4.8 A PERFORMANCE EVALUATION OF THE JUNEAU WIND HAZARD ALERT SYSTEM

Tressa L. Fowler*, Jamie T. Braid, and Matthew J. Pocernich
National Center for Atmospheric Research, Boulder, Colorado

1 INTRODUCTION

The Juneau Wind Hazard Alert System (JWHAS) is a prototype system in place at Juneau International Airport (JNU). An upgraded version of this system will form the basis for the Juneau Airport Wind System (JAWS). JWHAS is comprised of uncertified anemometers and profilers at a variety of locations in the Juneau area (See Juneau, 2002 for more information on the equipment and its placement). During fiscal years 2000 and 2003, field projects were conducted at JNU. The University of Wyoming (UW) King Air research aircraft collected turbulence and wind shear measurements during the program. These measurements will be used to improve and verify JWHAS. For the 2003 field project, the UW King Air and an Alaska Airlines 737 (ASA737) submitted Pilots' Reports (PIREPs) of turbulence to the operations director. These PIREPs are also used for verification of the system.

2 DATA

2.1 Hazard Areas

Twelve hazard areas on the approach/departure paths of the Juneau airport have been identified primarily by the Federal Aviation Administration (FAA) staff based on analyses provided by scientists at the National Center for Atmospheric Research (NCAR). The map (Figure 1) below shows the areas in two dimensions. The four boxes nearest the airport are Coghlin island (cog), Outer Point (opt), north Douglas island (ndi), and Lemon Creek (lmn). The four boxes shown in the Gastineau channel (gc) actually represent eight hazard areas (gc1, gc2, gc3, gc4, gca, gcb, gcc, gcd) as each is divided into two vertical regions. The lower areas (gc1 to gc4) cover from the surface to 2,000 feet and the higher areas (gca to gcd) cover 2,000 feet and above.

2.2 Wind Regimes

Three wind regimes are of interest for hazard detection in the Juneau area: southeast, Taku, and mixed. A southeasterly wind regime (SE) has winds aloft from the southwest (due to an approaching low), which are then turned to the southeast as they flow up the Gastineau channel towards the airport. A Taku wind regime (TK) has winds from the north or northeast that are caused by a very strong pressure gradient between glaciers to the north (high pressure) and Juneau (low pressure) in the south. For a true Taku event, the depth of the cold air over the glaciers has to be higher than the mountains in the area. These winds spill over the Salisbury Ridge into the Gastineau Channel.

* *Corresponding author address*: Tressa L. Fowler, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307 e-mail: tressa@ucar.edu

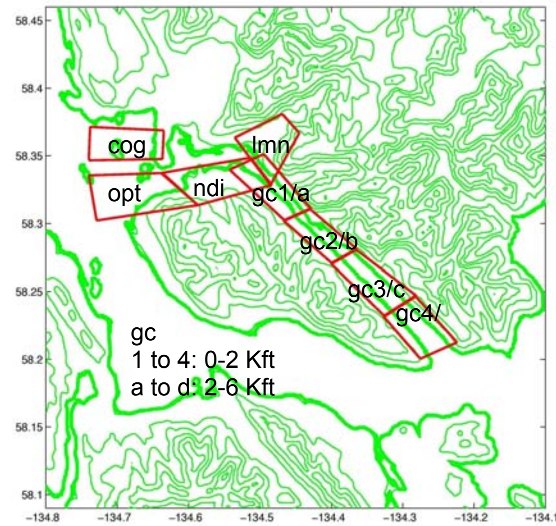

Figure 1: Map showing locations of hazard areas.

A Gap Flow wind regime is a weaker type of Taku event where the shallow pool of cold air to the northeast of Juneau is only deep enough to flow down the creek drainages (Salmon, Gold and Sheep), also known as the mountain "gaps". Mixed flow (MX) is a wind regime that is a combination of the Southeast and Gap Flow wind regimes.

