J2.5 VERIFICATION OF THE AVIATION WEATHER CENTER'S IN-FLIGHT AVIATION WEATHER ADVISORIES: THE METHODS, COMPLEXITIES, AND LIMITATIONS

Jennifer Luppens Mahoney[+], Judy K. Henderson[2], Barbara G. Brown*, Douglas Mathews[1], and Frederick Mosher[1]
NOAA Forecast Systems Laboratory, Boulder, CO

1. INTRODUCTION

An evaluation conducted by the Forecast Systems Laboratory and supported by the Federal Aviation Administration's Aviation Weather Research Program identified the need for verification of the in-flight aviation weather advisories produced by the Aviation Weather Center (AWC) of the National Weather Service (NWS).

Developing appropriate verification methods for the operational forecasting environment, particularly for the AWC, involves many limitations and complexities. These difficulties stem from the fact that AWC's aviation advisories must follow specific guidelines and formats. The advisories are textual products that need to be accurately decoded, and only nonstandard observations are available to verify many of these advisories. In addition to the limits imposed on the products, in some cases the basic nature of a particular product makes it difficult to verify. For instance, forecasters often issue or amend products in response to observations provided by the aviation community. Many of these same observations are then used to verify the advisory. Accordingly, verification methods must be developed so that the statistics, the forecasts, and the forecasters remain unbiased.

This paper briefly presents the statistical approaches that have been developed to verify the AWC's icing and turbulence advisories, as well as the Instrument Flight Rules (IFR) advisories, and describes the limitations and complexities encountered while developing these methods.

*.Research Applications Program. National Center for Atmospheric Research, Boulder, CO.

[1].National Weather Service's Aviation Weather Center, Kansas City, MO.

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

[2]. Collaboration with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

2. STATISTICAL APPROACH

In considering the quality of aviation weather advisories, two facets of the advisories are of interest: 1) the use and value of the information and 2) the scientific validity of the information. Data needed to evaluate the use of the information is difficult to obtain and to quantify. For instance, information on the product usage and on user actions and outcomes would be required to evaluate the value of the information. The focus of this work is the scientific verification of the advisories. This includes evaluating how well the advisory captures the weather it is forecasting, and it should not be interpreted to represent the usefulness or economic value of the advisory.

The most important goals for scientific verification are to apply methods that are objective and unbiased, and fairly represent the forecasts and the forecasting situation. However, for the aviation advisories, developing statistical approaches that fairly represent the forecast product, the forecaster, and the user was often difficult. To obtain methods that are as impartial and unbiased as possible, we chose to develop verification methods for the aviation advisories that follow, as much as possible, the verification framework developed by Murphy and Winkler (1987). This framework encompasses the characteristics of the forecasts, corresponding observations, and their relationship.

3. VERIFICATION METHODS

3.1. Aviation Weather Advisories

The AWC is an NWS operational aviation forecast office; hence, the advisories issued by the AWC must follow specified NWS guidelines and formats. For example, advisories are only issued for affected areas 3,000 square miles and greater. This stipulation may inhibit forecasters from issuing an Airman's Meteorological (AIRMETs) advisory for regions smaller than 3,000 square miles experiencing dangerous weather conditions.

The advisories are textual products that must be accurately decoded into usable data for verification. During the decoding process, much of the free-formatted guidance included in the advisory message, such as the slope of the freezing level and how it moves over time, is often lost. In some cases, attempts are made to use the free-formatted information by simplifying it (e.g., averaging).

Two forms of advisories are issued by the AWC: AIRMETs and significant meteorological advisories (SIGMETs). The verification methods described in the following sections were developed by simplifying many of their product limitations.

3.1.1. AIRMETs

AIRMETs are issued at four standard times each day and are valid for 6 h. AIRMETs are issued when the following conditions occur or are expected to occur and affect an area of at least 3,000 square miles:

- Moderate icing
- Moderate turbulence
- Ceilings less than 1,000 feet and/or visibility less than 3 miles affecting over 50% of an area at any time.

The textual AIRMET describes an outline of a forecast region using line segments that connect a series of Location Identifiers. Figure 1 shows an example of the forecast regions for icing outlined by the AIRMET segments. The weather content frequently includes complex altitude information describing the projected volume, such as a base or top that slopes from one part of the AIRMET to another and is not easily displayed or decoded for statistical computations because it is described in the text body of the AIRMET bulletin. This information is often simplified into an area or volume that remains constant over time.

