8.34 **NEAR REAL-TIME VERIFICATION AT THE FORECAST SYSTEMS LABORATORY: AN OPERATIONAL PERSPECTIVE**

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

Jennifer L. Mahoney
NOAA Research-Forecast Systems Laboratory
Boulder, Colorado

1. INTRODUCTION

Meteorological forecasts are created for a wide spectrum of uses in society. They range from forecasts of maximum expected temperature to predictions of rare and severe events that present a threat to human life and property. In order to provide the highest level of service possible, every forecast created should have a corresponding verification component that provides information concerning its performance. Both the creators and users of the forecast products benefit from such information, especially when the forecasts are verified thoroughly and systematically. To support such a premise, the Forecast Systems Laboratory (FSL) has developed an interactive, Web-based system called the Real-Time Verification System (RTVS) to support operational forecast verification.

The purpose of this paper is to discuss the design of operational verification systems, the history of the RTVS, and some of the challenges associated with systematic, operational forecast verification. First, an introduction to verification systems is given followed by a history of the RTVS. Issues and challenges encountered with real-time verification are then discussed followed by a summary and plans for the future.

2. VERIFICATION SYSTEMS

Upon initially considering forecast verification, most individuals think of it as a simple algebraic exercise involving the comparison of one set of numbers with another. In reality the modern verification process encapsulates a much broader set of interconnected components that often cut across numerous disciplines including meteorology, statistics, and computer science. A simple illustration of these components and the flow of information through them is shown in Fig. 1. Many of the components are taken for granted when performing verification; however, they are all necessary when creating a verification *system*.

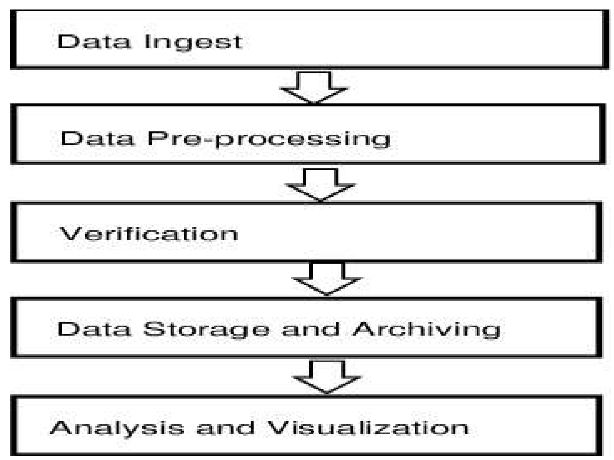

Fig. 1. Primary components of a verification system. Arrows indicate flow of information through the system.

Data ingest involves obtaining all of the forecast and observational datasets to perform the necessary comparisons. In many instances, the relevant forecast data comes from a local numerical model simulation while the observations, or comparison data are obtained separately. Data preprocessing potentially involves many steps that transform the forecasts and/or observations into a format that allows the verification to take place. Examples of processes included in preprocessing are transformation from text descriptions of forecasts into numerical equivalents, interpolation to a set of predefined grids, and conversion to a unified data format, such as network Common Data Format (netCDF).

* *Corresponding author address:* Michael P. Kay, R/FS5, Forecast Systems Laboratory, 325 Broadway, Boulder, CO 80305; e-mail: mike.kay@noaa.gov

The actual verification involving the comparison of forecast and observational data is performed only after the first two steps have occurred. The verification will generate one or more types of output which must then be stored for later retrieval as well as analysis. Example output data might include important stratifications and linear combinations of statistics stored in files or a relational database, as well as static graphical displays. A modern verification system should utilize a relational database management system (RDBMS) for the management of the postprocessed verification information. An RDBMS allows for the creation of more complex stratifications and aggregations of data to assess forecast performance than could reasonably be performed otherwise with flat files.