Two other wind regimes, unknown and calm, were planned for consideration. However, the unknown regime had no cases. The calm wind regime had too few cases to provide a basis for any analysis. Further, under the calm wind regime, the JWHAS system produces no warnings.

2.3 Algorithms

JWHAS and the OPS Spec are briefly described here. For a more complete description of the JWHAS, see the algorithm design description (ADD) (Juneau, 2002). The OPS Spec is described by the FAA Operations Specifications for Alaska Airlines Operations in Juneau, AK, C64, Effective 21 April 1998. A table listing the OPS Spec weather criteria is presented in Appendix 1.

Two versions of the JWHAS algorithm are verified. The first is "trained" on a subset of the cases from the field programs and verification is done on the remaining cases. The second algorithm is trained on all cases. This algorithm represents the best the algorithm could do with all of the available information. By comparing these versions of JWHAS, it can be determined how much different the algorithm would be if it included the verification cases.

OPS Spec: The OPS Spec uses wind direction with mountaintop and airport wind speeds to determine a Go/No-Go decision for some aircraft arrivals and departures under Federal Aviation Regulations, Part 121. Two versions of the OPS Spec are analyzed. The first is the version of the algorithm that was created in 1999. The second version, from 2002, is a modification of the original.

JWHAS (A) (Wp/Hydrotech): JWHAS was developed by NCAR/RAP, with funding provided by the FAA. The wind hazard warning algorithms are based on using linear regressions of the Juneau windprofiler and Hydrotech sensor data to estimate turbulence hazard levels (Juneau, 2002). This version of the algorithm is trained on a subset of the available cases and then verified on the remaining cases. Thus, the algorithm is independent of the verification data.

JWHAS (B) Complete (the cheater's version): This version of the JWHAS algorithm is the Wp/Hydrotech JWHAS algorithm "trained" on the complete set of observations from the field program. It is the "cheater's version" because it is verified on the same cases that were used in its development. This will allow us to determine how much difference the inclusion of the verification cases would have made in the JWHAS algorithm. With large enough sample sizes and random assignment of cases to the training and verification sets, this algorithm should not differ significantly from Wp/Hydrotech JWHAS.

2.4 Research Aircraft Observations

Observations for the verification were collected by the University of Wyoming King Air during the field projects in fiscal years 2000 and 2003. The UW King Air collected eddy dissipation rate (EDR) and wind shear measurements during the 2000 field project, from November 26, 1999 through February 14, 2000. For 2003, the UW King Air returned and collected observations of EDR and wind shear from October 14, 2002 to January 20, 2003, with breaks at Thanksgiving and Christmas. Voice PIREPs given by the research aircraft pilots are available from the FY 2003 field experiment from both the UW King Air and the Alaska Airlines 737.

During the FY 2003 field project, the ASA737 also collected numerical data samples with on board instruments that were subsequently processed into EDR measurements. These EDR measurements were intended to be combined with the King Air EDR measurements for regression development and performance verification. Data analysis showed that adding the ASA737 data degraded the overall accuracy of the aircraft EDR measurement set, despite extensive efforts at quality control of the ASA737 data. In order to develop the most accurate hazard detections possible, the JWHAS development team decided to use the King Air EDR measurements alone for regression development. While the ASA737 numerical data is not used, the ASA737 pilots' reports are still used as part of the verification data in this report.

The vertical components of the EDR measurements (ZEDR) taken by the aircraft can be considered a surrogate for observations of turbulence. The values used in this study were smoothed slightly (using running medians of length 3) to reduce the effect of extreme observations on the analysis. Higher values of ZEDR indicate greater amounts of turbulence. The ZEDR measurements are bounded below by zero. Although, in theory, the observations are unbounded above, the highest measurement taken was less than 1, with the great majority of the observations below 0.3. EDR measurements are considered to be a measure of turbulence that is independent of the aircraft experiencing it. Figure 2 shows box plots[*] of the measured ZEDR (i.e. hazard) in each hazard area for each of the three wind regimes. For a southeast wind regime (SE), the channel hazard areas had low measured hazards with a couple of outliers, the most extreme ones (near 0.4) in gcc and gc4.