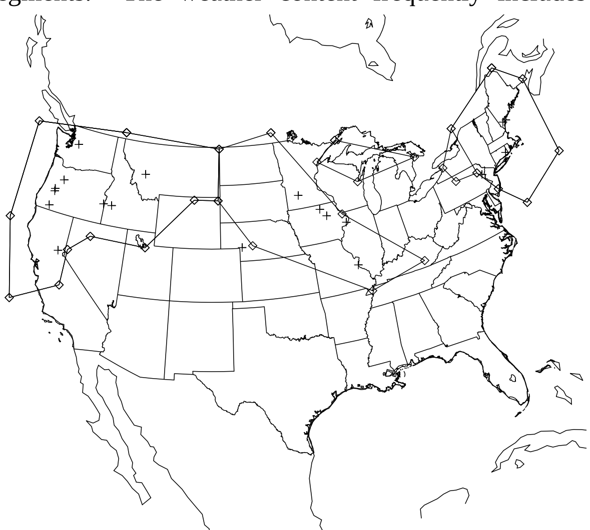

Fig. 1. Display of AIRMET line segments for icing conditions for 19 November 1996 with 1700 to 1859 UTC PIREPs. The "+" represents the PIREPs locations.

AIRMET amendments are issued as necessary to describe weather conditions not originally forecast or the cessation of conditions meeting the AIRMET criteria. In theory, the amended forecast supersedes the original or a previously issued amendment. However, users often base flight routes on previous AIRMET forecasts (i.e., the original or previously amended AIRMET) if only one of the previously issued AIRMETs is available at the time of their flight planning. Therefore, verification techniques were developed for the amendments, as well as, the standard AIRMETs, because of this manner in which they are used.

3.1.2. SIGMETs

The SIGMETs are aviation weather advisories that are often issued on the basis of an observation. They are issued hourly and are valid for up to 2 h. Each 1-h issuance supersedes and cancels the remainder of the previous advisory.

A SIGMET outlook is appended to the hourly bulletin and is valid from 2 to 6 h after its time of issuance.

3.2. Observations

Observations currently used to verify the icing and turbulence advisories are the pilot reports (PIREPs); observations of ceiling and visibility are used to verify IFR AIRMETs.

3.2.1. PIREPs

Numerous problems with using PIREPs for verification of icing and turbulence products have been identified and documented by Schwartz (1996), Kelsch and Wharton (1996), and Brown et al. (1997). For example, negative PIREPs are limited in frequency since pilots are required to report the existence of icing and turbulence conditions (*Yes* report), but not necessarily their absence. In addition, 1) the distribution of icing or turbulence reports is more indicative of air traffic routes than the true distribution of the weather phenomena; 2) the icing or turbulence reports are subjective and often related to the size of aircraft encountering the phenomena; and 3) severe events are undersampled since aircraft avoid areas of moderate and extreme weather events once the location of the weather is identified by other pilots.

Despite these problems, PIREPs remain the best data currently available for verifying icing and turbulence forecasts, and they are used to verify the icing and turbulence AIRMETs and SIGMETs. Both *"Yes"* and *"No"* PIREPs are used. The majority of the PIREPs are reported between the hours of 1200 and 0200 UTC (Brown et al., 1997). All PIREPs that report any turbulence or icing severity (e.g., light reports and greater) are included in the verification. Negative reports, where the pilot directly reports "No Icing" or "No Turbulence" are infrequent, but are included in the verification process.

3.2.2. Surface Observations

Surface observations of ceiling and visibility are used to verify the IFR AIRMETs. Observations taken 45 minutes before the forecast valid time to 45 minutes after the forecast valid time are used to verify that forecast. If a single station reports a ceiling and visibility observation more than once during the observation hour, a check is made to determine if any of those reports meet the AIRMET criteria. If several reports at one station meet the criteria, only one report from that station during the observation hour is added to the verification dataset.

3.3. Statistical Measures

Following Brown et al. (1997), Kelsch and Wharton (1996), and Mahoney et al. (1997), the preliminary verification measures used to verify AWC aviation weather advisories are POD*yes* and POD*no*. The probability of detection of *Yes* PIREPs (POD*yes*) is defined as the number of correct forecasts of *Yes* PIREPs with respect to the total number of *Yes* PIREPs. POD*ye*s ranges from 0 to 1. For more information refer to Brown et al. (1997). Due to characteristics of the PIREPs, it is not possible to compute the False Alarm Ratio (FAR) for icing and turbulence forecasts (Brown et al., 1997). However, FAR, a measure of overforecasting, can be computed for the IFR advisories. Thus, the *"Impacted Area"* and *"Impacted Volume"* are computed for icing and turbulence. The Impacted Area represents the total area encompassed by an AIRMET forecast, while the Impacted Volume represents the total volume. These methods provide a surrogate measure of overforecasting. The goal is to minimize the Impacted Area while maintaining a high detection rate.