The final part of a verification system is the component that allows for the analysis and visualization of the results. The analysis and visualization along with the verification component constitute what many people consider to be forecast verification; it is the scientific part of the sequence that allows one to discover the strengths and weaknesses of the underlying forecasts and observations. The interrogation component of a verification system should always be designed such that remote access is available, since all concerned parties may not be local.

The explosion of the Web as a communication medium allowed the widespread dissemination of verification information in a manner not available previously. By the late 1990s, the Web browser had clearly become the preferred tool for presenting forecast verification information to all users, including forecasters in both local and remote locations. Nurmi (2004) notes that Web pages and forms create the preferred interface to verification information in modern systems.

The importance of the union of the RDBMS with an interactive Web interface to create arbitrary, user-defined requests for information cannot be overstated, and has become the paradigm through which the RTVS and now several other verification projects display information (Nurmi 2004). Both tools represent radical expansions in capabilities from what was possible only a decade ago and allow for real-time activities in an operational setting that previously could only be attained in a research setting.

3. HISTORICAL ASPECTS OF RTVS

3.1 Initial Development

FSL has a long history of collaboration with the Federal Aviation Administration aimed at improving aviation meteorology. Beginning in the early 1990s, a verification program was created within the Aviation Division at FSL to evaluate the accuracy of a number of numerical model forecasts of direct variables, such as temperature and wind speeds, along with several derived variables, such as turbulence and icing. The initial work focused on understanding the problem domain and attempting to create a framework through which an organized verification effort could be undertaken (Mahoney et al. 1997). Early efforts were plagued by a number of problems such as difficulties in recreating necessary datasets, and limited and nonoverlapping time periods leading to small sample sizes that discouraged intercomparisons. These issues motivated the initial development of the RTVS in 1997, which attempted to account for all of these difficulties. Additionally, an effort was made to begin performing verification in realtime to eliminate some of the problems encountered with verification performed in a historical, or delayed mode. Throughout this paper the words “real-time” should be considered synonymous with “as soon as the forecasts and observations are available.”

The initial RTVS consisted of a single workstation that contained all of the aforementioned components except for the data ingest, which was already handled at the laboratory level. A version of RTVS was delivered to the Aviation Weather Center (AWC) in late 1997 to provide a long-term view of forecast performance and serve as a decision aid in the forecast process (Mahoney et al. 1998). AWC and the Federal Aviation Administration Air Traffic Control Strategic Command Center continue to utilize the RTVS to assess their forecasting performance that affect both strategic and tactical decision-making.

3.2 Maturity of RTVS

By the late 1990s, the RTVS had grown

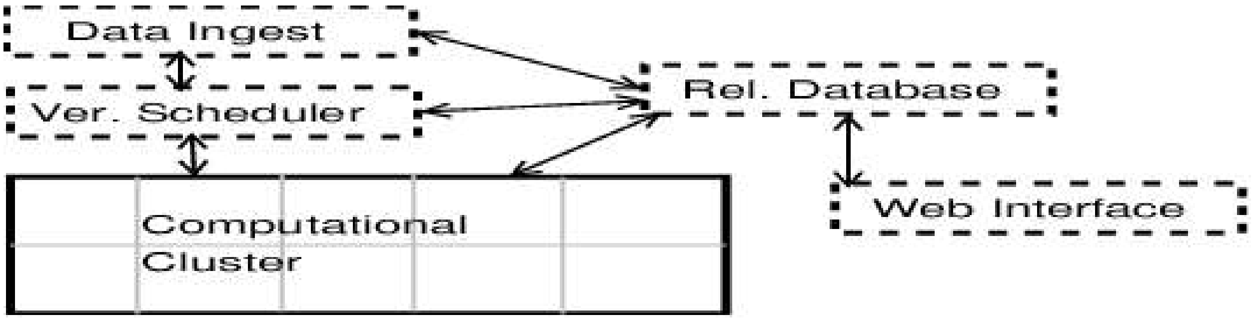

Fig. 2. RTVS hardware configuration. Dotted lines indicate machines which have automatic failover capabilities to redundant hardware.

from supporting a handful of forecasts of turbulence and icing to the verification of a large, diverse set of forecasts related to all aspects of aviation meteorology with a particular emphasis on human-generated forecasts (Mahoney et al. 2002). The increase in complexity necessitated changes to the RTVS infrastructure. Data preprocessing and verification were separated onto different computers and a large, unified underlying data storage infrastructure was implemented to deal with the increased amount of input and output data.

