4.12 BENEFITS OF USING FSL'S REAL-TIME VERIFICATION SYSTEM (RTVS) AT THE NWS/NCEP AVIATION WEATHER CENTER

Jennifer Luppens Mahoney*, Judy K. Henderson[1], Barbara G. Brown[2] and Craig. S. Hartsough[1],
Forecast Systems Laboratory, Boulder, CO

1. INTRODUCTION

The goal of the Federal Aviation Administration (FAA) Weather Research Program (WRP) is to provide the capability to generate more accurate and accessible weather observations, warnings, and forecasts (Sankey et al. 1997). Verification tools provide the mechanisms for assessing the quality and effectiveness of these services (Rodenhuis et al. 1997). As a result, the FAA WRP is funding the development FSL's Real-Time Verification System (RTVS) to assist the National Weather Service (NWS) Aviation Weather Center (AWC) in reaching these goals: improving performance, developing products that are more user-friendly and relevant to a broader user base, and ensuring that warnings are formatted and worded for ease of use by computers and humans (Kelly, 1998).

Since the implementation of the RTVS at AWC in November 1997, strides toward these goals are being realized. For the first time, AWC management is able to objectively monitor and evaluate the quality of the in-flight aviation weather advisories issued for icing, turbulence, convection, and Instrument Flight Rules (IFR). In addition, forecasters can also evaluate the quality of guidance products used to make their forecasts. Just as importantly than quality assessment, the RTVS is becoming the driving force behind evaluating what the forecast products represent and provide to the broader user base.

In this paper, we describe the benefits of using RTVS at AWC and its role in shaping the future of AWC forecast products.

2. BACKGROUND

Here is a brief description of the RTVS and its verification techniques.

2.1 RTVS System Description

The RTVS is a verification tool that is being developed to provide a statistical baseline for AWC in-flight advisories and model-based guidance products, support real-time forecast operations, and support model-based algorithm development and case study assessments. To this end, RTVS was designed to ingest AWC-issued and model-based forecasts for icing, turbulence, IFR, and convection in "real-time" (i.e. as soon as they become available), and to store the relevant information in local, long-term storage. The RTVS' real-time ingest and local storage allows: different models and algorithms to be easily compared to each other and to AWC forecasts, large sample sizes to be quickly gathered and exploited for statistically significant comparisons and long-term monitoring, and a "quick-look" at, and comparisons of, the accuracy of forecasts valid at the current time (Mahoney et al. 1997).

2.2 Verification Methods

The verification methods used in RTVS (Mahoney et al. 1998) closely follow those developed by Murphy and Winkler (1987) and Brown et al. (1997). Murphy and Winkler (1987) described a general framework for forecast verification that encompasses the characteristics of the forecasts, the corresponding observations, and their relationship. Brown et al. (1997) applies these concepts to aviation verification problems and defines the probability of detection (POD) and POD*no* as statistical measures used to evaluate the quality of forecast products. In the case of POD, the probability would be the probability of a "Yes forecast given a Yes observation," while for POD*no*, the probability would be a "No forecast given No observation." Due to limiting characteristics of the PIREPs (Schwartz, 1996), it is not possible to compute the False Alarm Ratio (FAR), which typically would be computed as a measure of overforecasting (Brown et al., 1997). Thus, the "Impacted Area" and "Impacted Volume" are computed. These methods represent the total area and volume encompassed by the advisory.

---

* *Corresponding author address*:
Jennifer Luppens Mahoney
R/E/FS5, 325 Broadway, Boulder, CO 80303;
email<mahoney@fsl.noaa.gov>

[1] Joint collaboration with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

[2]. Research Applications Program. National Center for Atmospheric Research, Boulder, CO

This provides a surrogate measure of overforecasting. The goal is to minimize the Impacted Area while maintaining a high detection rate.

To obtain the forecast/observation pairs used to compute POD, the observations are tested to determine whether they fall within the temporal and spatial constraints of the advisory and are reporting the forecasted weather conditions. If the observation is a "Yes" and lands within the boundaries of the advisory, then a Yes - Yes forecast/observation pair is recorded. If the observation is a "Yes" and falls outside the boundary, then a Yes - No pair is recorded, and so on until pairs of all four possible combinations (Yes-Yes, Yes-No, No-Yes, No-No) are obtained. All advisories that fall within the specified valid times are used to create the pairs. The forecast/observation pair combinations are collected according to the location and time of the observations, not by the forecasts.

3. BENEFITS

In this section, two specific case examples are used to describe the benefits of using the RTVS at AWC for quality assessment and product.

