FA 4.9 STATISTICAL EVALUATION OF EDITED ICING GRIDS USING FSL'S REAL-TIME VERIFICATION SYSTEM (RTVS)

Jennifer Luppens Mahoney[1], Judy K. Henderson[*], Patricia A. Miller, Lynn A. Sherretz,
Dennis M. Rodgers, and Mark Moore[+]
NOAA Forecast Systems Laboratory, Boulder, CO

1. INTRODUCTION

As part of a Federal Aviation Administration (FAA) sponsored project, the Forecast Systems Laboratory (FSL) and the National Weather Service Aviation Weather Center (AWC) conducted a multifaceted exercise involving AWC forecasters editing model-based forecasts of icing potential. The primary goals of the exercise were to:

- Introduce AWC forecasters to the use of three-dimensional gridded objective forecasts of icing potential.
- Provide forecasters the opportunity to view and edit the three-dimensional model-based forecasts using prototype tools and to critique those tools.
- Start the process of collecting and processing edited model-based forecasts of icing, and then begin the statistical evaluation of forecaster changes to these model-based forecasts.

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

[+]. Contract with the Gonzales Consulting Services, Inc., Denver, CO

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

This paper focuses on the third goal of the exercise. The progress made toward real-time verification capability is described and preliminary statistical results are presented.

The exercise successfully laid the groundwork for the longer-range goal of performing ongoing statistical evaluations of the value added by forecasters to model-based forecasts of aviation-impact variables (AIVs). The concept of "adding value" to model guidance products is certainly not a new one. Forecasters have been "adding value" to model-based output since the beginning of numerical weather prediction. Forecasters currently use a combination of observations and model-based data to determine how meteorological fields, such as temperature, moisture, and wind, change over time. During this process, the forecasters can perform "intuitive" adjustments to the model guidance based on observational data, recent performance of that model, and their forecasting experience. However, for this study, the forecasters were asked to apply their adjustments directly to the model-gridded fields by using their knowledge and experience to insert changes in the values at selected grid points. To determine how this was to be done in a quasi-operational mode, the following questions were addressed:

- What model times and forecast periods need editing?
- How much time is required for a forecaster to edit the forecast?

- What data should be available prior to editing?
- What are the "mechanics" of physically editing a model-based grid?

This exercise required that AWC forecasters edit model-based forecasts of icing potential in realtime. The original and edited grid pairs were then statistically evaluated, and the results were displayed in realtime on FSL's Real-Time Verification System (RTVS; (Mahoney et al. 1997, in this volume)). In addition, the numerical output and verification data collected and archived to the RTVS' local storage allows for detailed comparisons of the grid pairs.

2. DATA AND METHODOLOGY

This evaluation was conducted in two phases: 1) the generation of the edited grids and 2) the statistical comparison of the icing grid pairs (i.e., the original and edited icing grids). Following are descriptions of the model and the algorithm, the verifying observations, and the procedures which define the editing process.

2.1 Icing Forecasts

Objective forecasts of icing potential were generated using the Research Applications Program's (of the National Center for Atmospheric Research) in-flight icing algorithm (Thompson et. al. 1996) applied to the Rapid Update Cycle (RUC) model, running at the National Centers for Environmental Prediction (NCEP). Hourly forecasts out to 12 h were produced every 3 h on the CONUS 211 grid with 80-km horizontal grid spacing and 39 vertical levels.

2.2 Verifying Observations

Voice Pilot Reports (PIREPs) with an observation time of 1 h on either side of the model's valid time were used to evaluate the icing grid. For these analyses, PIREPs reporting at least light icing were considered to be *Yes* reports. The PIREPs were interpolated to the model grid location (see Mahoney et al. 1996 for more details).

Although, PIREPs continue to be problematic for verifying icing and turbulence, they were the only verification dataset available for this study (Brown et al 1993, 1996; Schwartz 1996 provide more details on these problems).