Perhaps the greatest change to the system occurred on the back-end of the system where all data, excluding static displays, were placed into an RDBMS, and the user interface was enhanced and migrated from a standalone application to a series of World Wide Web interfaces.

3.3 Current RTVS

The RTVS currently supports a wide variety of experimental and operational human-generated and model-based forecast products in the following core areas: convection, icing, turbulence, precipitation, ceiling and visibility. The RTVS is primarily used for three purposes: 1) to support the transition of forecast products through the FAA Aviation Weather Research program from research into operations, 2) to establish long-term baselines of performance for a wide variety of aviation hazard forecasts and 3) to support field projects and other outside exercises.

For all verification efforts, the goal is to provide verification results to all users in realtime, or as quickly as possible. Murphy and Daan (1984) describe a forecasting experiment in the Netherlands where extensive feedback was given to forecasters between the first and second years of the experiment. Marked improvement was seen in forecast performance in the second year, which was partially attributed to the feedback that the forecasters received. It is the goal of RTVS to provide an infrastructure that supports the potential for similar gains within a broader framework.

The RTVS is considered operational and runs 24 hours a day, 7 days a week without human intervention. It is important to realize that automation is crucial in the consistent, timely dissemination of verification information to all users. The RTVS Web site is segmented into areas for both operational and experimental forecasts. The operational piece is concerned with products that are stable in the sense that they are not changing from year to year and whose long-term quality assessment is important. The experimental area is where new techniques, interfaces, and forecasts are first implemented for evaluation. Experimental products are continuously verified in a real-time manner much like the operational part of the system. User interfaces are kept consistent between the operational and experimental areas wherever possible.

The computational infrastructure supporting the RTVS is shown in Fig. 2. Verification jobs are run in both scheduled and asynchronous modes on a 10-node, 20-CPU cluster running the Linux operating system. The current system processes more than 10 gigabytes of data per day necessary to support existing verification projects. Online storage capacity of the RTVS is currently approaching 7 terabytes. This storage is used to maintain both short- and long-term repositories of raw and preprocessed data, as well as static, plan view displays which cannot be generated efficiently in realtime. A very large amount of data storage is necessary to support any verification project, especially when there is the potential for either reprocessing using existing techniques or in the development of new techniques. Data storage requirements will continue to expand rapidly in the near future as greater numbers of gridded forecasts, especially numerical model forecasts, become available at increased spatial and temporal resolutions. The computational cluster is networked with redundant, high-availability Web and database servers with automatic failover capabilities to present output data to users. A development environment consisting of cluster, Web, and database machines is currently being established so that software development and testing can occur in a parallel environment to operations without impacting any of the operational systems.

4. CHALLENGES

As stated previously, the implementation of a forecast verification system is clearly a multidisciplinary undertaking. Significant challenges arise in all areas when attempting to design a system that satisfies the needs of all users. In the following section, some of the more important hurdles that have been encountered in the past nine years of operational forecast verification will be discussed with an emphasis on the verification and analysis components.

