NOAA Technical Memorandum OAR FSL-30

CURRENT ICING POTENTIAL FOR ALASKA (CIP-AK): QUALITY ASSESSMENT REPORT

M. Kelsch
J.L. Mahoney
T. Fowler
B.G. Brown
J. Henderson
C. Fisher

Forecast Systems Laboratory
Boulder, Colorado
September 2003

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Oceanic and Atmospheric Research Laboratories

NOAA Technical Memorandum OAR FSL-30

CURRENT ICING POTENTIAL FOR ALASKA (CIP-AK): QUALITY ASSESSMENT REPORT

Aviation Weather Research Program
Quality Assessment Product Development Team[1]

Matthew Kelsch[2, 3]
Jennifer Luppens Mahoney[3]
Tressa Fowler[4]
Barbara G. Brown[4]
Judy Henderson[3]
Chris Fisher[2, 3]

[1]Contact: M. Kelsh (Matt.Kelsh@noaa.gov), J.L. Mahoney (Jennifer.Mahoney@noaa.gov), or B.G. Brown (bgb@rap.ucar.edu)
[2]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado
[3]NOAA Forecast Systems Laboratory, Boulder, Colorado
[4]National Center for Atmospheric Research (NCAR), Boulder, Colorado

Forecast Systems Laboratory
Boulder, Colorado
September 2003

UNITED STATES DEPARTMENT OF COMMERCE
Donald L. Evans
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
VADM Conrad C. Lautenbacher, Jr.
Under Secretary for Oceans and Atmosphere/Administrator

Oceanic and Atmospheric Research Laboratories
Richard D. Rosen
Director

Notice

Mention of a commercial company or product does not constitute an endorsement by the NOAA Oceanic and Atmospheric Research Laboratories. Use of information from this publication concerning proprietary products or the test of such products for publicity or advertising purposes is not authorized.

For sale by the National Technical Information Service, 5285 Port Royal Road
Springfield, VA 22061

Contents

Current Icing Potential for Alaska (CIP-AK): Quality Assessment Report

Matthew Kelsch, Jennifer Luppens Mahoney, Tressa Fowler, Barbara G. Brown, Judy Henderson, and Chris Fischer

ABSTRACT. Current Icing Potential for Alaska (CIP-AK). CIP-AK was developed by the Inflight Icing Product Development Team of the Federal Aviation Administration Aviation Weather Research Program (FAA/AWRP), and is currently being considered for transition to an experimental product through the Aviation Weather Technology Transfer (AWTT) process.

This report concentrates on verification results for the CIP-AK computed for the periods 23. This report summarizes the assessment of icing conditions produced by the July–11 September 2002 and 1 January–30 May 2003. The evaluation considers performance of the algorithm over three domains in Alaska where observation concentrations are highest. Voice pilot reports (PIREPs) were used to evaluate the CIP-AK performance and were augmented with additional PIREPs supplied during 23 July–11 September 2002 by PenAir Airlines. Because of the sparse nature of PIREP observations over Alaska, overall results are presented in the report, while further stratifications of the results are excluded from the analysis.

The icing diagnoses were verified every 3-h using Yes and No icing observations from PIREPs indicating either "light or greater" icing severity or "no icing." CIP-AK diagnoses were evaluated as Yes/No icing diagnoses by applying a threshold to convert the algorithm output to a Yes or No value. A variety of thresholds was applied to the algorithm output in order to examine the full range of CIP-AK performance characteristics. The verification analyses were primarily based on the algorithm's ability to discriminate between Yes and No observations, as well as the extent of their coverage. To provide a standard of comparison, complimentary results for the Airmens' Meteorological Advisories (AIRMETs), the operational forecasts issued by the National Weather Service Alaskan Aviation Weather Unit (AAWU), are presented in Appendix A. Several thousand individual CIP-AK diagnoses were considered in this evaluation. The number of icing observations considered (Yes and No) in the evaluation ranged from 3,070 for the largest of the three domains, CIP-AK Land domain, to 528 for the Anchorage-area subdomain.