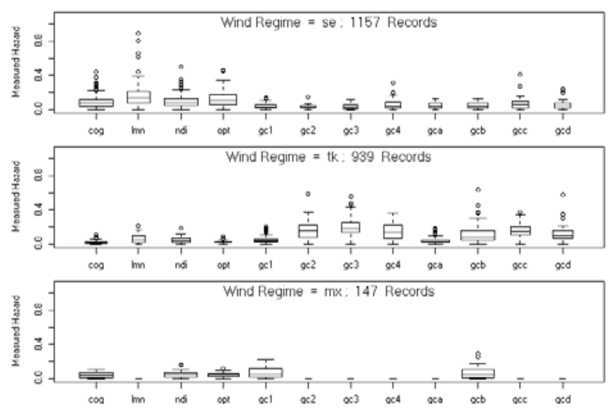

Figure 2: Box plots showing the distribution of aircraft measured hazards (ZEDR) for each hazard area by wind regime.

The four hazard areas nearer the airport, especially lemon creek (lmn) had the most severe events during the SE wind regime, several of them above 0.4. During Taku events (TK), the opposite was true. The hazard areas nearer the airport (cog, lmn, ndi, opt, gc1 and gca) had less extreme measured hazards than did the six channel boxes farthest from the airport (gc2, gc3, gc4, gcb, gcc and gcd). When the wind regime was mixed, the measured turbulence hazards in all areas

[*] Box plots show the distribution of values. The line at the center of each box is the median while the top and bottom of the box represent the 75^{th} and 25^{th} percentiles, respectively. Thus, the box shows the range of the center half of the data. The whiskers extend to the maximum and minimum values that are not outliers, each showing the range of the top and bottom quarters of the data. The dots above or below the whiskers are outliers.

were generally quite low, nearly all below 0.2. Seven of the boxes (lmn, gc2, gc3, gc4, gca, gcc, and gcd) had no observations for the mixed wind regime.

3 METHODS

3.1 Statistics

A forecast verification methodology outlined by Brown et al. (1997) treats the forecasts and observations as Yes/No values. This method can be extended to forecasts with values on a continuous scale using the approach outlined in Brown et al. (1999). In particular, forecasts with continuous output are converted to a set of Yes/No forecasts by application of a variety of thresholds. The thresholds have been chosen to span the range of values of the nowcasts and observations. Additionally, enough thresholds are used to give dense coverage over the range of nowcast skill. Further, threshold selection is somewhat arbitrary. The actual value of the threshold is not really meaningful, especially if bias is present. For instance, application of a threshold of 0.20 to JWHAS nowcasts would lead to a Yes nowcast for all JWHAS values greater than or equal to 0.20. The OPS Spec values are already binary (Go/No-Go), so use of thresholds is unnecessary. The continuous EDR measurements are also converted to binary observations by use of thresholds. The verification methods are based on the Yes/No two-by-two contingency table (Table 1), where the nowcasts are represented by the rows and the observations are represented by the columns.

The nowcast/observation pairs are divided up into the four cells shown in Table 1. A minimum of five observations per cell is required for estimating probabilities from these cells (Wilks, 1995; Mood *et al*, 1974). Thus, a minimum of 20 cases is required for this type of analysis.

Table 1: Basic contingency table for evaluation of dichotomous (e.g., Yes/No) nowcasts. Elements in the cells are the counts of nowcast-observation pairs.