4.1 Complete Knowledge of the Dimensions

Before verification can be performed on a forecast, all relevant information, or dimensions, must be known before one can provide a thorough assessment of the product. The complete set of dimensions is important for the design of the verification component as well as the way that the data fit within the broader verification system. Murphy and Brown (1984) enumerated the dimensions associated with the complete description of user requirements for nowcasts and very short-range forecasts. The information required includes a) the variable or event, b) the spatial domain, c) the temporal domain, d) lead time for the forecast, e) the form of the forecasts, f) the content and formation of information packages, and g) the communication or dissemination media. The reader is urged to consult Murphy and Brown (1984) for more information about each of the points listed above. The verification of forecasts requires the specification of the complete set of dimensions listed above along with two additional dimensions: h) the observation(s) and i) the verification technique(s). These additional dimensions are discussed below. Note that there is a further, more extreme, issue with dimensionality associated with the forecasts and observations related to the verification of all appropriate stratifications and combinations of the dimensions listed above. Murphy (1991) discusses the problems associated with the high dimensionality of forecast verification.

4.2 Verification Techniques

The term “verification techniques” refers to the methods through which forecasts and observations are brought together in order to generate resulting statistics. In some instances this is as simple as taking two temperature grids, a forecast and some measure of truth, and computing the mean squared error for the entire grid. For data such as precipitation, one must choose how to treat the variable before any attempt at verification can be made. Precipitation is a continuous variable and can be verified using continuous statistics such as magnitude bias or mean squared error. The data could also be transformed into a dichotomous format such as whether or not 0.01 in of precipitation occurs in the forecast period. Skill and summary scores can then be computed on these data. Yet another method is to treat contiguous areas of precipitation as individual entities and to decompose the errors found in a variety of ways (Ebert and McBride 2000). Each of these methods provides different information concerning forecast performance. In many other instances, novel approaches have been developed to account for systematic deficiencies in available forecasts and observations (Brooks et al. 1998). In our experience, the most fluid part of the system is the verification component where methods are developed and modified.

4.3 Forecast Definition

Quite frequently, forecasts are defined in such a way so as to make verification either impossible or very difficult. One situation where