Results of the evaluation indicate that CIP-AK is skillful at discriminating between Yes and No icing conditions. CIP-AK provides relatively efficient diagnoses, covering comparatively small volumes for a given icing detection rate. Using a threshold of 0.10, CIP-AK correctly classifies 68% of the Yes PIREPs and 66% of the No PIREPs, while covering approximately 13% of the airspace volume over the CIP-AK Land domain. The forecast quality is maintained up to about 15,000 ft. The quality of CIP-AK diagnoses is relatively insensitive to variations in the PIREPs used for the analyses, and performance was best between 3,000 and 15,000 ft where the number of observations was greatest. Detection rates and volumes covered varied, with the most favorable numbers for the larger of the three domains (CIP-AK Land). Trends in CIP-AK performance over the two evaluation periods suggest that CIP-AK performs similarly in summer 2002 as in winter/spring 2003. However, the summer 2002 dataset had supplemental PIREPs from PenAir Airlines.

1. INTRODUCTION

This report summarizes basic results of an evaluation of the Current Icing Potential algorithm for Alaska (CIP-AK). This algorithm is under consideration for transition to experimental status through the Aviation Weather Technology Transfer (AWTT) process. CIP-AK is designed to diagnose in-flight icing conditions over Alaska. The In-flight Icing Product Development Team, through real-time forecasting exercises, has subjectively evaluated the CIP-AK over several seasons. This objective evaluation, performed by the Quality Assessment Product Development Team (QA PDT) of the Federal Aviation Administration Aviation Weather Research Program (FAA/AWRP), follows those verification techniques used to evaluate the quality of the CONUS CIP and Forecast Icing Potential (FIP) algorithms (Brown et al. 2001, 2002).

The analyses presented in this report focus primarily on evaluations of CIP-AK diagnoses from 23 July–11 September 2002 (summer 2002) and 1 January–30 May 2003 (winter/spring 2003). These two periods represent a combined total of roughly 30 weeks. The evaluation considers performance of the algorithm over three domains in Alaska, where observation concentrations are highest. The largest domain covers the land-area portion of the CIP-AK domain, and the other two are subdomains. Voice pilot reports (PIREPs) were used to evaluate the CIP-AK performance and were augmented with additional PIREPs supplied during 23 July–11 September 2002 by PenAir Airlines. Because of the sparse nature of PIREP observations over Alaska, overall results are presented in the report, and further stratifications of the results are excluded from the analysis.

Complementary results for the Airmens' Meteorological Advisories (AIRMETs), the operational forecasts issued by the National Weather Service Alaskan Aviation Weather Unit (AAWU), are presented in Appendix A to provide a standard of comparison for the CIP-AK.

The report is organized as follows. The study approach is presented in Section 2. Section 3 briefly describes the algorithm and PIREPs that were included in the evaluation. The verification methods are described in Section 4, results of the study are presented in Section 5, and conclusions in Section 6. Appendix A contains verification results for the AAWU AIRMETs, and Appendix B considers the characteristics of PIREP data in Alaska.

2. APPROACH

The CIP-AK diagnosis is based on output provided by the Eta model running operationally over Alaska. CIP-AK diagnoses issued every 3h from 23 July–11 September 2002 (summer 2002) and 1 January–30 May 2003 (winter/spring 2003) were included in the verification study. Statistics for CIP-AK were computed for three regions, which are shown in Fig. 1. These are, from largest to smallest, (1) the CIP-AK domain with oceanic areas removed due to lack of observations and is called *CIP-AK Land*, (2) a subdomain focused on three AAWU forecast zones that we will call *Zones*, and (3) a subdomain focused on the relatively high-traffic areas around Anchorage that we will call *Anchorage.*