3.3.1. Forecast/Observation Pairs

Observations (i.e. PIREPs and surface observations) are matched to the AIRMET forecasts for verification. This process includes, first, determining whether an observation falls within the temporal and spatial constraints of the advisory, and second, evaluating whether the observation report includes the necessary weather conditions. If the observation is *"Yes"* and falls 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 *No-Yes* pair is recorded, and so on until counts of all four possible combinations (*Yes-Yes*, *Yes-No, No-Yes, No-No)* for the pairs are obtained. These pairs are required to compute the statistical measures of POD*yes* and POD*no* for icing, turbulence, and IFR AIRMETs, as well as, FAR for IFR AIRMETs. All advisories that fall within the specified valid times are used to create the pairs.

3.3.2. Impacted Area and Volume

The Impacted Area is computed for specified valid times by projecting a grid onto all areas of the advisory valid at that time and summing the areas of the grid boxes that fall within those horizontal boundaries. The volume is computed by multiplying the total areas by the vertical extent of the AIRMET boundary.

3.4. Applications

3.4.1. Application to Icing and Turbulence AIRMETs

For verification of icing and turbulence AIRMETs, the forecast/observation pairs are computed using PIREPs for the period 1-h before and 1-h after a specified valid time. This approach allows verification of the AIRMETs at any specified time (e.g., for comparison to model-based forecasts) and provides a way of incorporating AIRMET amendments into the evaluation. For example, icing AIRMETs valid from 1500 to 2100 UTC on 19 November 1996 are plotted along with the PIREPs reported between 1700 and 1859 UTC (Fig. 1). The forecast/observation pairs are accumulated over a period of time by comparing PIREPs with all AIRMETs valid at the PIREP time. Each PIREP is tested to determine if its location falls within any of the AIRMET boundaries. For instance, the PIREP located over southwestern Montana resides within an AIRMET boundary and reports icing conditions. As a result, the forecast/observation pair is a *Yes-Yes* match. The pairs are obtained from the perspective of the PIREP (as opposed to the AIRMET). In other words, each PIREP is tested against an AIRMET, as opposed to each AIRMET being verified by a single PIREP.

This approach of using the PIREPs to drive the verification of these advisories complements that used by forecasters who monitor and amend their forecasts. For instance, in an operational environment, forecasters use displays of AIRMETs overlaid with PIREPs to make judgments of where and when to amend their forecasts. Often the guidance for making these decisions is only provided by the PIREPs, particularly in the clear air turbulence cases, where visual observations and model-based guidance are absent.

An example, a height series plot of POD*yes* and POD*no* for icing AIRMETs valid at 1800 UTC from September 1996 - September 1997 is shown in Fig. 2. In this case, POD*yes* and POD*no* values were computed at each 3,000 ft level from all forecast/observation pairs where PIREPs occurred between 1700 UTC and 1859 UTC for the entire 13-month period.

The forecast/observation pairs can be accumulated to compute statistics over any specified time period (e.g. day, month, year). For example, POD*yes* values computed for each month from September 1996 - September 1997, are shown in the time series plot in Fig. 3. In this instance, POD*yes* values were computed, using all flight levels from the forecast/observation pairs, and are combined to form monthly values.

3.4.2. Application to IFR AIRMETs

A somewhat different method is used to compute the forecast/observation pairs for the IFR AIRMETs. In particular, because the surface observations are more nearly systematic, compared to the PIREPs, the IFR AIRMETs can be verified more completely and directly.

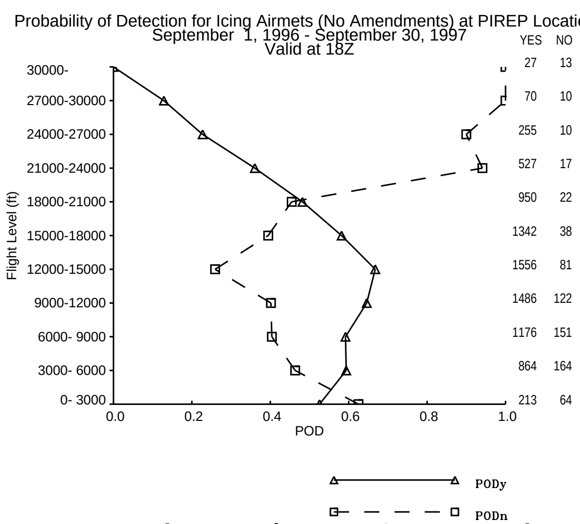

Fig. 2. Height series for icing AIRMETs without amendments valid at 1800 UTC for September 1996 - September 1997. Flight levels are in ft; number of yes and no PIREPS is listed along the right side of Fig.; solid line is POD*y* and dashed is POD*n*.