2.3 Procedures

During the statistical evaluation period, interaction between AWC and FSL was critical. To handle this, we designed a closely-monitored system of data transfer processes among NCEP, FSL and AWC. These processes required the transfer of the RUC model from NCEP to FSL every 3 h for generation of the icing Aviation Impact Variable (AIV). From FSL, the icing output was transmitted to the AIV Editor, a prototype software system which allows for AWC users to easily easy manipulate model gridded fields (Sherretz et. al. 1996).

Forecasters at AWC were directed to edit the 9-h icing forecasts produced by the 0900, 1200, and 1500 UTC RUC model runs. The model valid times were chosen to correspond to the large number of PIREPs available for verification. The editing took place Monday - Friday from 15 April - 25 May 1996. To produce the edited icing grid, the forecasters used the AIV Editor (Sherretz et al. 1996[1]; FSL and AWC 1996). Forecasters were given satellite imagery and PIREP data, valid 3 h prior to the model validation time, to be used as background information to help them identify areas

1. Detailed information on the AIV Editor can be found at http://www.fsl.noaa.gov/~osborn AIV_Editor_User's _Guide.html

of potential icing. Once the forecasters intuitively assimilated the data, they were instructed to edit each original icing grid, using the AIV Editor, 3 h prior to the model valid time. The forecasters were given the full 3 h to perform their edits. However, in the majority of the cases, the forecasters completed their edits within an hour. These edited icing gridded fields were then transmitted back to FSL for verification on the RTVS.

Twenty-one regular duty AWC forecasters were able to complete at least one forecasting session for this study. The same forecaster could produce 2 or more forecasts on the same day. However, usually 2 or 3 forecasters produced edited icing grids on the same day. A total of 60 grid pairs were collected during this evaluation period.

2.4 Verification Methods

Due to the serious pitfalls associated with the PIREP data (Brown et al 1993, 1996; Schwartz 1996), verification analyses of the original and edited icing grids are limited to the computations of probability of detection (POD) based on *Yes* reports (POD*y*) of icing, PODs based on *No* reports (POD*n*) of icing, grid areas (computed for each model-based forecast by projecting the grid point to the earth's surface and summing the areas associated with those grid points), and grid volumes (computed the same as grid area except that the grid points in a volume with a Yes forecast are summed together). "Perfect" forecasts of icing are characterized by POD=1 and small areas and volumes. Mahoney et. al (1996) describes the methods by which PODs and areas and volumes are computed in the RTVS. For this study, confidence intervals were also computed for the median of the daily POD values, areas and volumes. The methods by which confidence intervals are computed follow work done by Brown et. al (1996).

3. RESULTS

Overall results for the verification of the icing grids are shown in the Table. The statistics indicate that for the 6-week period the overall POD values for the edited icing grid increased slightly when compared to the POD values of the original icing grids. In addition, the average area of the edited grid decreased, while the average volume increased.

The 95% confidence intervals computed for the daily medians are displayed as notched boxplots in Figs. 1-3 (the notch is the horizontal "v" shape extending above and below the medians). If the notched regions for the two medians do not overlap, the differences in the medians are considered to be statistically significant.

TABLE. Overall verification statistics for the original and edited icing grids.

	Original	**Edited**
POD*y*	0.4132	0.4663
POD*n*	0.7471	0.7447
Average Area (Km^2)	2.39×10^6	2.27×10^6
Average Volume (Km^3)	3.68×10^6	4.57×10^6

Boxplots representing the distribution of POD*y* values for the entire period are shown in Fig. 1. As was indicated by the overall results in the Table, the median POD*y* increased slightly for the edited grid. However, the median daily POD value for the edited grid is not significantly different from the median value of the original grid, as indicated by the overlap in the notches. The median daily POD*n* values (not shown) for the original and edited grids were also not statistically significantly different.