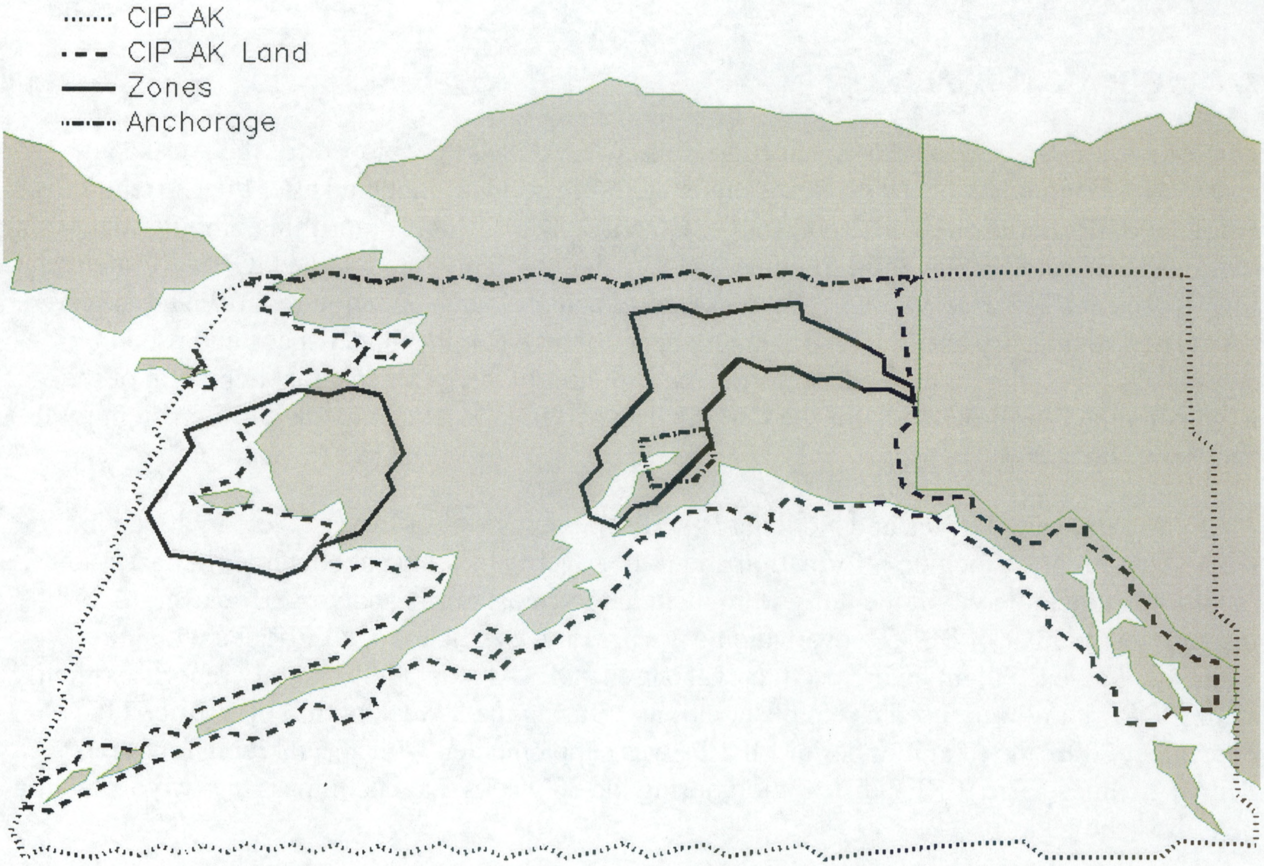

Figure 1: CIP-AK domains used in the evaluation: (1) CIP-AK Land (dashed), (2) Zones (solid), and (3) Anchorage (dash-dot), and the overall CIP-AK domain (dotted) which is not used in the evaluation.

The verification approach applied in this study is identical to the approach taken in previous studies associated with the CONUS versions of CIP and FIP (e.g., Brown et al. 2001, 2002). In particular, the algorithm diagnoses were verified using Yes and No PIREPs of icing. The algorithm diagnoses were transformed into Yes/No icing fields by determining if the algorithm output at each model grid point exceeded or was less than a pre-specified threshold; a variety of different thresholds were utilized to examine the full range of performance of the algorithm. The Yes/No diagnoses were evaluated using standard verification techniques available for Yes/No forecasts, whose observations are based on PIREPs. In addition, the amount of airspace impacted by the forecasts was considered. For the analyses, PIREPs reporting at least "light" icing severity were included as Yes reports.

Analyses were stratified by PIREP type, forecast domain, and time period. However, to ensure an adequate number of observations for evaluating the CIP-AK, the results presented in this report incorporate the statistical information from both evaluation periods and both PIREP types into overall statistical results.