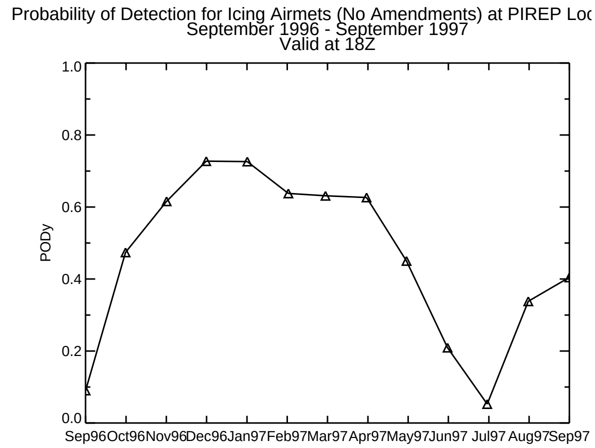

Fig. 3. Time series of POD*y* for icing AIRMETs without amendments valid at 1800 UTC for September 1996 - September 1997.

For these advisories a grid is projected over the whole country. Surface observations within each 3,000 ft grid box are checked for reports that meet IFR criteria (i.e., ceilings less than 1,000 ft and/or visibility less than or equal to 3 miles). The 3,000 ft grid size was chosen to reflect the area criteria defined by the NWS-specified guidelines for issuing AIRMETs. If at least 1 report in the box meets the IFR criteria and is within the AIRMET boundary, then it is recorded as a *Yes*-*Yes* pair, and so on until each grid box is checked. The forecast/observation pairs (based on the grid boxes as opposed to observation sights) are tallied and used to compute the statistics (e.g., POD*yes*, POD*no*, FAR).

3.4.3. Application to Icing and Turbulence SIGMETs

The methods for collecting the forecast/observation pairs and the statistics for the icing and turbulence SIGMETs are consistent with those used for the AIRMETs. However, in this case, only PIREPs reporting in convective-free regions are used to verify the SIGMETs. This distinction is made because the convective regions generally should be identified by a convective SIGMET rather than an icing or turbulence SIGMET.

Due to the nature of the forecast length of the nonconvective SIGMET, forecast/observation pairs are computed for 1-h and 2-h forecast periods. We assume that each SIGMET is a short-term forecast or nowcast, not an observation, although it may be often issued in response to observations. For example, forecasters do project the location and intensity of the weather as they formulate the 1-h advisories.

4. SUMMARY

The statistical approaches developed to verify AWC's icing, turbulence, and IFR AIRMETs, and icing and turbulence SIGMETs have been presented. Developing appropriate verification methods often involved many complexities and limitations, as summarized below.

- The AIRMET message is simplified to extract the specific information needed to evaluate the advisory. However, much of the specific detail on extent and movement of the advisory is averaged or lost, such as conforming a slanting moving AIRMET into a square volume that remains in place.
- PIREPs, a somewhat flawed observation dataset, are the only observations currently available to evaluate the icing and turbulence advisories. Techniques were developed to compensate for the weaknesses associated with this dataset, such as using Impact Area and Impacted Volume as surrogate measures of overforecasting.
- The verification techniques developed to evaluate these advisories often complemented those used to issue the advisories. One example is how forecasters interpret the quality of their forecasts by visualizing the displays of AIRMETs overlaid with PIREPs compared to what the objective statistics may indicate.
- Multiple verification methods were developed to compensate for the various interpretations of the advisories. Examples include verifying the original AIRMETs and the AIRMETs with the amendments, and verifying the length of the 1 and 2-h forecasts for SIGMETs.

Future work includes developing verification methods for convective SIGMETs and cloud forecasts.

5. ACKNOWLEDGMENTS

This research is funded by the Federal Aviation Administration's Aviation Weather Research Program. The views expressed are those of the authors and do not necessarily represent the official policy or position of the U.S. Government. We thank Randy Bullock for his contribution to the AIRMET "is-inside" methods. We also thank the reviewers for their helpful comments.

6. 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,** in press.

Kelsch, M. and L. Wharton, 1996: Comparing PIREPs with NAWAU turbulence and icing forecasts: Issues and results. *Wea. Forec*asting, **11**, 385 - 390.

Mahoney, J.L., J.K. Henderson, P.A. Miller, L.A. Sherretz, D.M. Rodgers, M. Moore, and D. Sims, 1997: Using the FSL prototype AIV editor to edit icing grids: A preliminary evaluation using the RTVS. NOAA Technical Memorandum ERL FSL-20, Boulder, CO, 14p.

Murphy, A.H. and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Review*, **115**, 1330-1338.

Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting.*, **11**, 372-384.