Nowcast	***Observation***		***Total***
	Yes	***No***	
Yes	YY	YN	YY+YN
No	NY	NN	NY+NN
Total	YY+NY	YN+NN	YY+YN+NY+NN

Table 2 lists the verification statistics that are used in the JWHAS evaluation. PODy and PODn are the primary verification statistics. PODy and PODn are estimates of the proportion of *Yes* observations that are correctly forecast and *No* observations that are correctly forecast, respectively (Brown et al. 1999; Brown et al. 1997). The True Skill Statistic (TSS) (Doswell et al. 1990) is a measure of the ability of the forecasts to discriminate between *Yes* and *No* observations, and is also known as Hanssen-Kuipers discrimination statistic (Wilks 1995). The False Alarm Ratio (FAR) can be used to assess over warning. The FAR is the proportion of forecast hazards that failed to occur in the observed data. The % of forecast hazards is the proportion of events that a system calls a hazard. This measure is independent of the observation.

4 RESULTS

4.1 Overall Results

Overall verification statistics combining cases from all hazard areas and wind regimes using a threshold of 0.1 for the JWHAS and aircraft values are presented in Table 3. Clearly, the JWHAS system (i.e. both versions of the algorithm) performs better in terms of correctly detecting both events and non-events than either version of the OPS Spec. In fact, it correctly classifies over twice as many events and non-events as the OPS Spec. It also has less than half the false alarms and much greater skill. In fact, the OPS Spec shows negative skill. The statistics for JWHAS (A) are probably a better estimate of system performance than those for JWHAS (B) since the former are verified on independent cases.

Table 2: Overall verification statistics for JWHAS (A), JWHAS(B), OPS Spec 99, and OPS Spec 02 using a 0.1 threshold.

	PODy	PODn	FAR	TSS
JWHAS (A)	0.772	0.883	0.292	0.655
JWHAS(B)	0.832	0.881	0.241	0.713
OPS Spec 99	0.353	0.353	0.793	-0.294
OPS Spec 02	0.37	0.34	0.788	-0.291

For the JWHAS B data, errors between the nowcast EDR value and the observed ZEDR value are generally quite small. In fact, only 1% of the errors have a mangitude of greater than 0.17 and only 5% have magnitude greater than 0.9. Fully 75% of the errors are below 0.035. Figure 4 shows box plots of these errors. A similar figure for JWHAS (A) was produced, and the results are quite similar. However, since only the non-training set cases are included for JWHAS (A), some of the outliers, i.e. large errors, do not appear.

Generally, larger errors occur in conjunction with larger ZEDR values, and vice versa. For the MX wind regime, the errors are all quite small, as were the aircraft ZEDR measurements. In a SE wind regime, large errors are in the cog, lmn, ndi, and opt boxes, where the larger hazards were measured, while the remaining boxes have smaller errors. During TK winds, the larger errors and hazards are in the channel boxes along while the boxes nearer the airport had smaller errors and hazards. The center of each box is near 0, indicating that the typical (i.e. median) value of the errors is near 0. Thus, the nowcasts appear to be relatively unbiased.

Generally, the JWHAS system issues high EDR values for the cases with high EDR values as measured by the aircraft. However, JWHAS still tends to underestimate the actual EDR value. Thus, though the JWHAS system indicates a value high enough to warn of a turbulence hazard, the actual numerical forecast still may have a fairly large error. If a threshold is applied to the JWHAS values to create a categorical alarm warning system, then this underestimation should not affect the final system.

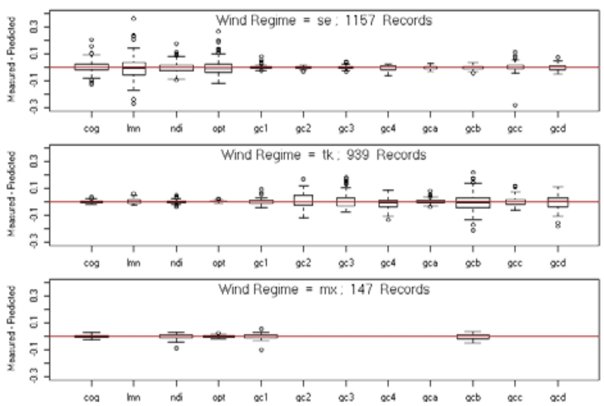

Figure 3: Box plots showing error (measured - predicted) in EDR for JWHAS (B).