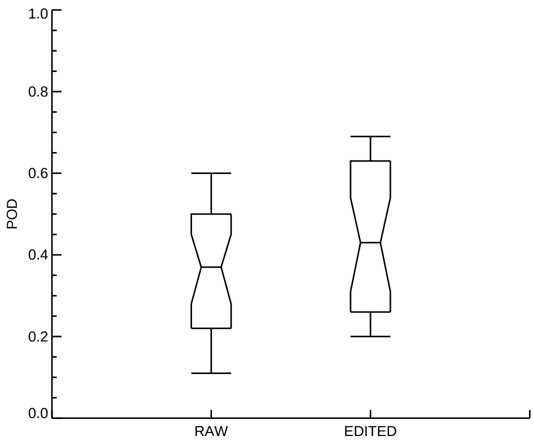

FIGURE 1. Distribution of daily POD*y* values computed for the 6-week period. The notches indicate confidence intervals; the horizontal lines of the boxplot, starting with low numbers and going to high numbers, are the 10th, 25th, 50th (median), 75th, and 90th percentiles.

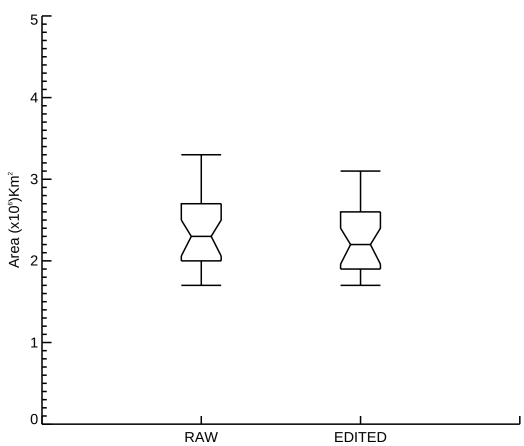

FIGURE 2. Distribution of areas computed from each grid for the 6-week study period. The boxplots are described in Fig. 1.

The areas of the edited icing grids decreased when compared to the original grid, as shown in the Table and by the medians in Fig. 2. However, as indicated by the notches in Fig. 2, this decrease is not statistically significant. In contrast to the decrease in area, a substantial increase in volume was detected in the edited grid when compared to the original icing grid. The results are apparent in both the Table and Fig. 3. However, these findings are also not statistically significant, as shown by the overlapping notches in Fig. 3.

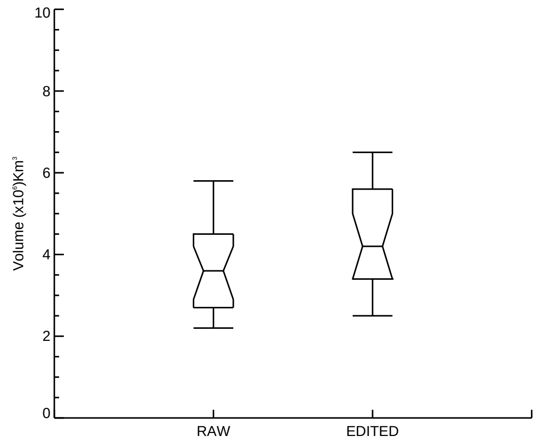

FIGURE 3. Same as Fig. 2 except for volumes.

Figs. 4a and 4b show the POD*y* and POD*n* values of the original and edited icing fields, respectively, plotted as a function of flight level. The edited grid (Fig. 4b) shows a slight increase in the POD*y* between 3,000 ft. and

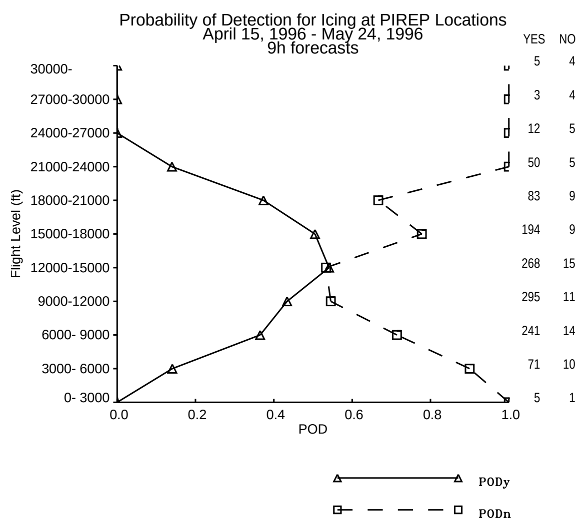

FIGURE 4a. Height series plot of POD*y* and POD*n* values for the original icing grid. The solid line indicates PODy and the dotted line PODn. The number of PIREPS reporting icing and no icing explicitly are listed along the right side of the plot. Flight levels are in ft.