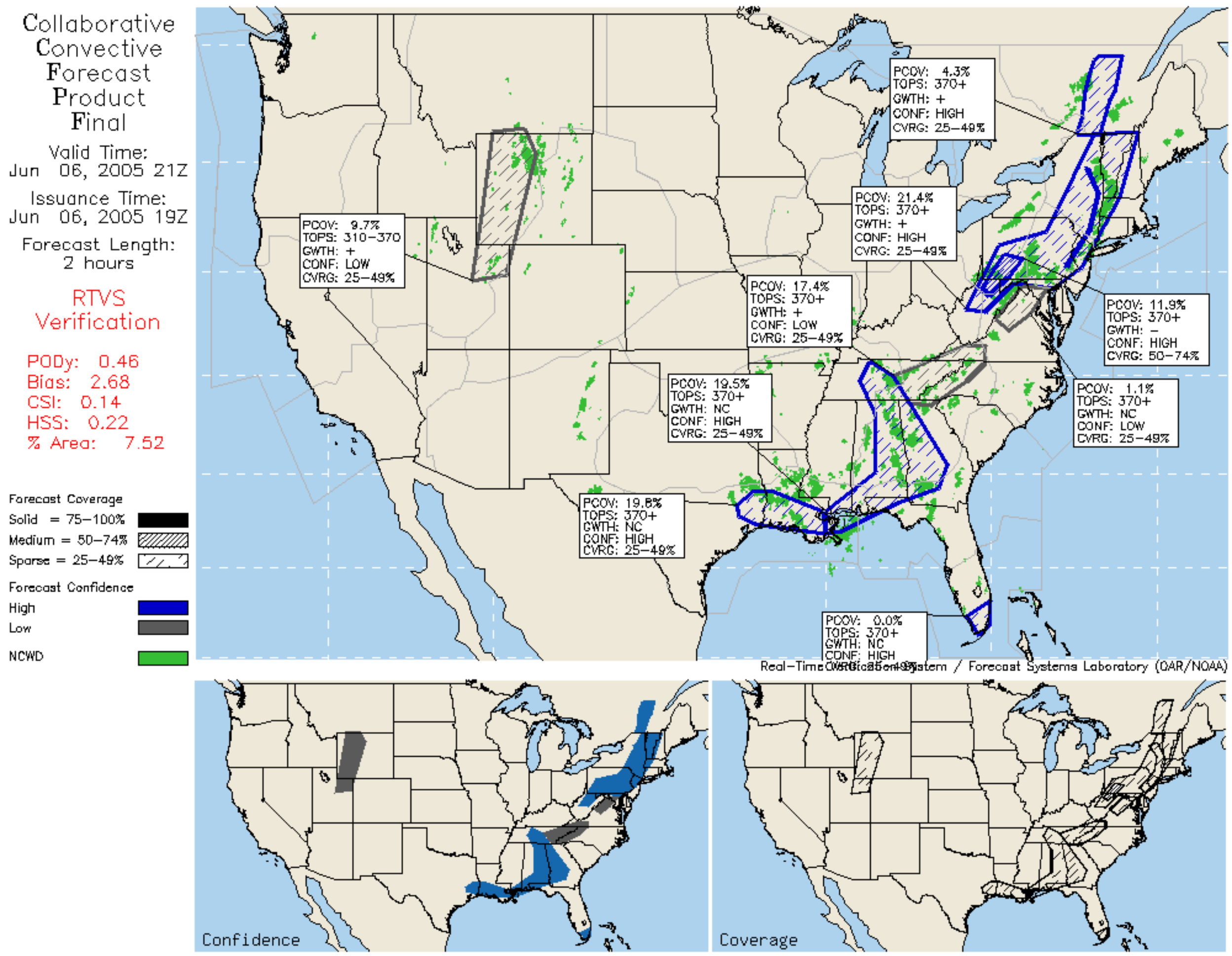

Fig. 3. Sample convective verification display from RTVS illustrating forecasts, observations, and resultant verification statistics. Lower panels redisplay the primary forecast attributes, expected areal coverage and forecaster confidence, to aid comprehension.

this frequently occurs is for forecasts that convey information for which no direct verifying observation exists. Free-format forecasts which may be partially or entirely composed of

nonstandardized data often cannot be parsed to provide more than a cursory description of a more rich set of attributes that exist within the forecast. These types of problems can often be eliminated when verification is considered in the initial design of forecasts (Kay and Brooks 2000). Categorically worded formats using words like “moderate” or “likely” continue to be prevalent without well-defined numerical equivalents that can be related to observational data. See Vislocky et al. (1995) for more examples of forecasts whose communication and dissemination format precludes an unambiguous interpretation.

4.4 Observational Datasets

One of the greatest challenges faced by those performing meteorological verification will always be finding appropriate independent observational datasets. The assessment of the performance of nonsurface-based forecasts is very difficult where nonsystematic observations and data-sparse regions dominate. Forecasts for which only remotely sensed data are available for verification create significant problems and limit analyses to intercomparisons where “truth” is often unknown. A more recent development that will remain a future problem is the lack of independence between forecasts and observations when the forecasts are algorithms that combine numerous datasets along with sophisticated logic to create a new product.