3.1 Quality Assessment

3.1.1 Definition of Quality

Before being able to evaluate the "quality" of AWC products, the definition of quality must be fully understood. Murphy (1993) describes quality as the degree of correspondence between the forecasts and the observations. Thus, high quality forecasts exhibit a close correspondence with the observations. However, forecast quality is inherently multifaceted in nature. As stated by Brown et al. (1997), evaluating only one characteristic of forecast quality can lead to mistaken conclusions about forecast performance. For instance, Impacted Area and Volume provide some guidance, but they cannot alone be used to define the quality of the forecast. One way to handle this problem is for the forecast users to specify a minimum acceptable criterion for at least one of the verification statistics. Then, the best forecast might be the one that meets the minimum criterion for that statistic and has the highest efficiency value. Efficiency is measured by the ratio of POD with respect to Impacted Area.

3.1.2 Seasonal Comparison of IFR AIRMETs

Verification results from RTVS are used to compare the quality of IFR AIRMETs during a summer (May – October 1997) and winter season (November 1997 – April 1998). Time series plots of average monthly POD and POD*no* for the summer and winter seasons are shown in Figs. 1 and 2. Overall, the POD values during the summer (Fig. 1) are lower than those computed for the winter season (Fig. 2). For example, POD values range between 0.59 – 0.69 for the summer case (Fig. 1) with the lowest values in June and July. During the winter, POD values range from 0.69 – 0.83, almost 20% higher for the winter (Fig.2) than the summer with the highest values occurring in January.

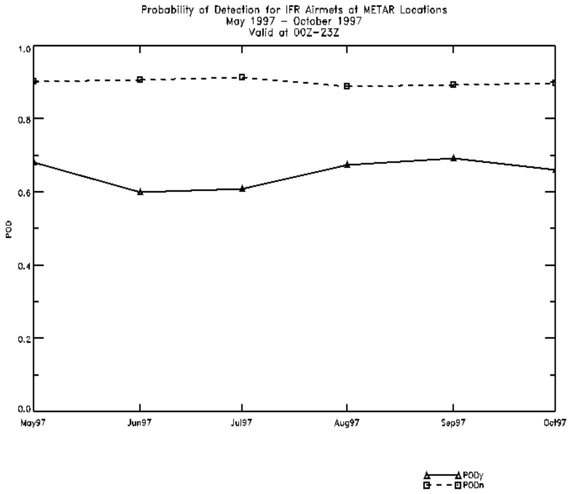

Figure 1. Time-series plot of monthly averages for POD and POD*no* for a summer case from May – October 1997. Legend: Solid line is POD and dotted line is POD*no*. POD values are represented as a percentage.

The POD*no* values indicate the opposite, with forecasts in the summer (Fig. 1) generally better in identifying areas of *no* IFR conditions than in the winter (Fig. 2). These results suggest that IFR weather conditions are more difficult to forecast in the summer than in the winter possibly due to the localized nature of the phenomena in the summer.

Scatterplots of hourly POD and percent of area are shown in Figs. 3 and 4. For the winter case (Fig. 4), the points cluster closer to the upper left-hand corner (i.e. high POD and low area) than for the summer case (Fig. 3). These results suggest that along with higher PODs in

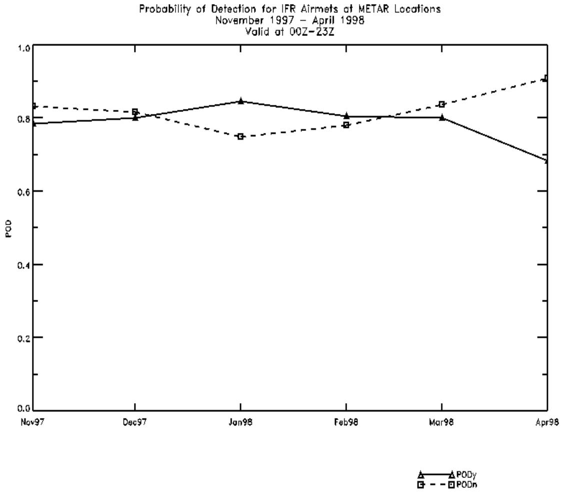

Figure 2. Same as Fig. 1, except for winter case from November 1997 – April 1998.

the winter, the area over which the conditions are forecast to occur in the winter is larger than during the summer. In addition, these areas are also more variable in the winter than summer as shown by the orientation of the regression line on the scatter plots. During the summer (Fig. 3), the percent of area ranges between 5 – 20%, which is less than that covered during the winter season where the area ranges between 5 – 40%. This variance is likely due to differences in the predictability of the basic weather phenomena between summer and winter, which leads to more variable areas in the winter.