3. CIP-AK AND PIREPs

CIP-AK: An in-flight icing diagnostic algorithm, CIP-AK, was developed for the Alaska air space (McDonnough et al. 1999, Bernstein et al. 2003) and was modeled after the Current Icing Potential (CIP; Bernstein et al 2001) for the CONUS. The CIP-AK algorithm combines data from the Alaska ETA (AK ETA) 45-km numerical weather prediction model, the GOES-10 satellite imager, and surface observations from the Alaska meteorological aviation reports. The observed data (satellite imagery and METARs) are mapped to the AK ETA model space and used in combination with the model soundings to create an hourly three-dimensional depiction of the icing potential. Further details for the CIP-AK can be found in McDonough and Bernstein (1999) and Bernstein (2003).

PIREPS: All available Yes and No icing PIREPs in the regions of interest were included in the study. These reports include information about the severity of icing encountered. In past studies, reports of moderate to extreme icing were included for most of the analyses. However, to increase the number of PIREPs available for verification over Alaska, all PIREPs with an intensity of at least "light" were used to evaluate the CIP-AK. In addition, only reports explicitly stating "No" icing were used as No observations. During the 8-week period of 23 July-11 September 2002, the total number of PIREPs was supplemented with special icing reports from PenAir Airlines. The PIREP distribution during the 30 weeks that encompass the two evaluation periods is shown in Fig. 2.