4.2 Percent of Nowcast Hazards – An assessment of overwarning

The percent of nowcast hazards can be thought of as a sample climatology of alarm frequency. While use of JWHAS (A) is recommended for assessment of system performance via verification statistics, it would be inappropriate to use it for a sample climatology because of the resampling of cases. For JWHAS (B), each case is used exactly once. For JWHAS (A) some cases are not used at all while others are included multiple times.

The OPS Spec (both 1999 and 2002) hazards are the “NO GO” decisions. For the JWHAS (B), a hazard is any JWHAS ZEDR value of 0.2 or greater. This threshold was selected since it represents a typical (i.e. approximately median) value of JWHAS nowcasts when moderate or greater turbulence is observed. In the Coghlin Island hazard area, the OPS Spec does not produce a go/no-go decision, so these table cells are empty.

Accumulated across all wind regimes and hazard areas, the JWHAS (B) nowcasts a turbulence hazard with an EDR value of 0.2 or greater in almost 9% of the cases. The OPS Spec (99) gives a “NO GO” in 57% of cases while the OPS Spec (02) gives “NO GO” in 55% of cases. In mixed and southeast wind conditions for all hazard areas, the JWHAS system nowcasts fewer hazards than either version of the OPS Spec. During Taku wind conditions, the JWHAS nowcasts no turbulence hazards in the four areas near the airport, but nowcasts more hazards in the channel areas gc2, gc3, gc4, gcc and gcd than the OPS Spec. Thus, the OPS Spec is, on average, much more conservative than the JWHAS in that it restricts use of the airspace over five times as often. This is also evidenced in the high rate of false alarms for both versions of the OPS Spec.

4.3 Verification with PIREPs from the research aircraft.

Select King Air and ASA737 PIREPs from the 2003 field program are compared to the nowcast turbulence hazard values from JWHAS (B) and to the ZEDR measurements from the UW King Air. Many of the flights passed through the same airspace within short time intervals. Since atmospheric conditions tend to persist over short lengths of time, the observations in the same airspace at times close together are not independent. To mitigate the non-independence in the data and to make the analysis of the PIREPs manageable, a 30 minute time window for PIREPs was used. Thus, observations of the same type of event (e.g. no turbulence) in the same hazard area within the 30 minute window were considered to be a single report. PIREPs of different conditions in the same area within a 30 minute time window were considered separate reports as were PIREPs more than 30 minutes apart. Then, a time window of ±5 minutes was used to match the resulting PIREPs to aircraft observations and JWHAS nowcasts. Only PIREPs of “no turbulence”, “moderate turbulence”, or “severe turbulence” were considered. PIREPs of “light turbulence” were ignored for two reasons. First, there were a lot of them and the matching of PIREPs to JWHAS values had to be done by a person rather than a computer. Secondly, differentiating between light turbulence and no turbulence or between light turbulence and moderate or greater turbulence is simultaneously more difficult and less important than differentiating between no turbulence and moderate or greater turbulence.

The PIREPs were collected in all twelve hazard areas and all three wind regimes. However, due to the sample size of the PIREP data, dividing the analyses by hazard area and wind regime (36 possible combinations) is not feasible. The resulting number of cases would be too small in many of the hazard area/wind regime combinations. Thus, this data is used collectively to provide an overall verification of the JWHAS system. Table 7 displays the counts of “No”, “Moderate”, and “Severe” events reported by each aircraft and used in this analysis.

Table 3: Counts of turbulence PIREPs by severity and aircraft type.