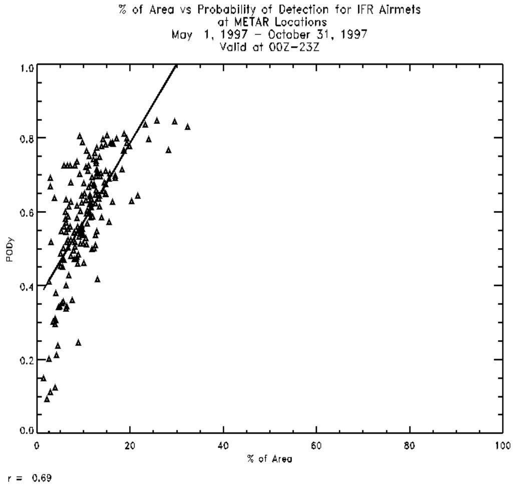

Figure 3. Scatter plot of POD and Impacted Area for the summer case from May – October 1997. Legend: Triangles represent hourly values of POD and Impacted Area. Impacted Area is represented as a percentage of the total area.

3.2 Product Evaluation

In support of AWC's attempt to develop products that are user-friendly and relevant to the broader user base, statistics generated by RTVS were used to evaluate the quality of AWC experimental convective SIGMET outlooks (Hartsough et al. 1999) and the convective SIGMETs. Only examples from the SIGMET outlook evaluation are presented here.

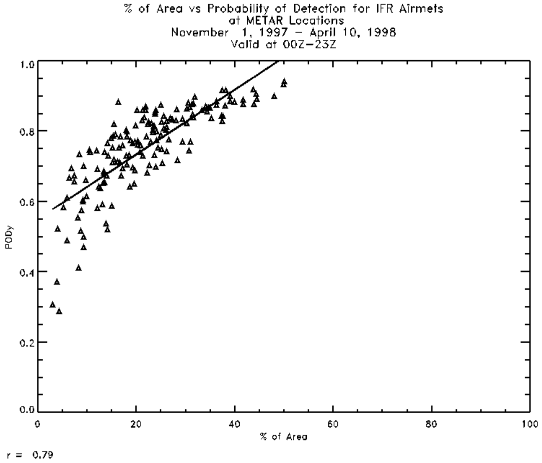

Figure 4. Same as Fig. 3, except for winter case from November 1997 – April 1998.

Experimental convective SIGMET outlooks were generated for 2-, 4-, 6-, and 9-h forecast periods during a joint AWC/Northwest Airlines convective SIGMET outlook evaluation conducted from 6 June – 14 August 1998.

Developing verification methods for the SIGMET outlooks involved investigating the product's purpose; for example, whether it was a forecast for convection or for convective SIGMETs. After thorough discussions with AWC staff, the outlook forecast was defined as a forecast for developing convective SIGMETs.

Convective SIGMETs were used to evaluate the quality of the outlooks as shown in Fig. 5. A 40-km grid was laid over the outlooks and SIGMETs and each grid box was inspected to determine whether it fell inside the outlook, SIGMET, both, or neither. The forecast/observation pairs were generated from that information, and verification techniques were developed using the suggestions stated in Section 2.

Overall results from the evaluation of the standard, 2-, 4-, 6-, and 9-h forecasts for POD and area are shown in the scatterplot in Fig. 6. The 2-h experimental forecast produced the highest efficiency, in terms of POD and Impacted Area. Moreover, the experimental 4-h forecasts often had comparable POD values to the 2-h forecast, but produced higher areas. The areas for the 9-h forecasts were the smallest with POD values generally below 0.3. Some exceptions for the 9-h forecast were noted, however, in a few cases where PODs were greater than 0.6 and Impacted Areas were less than 20%.

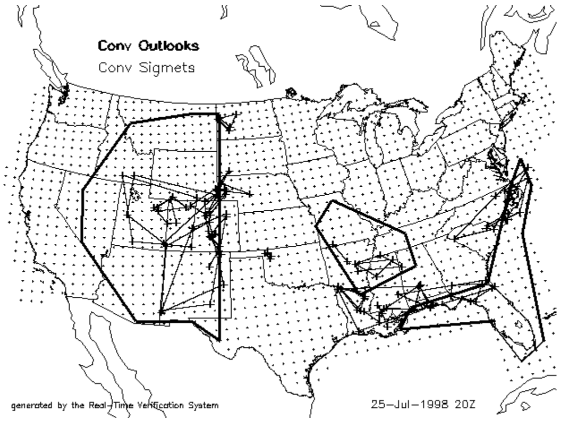

Figure 5. Map of convective SIGMET outlooks, convective SIGMETs, and 40-km grid. Legend: Bold solid lines are convective SIGMET outlooks, light solid lines are convective SIGMETs, and dots represent the 40-km grid.

