as well. It should be mentioned that since this dataset involves all valid times, the diurnal variation in convection may actually mask any effect which may be present in this data.

4. CONCLUSIONS

This study has shown how forecast consistency, defined as the average correlation between a set of areal forecasts valid at the same time, may be used to provide information to users. The development was approached in such a way that its use in operational settings was considered a priori. Such an approach to new products and tools is absolutely critical as forecasters continue to be inundated with ever-increasing data streams while working under unchanging time constraints. Care was taken to design graphical displays of the information for the forecasters so that they may relate the result, a simple number, to a variety of other information important to the context within which the value should be understood.

Forecasters have long sought means that provide rapid and intuitive summarizations of numerical model forecasts that can help to interpret forecast information (Hamill 2003). Hamill (2003) and Dalcher et al. (1988) both found that for numerical models, care must be taken not to associate d (prog)/dt with forecast skill and use them interchangeably. This study has found a similar result for human-generated forecasts. Perhaps the most important result from this exercise is that while large values of forecast consistency may have little relation to accuracy of the forecasts, the times when forecast consistency is very low should serve as an indicator to users that heightened awareness in the current situation is warranted.

5. ACKNOWLEDGMENTS

Thank you to Sean Madine for numerous insightful discussions on this work, Nita Fullerton for her editorial review, and to Joan Hart for providing access to the CCFP forecast data.

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

6. REFERENCES

Dalcher, A., E. Kalnay, and R.N. Hoffman, 1988: Medium range lagged average forecasts. Mon. Wea. Rev., **116**, 402-416.

Doswell, C.A. III, 1986: The human element in weather forecasting. Natl. Wea. Dig., **11**, 6-18.

Gilovich, T., 1993: How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life. Free Press, 216 pp.

Hamill, T., 2003: Evaluating Forecasters' Rules of Thumb: A Study of d(prog)/dt. Wea. Forecasting, **18**, 933-937.

Huberdeau, M., J. Gentry, 2004: Use of the Collaborative Convective Forecast Product in the Air Traffic Control Strategic Planning Process. Jour. Of Air Traffic Control, 46, 9-14.

Mahoney, J.L., J.K. Henderson, B.G. Brown, J.E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its Application to Aviation Weather Forecasts. 10th Conf. On Aviation, Range, and Aerospace Meteorology, 13-16 May, Portland, OR.

National Research Council, 2003: The Aviation Community's Weather Forecast Needs. Weather Forecasting Accuracy for FAA Traffic Flow Management: A Workshop Report. 68 pp. Available from http://www.nap.edu/openbook/0309087317/html/1.html.

Tufte, E.R., The Visual Display of Quantitative Information. Graphics Press, 197 pp.

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