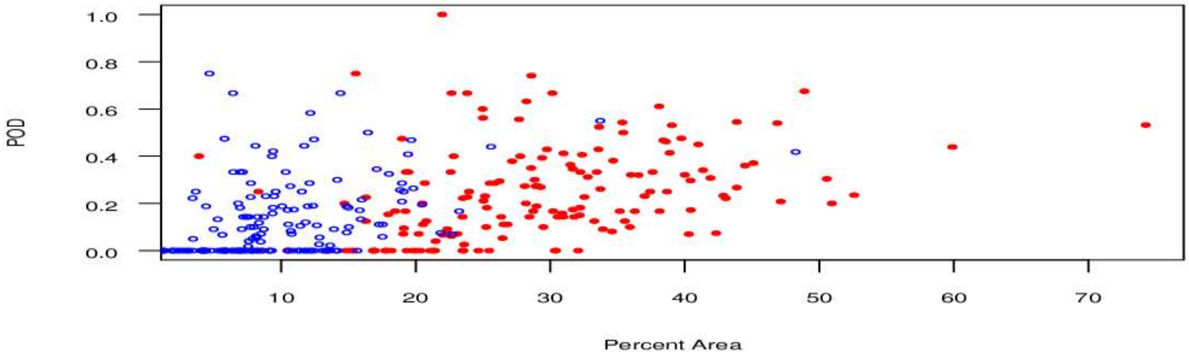

Fig.4. Scatterplot of probability of detection (POD) as a function of observed fractional areal coverage (percent area) of turbulence in the 20000-40000 ft layer from two numerical forecast algorithms from 1 January 2005 to 12 June 2005. Algorithm 1 is indicated by open blue circles and Algorithm 2 is shown by the solid red dots.

4.5 Appropriate Information for Short-Term Decision-Making

The authors are unaware of any formal studies on how to integrate verification information directly into the forecast process itself. In the first author's experiences at the United States National Weather Service Storm Prediction Center (SPC), forecasters actually do integrate subjective forecast verification directly into the forecast process of nowcasts and short-term forecasts. Kain et al. (2003) have discussed some of the benefits associated with subjective forecast verification. SPC forecasters produce a number of spatial forecasts that are similar in spirit to those produced by the AWC. SPC forecasters are highly adept at reviewing, subjectively, maps associated with forecasts from both the previous day as well as the current day and using this information in the preparation of future forecasts.

There are several indications that AWC forecasters are performing in a similar manner. First, The AWC has requested that the shortest lead-time, 2-hour Collaborative Convective Forecast Product verification graphic showing the forecast, observations, and corresponding summary statistics be made available in order that the information is available for the next product issuance. Simple, targeted displays help reduce the need to expend time and resources to bring verification information into their workflow (Fig. 3). Graphics produced in RTVS such as Fig. 3 are displayed within an AWC web page directly with the corresponding real-time forecasts so that users have an integrated information source for decision making. The second indication that the RTVS is being used for decision-making comes from an analysis of Web server log statistics for RTVS from May 2004 through May 2005. Logs have been analyzed to determine what sources of information users request most frequently. The results (not shown) suggest that users of RTVS are most often interested in plan view maps of convective forecast verification. Requests for plan view displays of convective forecasts accounted for 20% of all RTVS Web requests during the 13-month period. While the motivations for this result are unknown, it is plausible that the behavior of AWC forecasters is similar to that of the SPC.

RTVS provides two basic types of verification tools: display tools combining maps and displays of forecasts and observations, and interrogation tools that allow for dynamic, user-driven, aggregate statistical results such as scatterplots and time series for arbitrary time periods, regions, statistics, etc. The interrogation tools appear to be most useful for postmortem verification exercises including long-term baselining and assessing overall behavior of the forecasts. The turbulence interrogation tool was the second most requested interface in RTVS. This tool is used by both algorithm developers and forecasters to validate and intercompare approximately 20 different numerical algorithms for turbulence prediction. The shift in usage statistics appears to correlate well with a change in the primary audiences. The convective forecasts are well-understood and forecasters are using verification information to improve their performance, whereas the turbulence users are primarily developers interested in understanding the behavior of a large set of competing, rapidly evolving set of forecasts.

Fig. 4 shows a scatterplot produced through the RTVS turbulence interrogation tool. Displays such as this are useful in determining systematic behaviors in a non-realtime setting. Information gleaned from comprehensive analyses of plots such as these can then be fed back to forecasters for realtime forecast situations. The scatterplot in Fig. 4 depicts the probability of detection as a function of observed coverage of upper-level turbulence for the first part of 2005. One might expect the algorithms to perform better in situations when turbulence occurs over broad areas. Neither algorithm shows significant improvement in situations where turbulence is widespread. Additionally we learn that the different algorithms forecast quite different areal extents of turbulent conditions in the upper part of the troposphere. Information such as this will be of interest to algorithm developers and users. The RTVS interrogation tools provide a framework for these types of systematic comparisons for user-defined date ranges, forecast periods, and other stratifications such as turbulence severity.