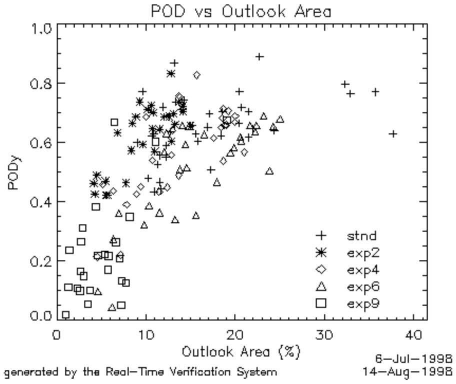

Figure 6. Scatterplot of POD and Impacted Area for experimental convective SIGMET outlooks. Legend: Plus sign, standard outlooks; star, 2-h; diamond, 4-h; triangle, 6-h; and square, 9-h experimental forecasts. Impacted area is a percentage of total area**.**

Evidently, more skill is shown in producing to produce smaller areas with a shorter rather than longer forecast length, possibly influenced by the short forecast length provided by model-based guidance produces. See Hartsough et al. (1999) for further details regarding this evaluation.

4. SUMMARY

Consistent with the goals of the FAA WRP to generate more accurate and accessible weather forecasts and the focus of NWS AWC to improve performance and develop user-friendly products, the benefit of the RTVS in an operational environment is being realized. First, the system is providing a mechanism for managers and forecasters to evaluate the quality of the current products, as demonstrated with the IFR AIRMETs. Once a statistical baseline is developed for AWC products, improvements to those products can be tracked. Second, the system has been a catalyst for product change. For instance, during the process of developing verification methods for the products, deep investigation into the product's characteristics and purpose has taken place, as shown for the convective SIGMET outlooks. Third, the system has provided tools used to support the generation of experimental products. This objective feed back is important in improving the products to better meet the needs of the broader user community.

ACKNOWLEDGMENTS

The author would like to provide a special thanks to the AWC forecasters, staff, and management for their contribution to the RTVS and the verification project. The author would also like to thank Dennis Rogers and Nita Fullerton of FSL for their helpful reviews. This research is in response to requirements and funding by the Federal Aviation Administration. The views expressed are those of the author and do not necessarily represent the official policy or position of the FAA.

REFERENCES

Brown, B. G., G. Thompson, R. T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of In-Flight Icing Algorithms. Part II: Statistical Verification Results. *Wea. Forecasting*, **12**, 890-914.

Hartsough, C. S., J. L. Mahoney, J. K. Henderson, 1999: Verification of the Aviation Weather Center's Convective SIGMET Outlooks using RTVS. *8th Conference on Aviation, Range, and Aerospace Meteorology,* Dallas Texas, (this volume), Amer. Meteor. Soc.,

Kelly, J. J. Jr., 1998: NWS Role in Tornado Warnings. Remarks during the National Tornado Forum. Hosted by the Federal Emergency Management Agency, Washington, D.C.

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

Mahoney, J. L., B. G. Brown, D. Mathews, and F. Mosher, 1998: Verification of the Aviation Weather Center's In-Flight Aviation Weather Advisories: The Methods, Complexities, and Limitations. *14th Conference on Probability and Statistics in the Atmospheric Sciences*, Phoenix, Arizona, Amer. Meteor. Soc., J28-J32.

Murphy, A. H., 1993: What is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Wea. Forecasting*, **8**, 281-293.

__________, and R. L. Winkler, 1987: A General Framework for Forecast Verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

Rodenhuis, D. R., J. H. Henderson, M. D. Mathews, F. R. Mosher, 1997: Interactive Information Processing at the Aviation Weather Center. *7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, CA, Amer. Meteor. Soc., J3-J8.

Sankey, D., K. M. Leonard, W. Fellner, D. J. Pace, K. L. Van Sickle, 1997: Strategy and Direction of the Federal Aviation Administration's Aviation Weather Research Program: *7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, CA, Amer. Meteor. Soc., 7-10.

Schwartz, B., 1996: The Quantitative Use of PIREPs in Developing Aviation Weather Guidance Products. *Wea. Forecasting*, **11,** 372-384.