	King Air	ASA737	Total
No	210	154	364

Moderate	100	30	130
Severe	2	0	2

The JWHAS system and the aircraft measurements both show an uncanny ability to discriminate between PIREPs of no turbulence and PIREPs of moderate or severe turbulence, as illustrated by Figures 4 and 5. These figures are known as discrimination plots. They show the distribution of the JWHAS nowcasts or aircraft measurements that are paired with PIREPs of “no turbulence” or “moderate or severe turbulence”. If these distributions had a large overlap, then the nowcasts or measurements would not “discriminate” between no-turbulence events and moderate–or-severe turbulence events. However, the distributions of the EDR values from both the JWHAS and the aircraft for no turbulence and for moderate or severe turbulence overlap only slightly. Separate plots were produced for the PIREPs from the King Air and 737, respectively. However, these plots were very similar to each other and to Figures 4 and 5. Thus, only the plots with the combined PIREPs are presented here.

The distribution of JWHAS nowcasts and aircraft measurements for no-turbulence PIREPs peaks very near 0 then tapers off dramatically near 0.06. For moderate-or-severe PIREPs, the JWHAS nowcast and the aircraft measured EDR values peak just above 0.1 with the medians of these conditional distributions at .17 and .19, respectively. Thus, because the distributions are fairly distinct, both the JWHAS system and the aircraft measurements show ability to discriminate between events and non-events. Also, the distribution of the JWHAS values appears very similar to the distribution of the aircraft values.

A dividing point of JWHAS nowcast EDR (0.06, 0.07, 0.1) can be selected such that the JWHAS system can correctly classify 95% (90% / 80%) of the events simultaneously as it correctly classifies 75% (82% / 91%) of the non-events. These dividing points (thresholds) are marked on Figures 4 and 5 with the associated levels of error alpha (α, e.g. 1-PODy) and beta (β, e.g. 1-PODn).

While the presented percentages of error may not be the “best” for operational use, the users can determine the PODy they consider acceptable and determine the level of PODn that results. The levels of error achieved by this system are quite good. In fact, when the alpha and beta errors are fixed in an experimental design to determine the necessary sample size, a common selection is 5% alpha error and 20% beta error (Mood et al., 1974). To achieve error levels so near these “textbook choices” is surprising in an operational system.

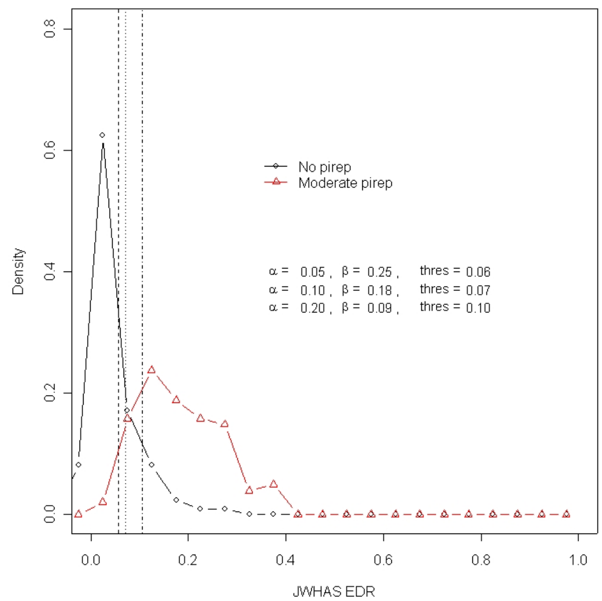

Figure 4: Discrimination plot of JWHAS nowcast EDR values for PIREPs of "No turbulence" vs. "Moderate or severe turbulence".