4.6 Target Audience

The diversity of the different user groups of verification information presents a challenge to those designing information interfaces. We believe that the primary users of RTVS can be segmented into three groups: forecasters, managers, and product developers. Each of these user groups have very different needs and expectations. Requirements for each group remain poorly understood. Displays and maps, produced in realtime, appear to be the most useful sources of information for forecasters. Forecasters are already overwhelmed with information that must be condensed and interpreted in their forecast process. Managers are most likely interested in long-term, overall performance augmented by individual forecast performance, while product developers benefit from a comprehensive analysis of both short- and long-term data.

5. SUMMARY

FSL's Real-Time Verification System has been utilized for over eight years for a wide range of verification tasks ranging from long-term baselines of operational products to verification in support of field and research projects. The system has been designed around a basic framework for verification systems that specifies a set of interrelated components that take raw forecasts and observations and lead to summary statistics and displays that can be used for decision-making in realtime. Our experience performing operational verification has shown that there are some difficulties that one must consider when implementing these types of systems, including understanding the complete dimensionality of the forecasts and observations, choosing appropriate techniques, and knowing the intended audience so that information can be tailored to best suit its needs. The challenges are daunting, but the premise that such a system can be used to improve operational forecasts makes the endeavor very worthwhile.

The RTVS can be accessed on the Web at

http://www-ad.fsl.noaa.gov/fvb/rtvs/

6. REFERENCES

Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints*, *19th Conference on Severe Local Storms,* Minneapolis, MN, Amer. Meteor. Soc., 552-555.

Ebert, E.E., and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Kain, J.S., M.E. Baldwin, P.R. Janish, S.J. Weiss, M.P. Kay and G.W. Carbin. 2003: Subjective Verification of Numerical Models as a Component of a Broader Interaction between Research and Operations. *Wea. Forecasting,* **18**, 847–860.

Kay, M.P., and H.E. Brooks, 2000: Verification of probabilistic severe storm forecasts at the SPC. *Preprints, 20th Conference on Severe Local Storms,* Orlando, FL, Amer. Meteor. Soc., 285-288.

Mahoney, J. L., J.K. Henderson, and P.A. Miller, 1997: A description of the Forecast Systems Laboratory's Real-Time Verification System (RTVS). *Preprints, 7th Conference on Aviation,* *Range, and Aerospace Meteorology*, Long Beach, CA, Amer. Meteor. Soc., J26-J31.

Mahoney, J.L., J.K. Henderson, B.G. Brown, C.S. Hartsough, 1999: Benefits of using FSL's Real-Time Verification System (RTVS) at the NWS/NCEP Aviation Weather Center. *Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, Amer. Meteor. Soc., 95-99.

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 Forecast. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., 323-326.

Murphy, A.H., 1991: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev*., **119**, 1590-1601.

Murphy, A.H. and B.G. Brown, 1984: User requirements for very-short-range weather forecasts: Some key issues. *Proceedings, Nowcasting-II Symposium*, 3-7 September, Norkopping, Sweden, 491-494.

Murphy, A.H., and H. Daan, 1984: Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee Experiment. *Mon. Wea. Rev.*, **112**, 413-423.

Nurmi, P., 2004: Design of operational verification systems. Presentation, International Verification Methods Workshop, Montreal, CN,

available at
http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/Workshop2004/MeetingProgram.html.

Vislocky, R.L., J.M. Fritsch, and S.N. DiRienzo, 1995: Operational omission and misuse of numerical precipitation probability expressions. *Bull. Amer. Meteor. Soc.,* **76**, 49-52.