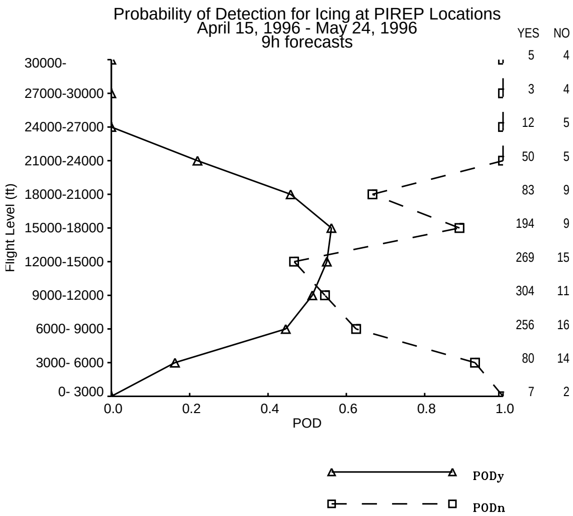

FIGURE 4b. Same as Fig. 4a except for the edited icing grid.

21,000 ft. over the original grid (Fig. 4a). However, analysis of confidence intervals at each height again indicated no statistical significance (not shown).

Figs. 5a and 5b show a time series of POD*y* values computed for each day (POD values were computed by using all available grids for a particular day) over the 6 week period. The disconnects indicate a weekend when no data

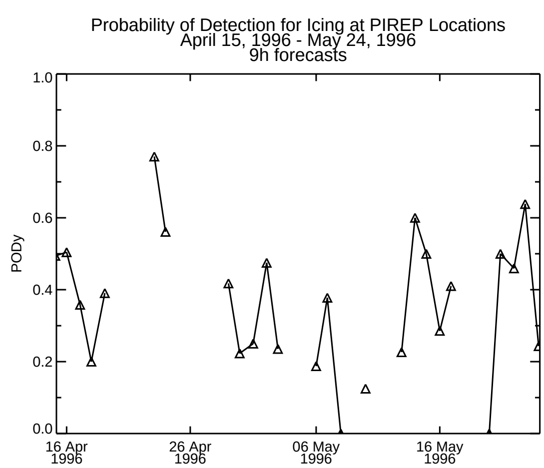

FIGURE 5a. Time series plot of the original icing PODy values.

was collected, an operational day when the data was not available from NCEP, or prob-

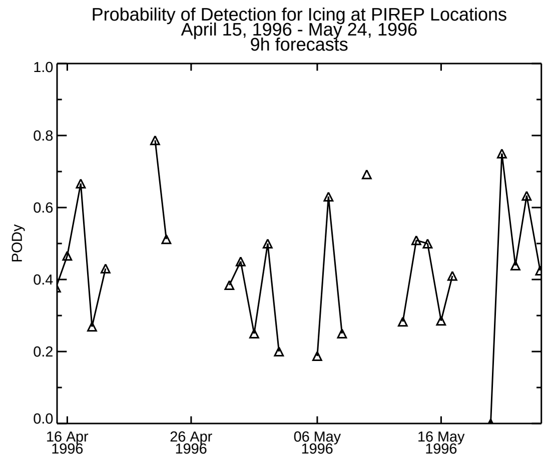

FIGURE 5b. Same as Fig. 5a except for the edited icing grid.

lems with the data transmission to AWC. Close examination of the time series plots indicate that out of the 27 days when edited grids were generated, POD*y* values were higher for the original grid (Fig. 5a) 45% of the time, while 55% of the time the values were greater for the edited grid (Fig. 5b).