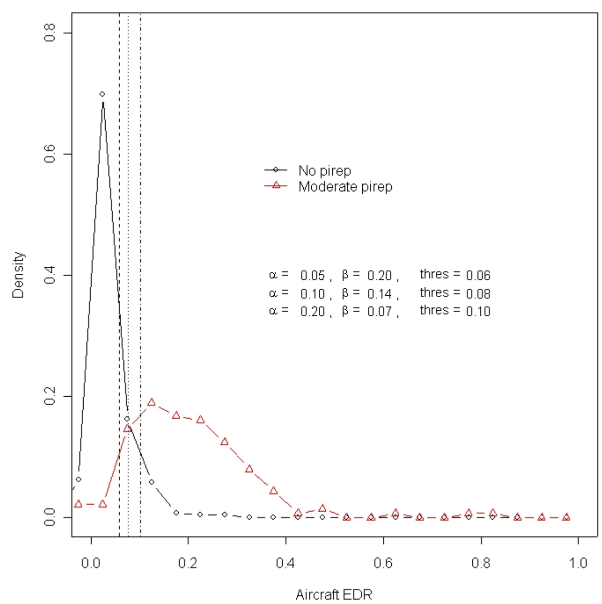

Figure 5: Discrimination plot of aircraft measured ZEDR values for PIREPs of "No turbulence" vs. "Moderate or severe turbulence".

5 CONCLUSIONS AND FUTURE WORK

Generally, the JWHAS performs well in identifying turbulence hazards in a variety of wind regimes and hazard areas. However, the performance does vary considerably across hazard areas and wind regimes. In all cases, it outperforms both the 1999 and 2002 versions of the OPS Spec. JWHAS tends to underestimate the EDR value of large hazards, though it does indicate the presence of a hazard. If the final system produces a binary alarm/no-alarm warning, then this underestimation is unlikely to affect the system.

Due to a lack of high EDR hazards in some combinations of wind regime and hazard area, the JWHAS system is not trained to detect these types of events. However, it may be that these events are extremely unlikely to occur. Thus, the JWHAS system may be accurately representing the range of possible (or likely) conditions.

The OPS Spec does show some ability to identify locations with turbulence hazards, but it has a very high false alarm rate. Therefore its overall skill, as measured by the true skill statistic, is negative. Further, the OPS Spec (both the 1999 and 2002 versions) alarms over five times more often than the JWHAS system and thus restricts use the airspace. However, the OPS Spec has less ability to detect events than the JWHAS system.

The mixed wind regime component of the JWHAS system has few cases, almost no events, and shows little skill. It has been suggested that it may make sense to combine this wind regime with the “calm” wind regime, under which no JWHAS warning is produced.

Overall, the JWHAS system shows excellent ability to discriminate between “no turbulence” PIREPs and “moderate” or “severe” turbulence PIREPs. Discrimination plots suggest that use of an EDR threshold near 0.1 may yield favorable results. Though this threshold may seem low, the JWHAS system is uncalibrated and thus the 0.1 value from the JWHAS system may not represent an observed 0.1 EDR value. Further, the EDR values measured by the aircraft were smoothed prior to use, which is likely to result in an average value.

ACKNOWLEDGEMENTS

NCAR is sponsored by the National Science Foundation. This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy 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. Weather and Forecasting, 12, 890-914.

Brown, B.G., T.L. Kane, R. Bullock, and M.K. Politovich, 1999: Evidence of improvements in the quality of in-flight icing algorithms. Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology, Dallas, TX, 10-15 January, American Meteorological Society (Boston), 48-52.

Doswell, C.A., III, R. Davies-Jones, and D.L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. Weather and Forecasting, 5, 576-585.

Fowler, T.L. and B. G. Brown, 2003: Verification Plan: Juneau Wind Hazard Alert System (JWHAS). Submitted to JWHAS program office.

Juneau FY2002 Algorithm Design Description (ADD), September 13, 2002, DTFA01-98-C-00031, Modification 19.

Mood, A. M, F. A. Graybill, and D. C. Boes, 1974: Introduction to the Theory of Statistics. McGraw Hill, Boston, Massachusetts.

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