4. DISCUSSION

Comparative analysis of the original to the edited icing grid suggests that some value is added when the forecasters edited the grid, as indicated by the increase in overall POD values and decrease in grid areas. However, these results are not statistically significant. A larger sample size would help determine the statistical significance of the results.

The time series of daily POD*y* values suggest little improvement in the icing forecasts when compared each day. However, the increase in POD*y* at heights between 6,000 and 18,000 ft. suggests that forecasters may be particularly able to correctly identify areas of icing at these levels, and adjust the grid accordingly.

The increase in volume of icing potential is probably related to the design of the AIV Editor user interface as documented in the FAA/FSL evaluation on the editor (FAA and FSL 1996). The FAA reported that when editing icing grids, the users found the flight level selector buttons to be tedious. Users suggested providing the option of entering the flight levels via the keyboard.

5. CONCLUSIONS

The statistical findings reported in this paper, coupled with the findings and recommendations of the FAA/FSL evaluation of the AIV editor (FAA and FSL 1996), will help guide the development of the editor and techniques required to utilize model-based forecasts of AIVs in the operational environment of AWC.

The success of the exercise also demonstrates the feasibility of real-time evaluation/verification of operational forecast products from AWC using the RTVS. This objective feedback to operations holds may benefits in the improvement of AWC forecast services and of individual forecaster techniques.

6. ACKNOWLEDGMENTS

The authors would like to thank AWC forecasters for their willing participation in this study, Laura Heigl and Dave Osburn of FSL for helping set up the computer environment needed to fulfill the requirements for this study, Barbara Brown of NCAR/RAP for her insights and contributions to the verification procedures, Randy Bullock and Greg Thompson of NCAR/RAP for their contributions to this study, Ed Tollerud of FSL for his statistical expertise, Barry Schwartz for his work on the icing grid conversions, and to Nita Fullerton for her helpful review.

7. REFERENCES

Brown, B. G., T. L. Fowler, B .C Bernstein, and G.S. Forbes, 1993: Use of pilot reports for verification of aircraft icing diagnoses and forecasts. *Preprints, Fifth International Conference on Aviation Weather Systems*, Vienna, VA, 2-6 August, American Meteorological Society, Boston, MA, 277-281.

Brown, B.G., G. Thompson, R.T. Bruintjes, R. Bullock, and T. Kane, 1996: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. and Forecasting*., in review.

Federal Aviation Administration and Forecast Systems Laboratory (FSL), 1996: Aviation Impact Variable (AIV) Editor Evaluation Report 25 June, 19 pp. (Copy available from Federal Aviation Administration Engineering and Test Division, Weather Branch (ACT-320), Atlantic City International Airport, Atlantic City, NJ, 08405,

FSL and NWS/AWC: 1996. Aviation Impact Variable Editor User's Guide, Version 3.0, 36 pp. (Copy available from Forecast Systems Laboratory, Aviation Division, 325 Broadway, Boulder, CO 80303).

Mahoney, J. L., J. K. Henderson, P. A. Miller, 1996: RTVS: Conceptual design (unpublished). (Copy available from Forecast Systems Laboratory, Aviation Division, 325 Broadway, Boulder, CO, 80303)

Mahoney, J.L., J.K. Henderson, 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*, 2-7 February, Long Beach, CA, Amer. Meteor. Soc., Boston, MA.

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

Sherretz, L., D. Rodgers, R. Olson, 1996: Enabling NCEP aviation forecasters to edit grids generated by the aviation gridded forecast system. *Preprints, 12th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Atlanta, Georgia, 28 January - 2 February, American Meteorological Society, Boston, MA, 252-255.

Thompson, G., R.T. Bruintjes, B.G. Brown, F. Hage, 1996: Intercomparison of in-flight icing algorithms. Part 1: WISP94 real-time icing prediction and evaluation program. *Wea. and Forecasting*., in review.