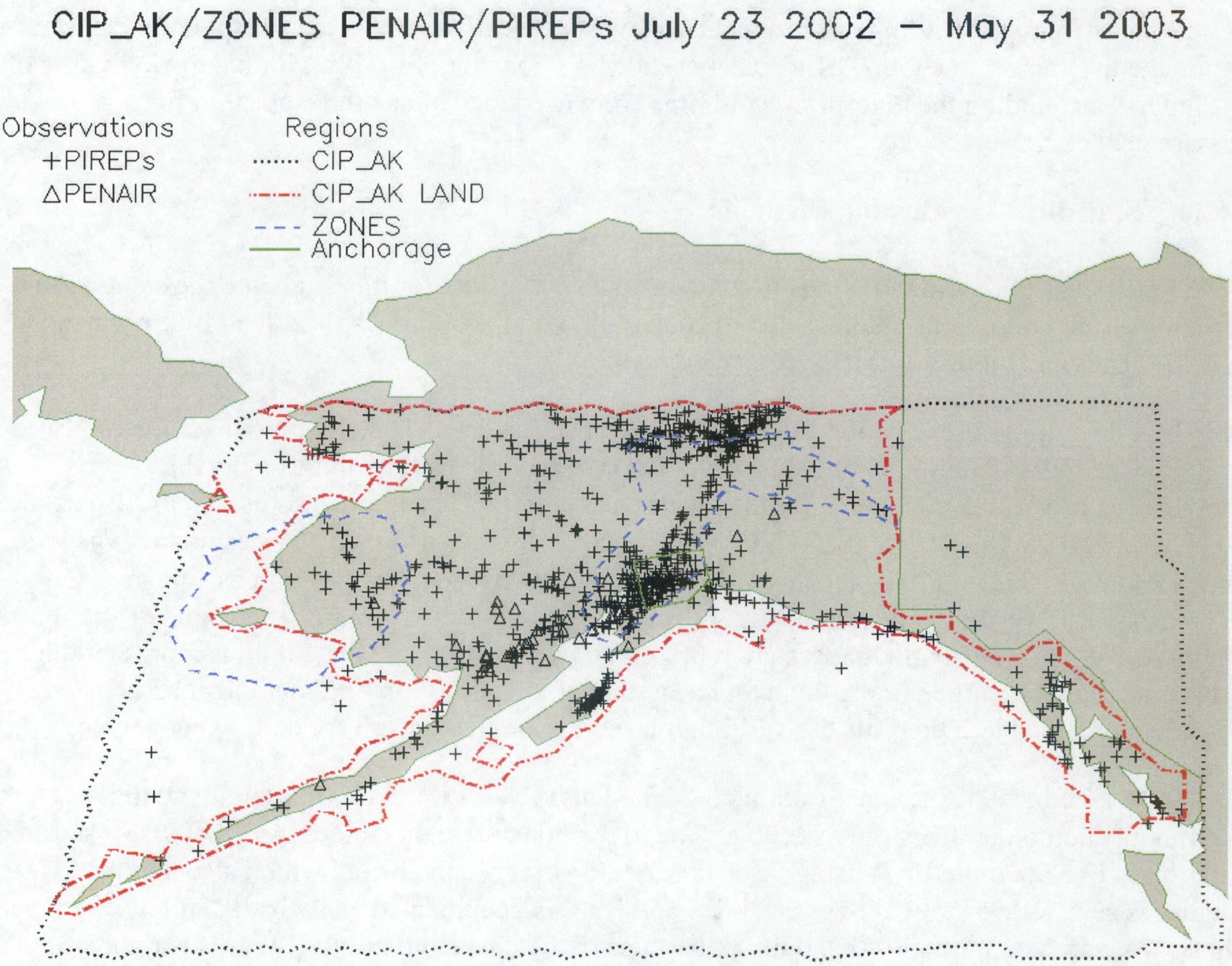

Figure 2: Distribution of observations used in the CIP-AK evaluation. The legend provides information about the type of PIREPs. PenAir PIREPs were only available during the summer 2002 period.

4. METHODS

This section summarizes methods that were used to match CIP-AK diagnoses and observations, as well as the various verification statistics that were computed for the evaluation.

4.1 Matching Methods

The methods used to connect the PIREPs to the CIP-AK diagnoses are the same as have been used in previous evaluations of CIP and other in-flight icing algorithms (e.g., Brown et al. 1997, 2001, 2002). In particular, each PIREP is connected to the CIP-AK diagnoses at the nearest 8 grid points (four surrounding grid points; two levels vertically). A bilinear interpolation is used to compute the appropriate CIP-AK value at each PIREP location. PIREPs within a time window of ±2 hours surrounding the algorithm valid time were used to evaluate the algorithm for this assessment exercise.

4.2 Statistical Verification Methods

The statistical verification methods used to evaluate the results for this study are the same as the methods used in previous studies and are consistent with the approach described by Brown et al. (1997). These methods are briefly described here.

CIP-AK icing diagnoses and the PIREP observations are treated as dichotomous (i.e., Yes/No) values. The CIP-AK icing potential values are converted to a variety of Yes/No values by application of various thresholds for the occurrence of icing. The thresholds used for CIP-AK are 0.005, 0.02, 0.05, 0.10, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85. For example, when the threshold of 0.10 is applied, then any CIP-AK values ≥0.10 are treated as a "Yes" diagnosis. Thus, the basic verification approach makes use of the two-by-two contingency table (Table 1). In this table, the CIP-AK diagnoses are represented by the rows, and the columns represent the observations. The entries in the table represent the joint distribution of icing diagnoses and observations. Table 2 presents the verification scores computed from the contingency table.

It will be noted that Table 2 does not include the False Alarm Ratio (FAR), a statistic that is commonly computed from the 2x2 table. Due to the nonsystematic nature of PIREPs, it is not appropriate to compute FAR using these observations. This conclusion, which also applies to statistics such as the Critical Success Index and Bias, is documented analytically and by example in Brown and Young (2000). In addition, because of the characteristics of PIREPs and their limited numbers, other verification statistics (e.g., PODy and PODn) should not be interpreted in an absolute sense, but can be used in a comparative sense, for comparisons among algorithm diagnostics and forecasts. Moreover, PODy and PODn should not be interpreted as probabilities, but rather as *proportions of PIREPs that are correctly diagnosed by CIP-AK.*

Together, PODy and PODn measure the ability of the CIP-AK diagnoses to discriminate between (or correctly categorize) Yes and No icing observations. This ability to discriminate is summarized by the True Skill Statistic (TSS), frequently called the Hanssen-Kuipers discrimination statistic (Wilks 1995). Note that it is possible to obtain the same value of TSS for a variety of combinations of PODy and PODn. Thus, it is always important to consider the PODy and PODn, as well as TSS.

As shown in Table 2, three other statistics are utilized for verification of the icing diagnoses: % Area, % Volume, and Volume Efficiency (VE). The % Volume statistic is the percent of the total possible airspace volume that has a Yes diagnosis. The % Area indicates the proportion of the surface area of the domain that is associated with a Yes icing diagnosis. Remember that the % Area is the diagnosed area at each level projected onto one plane. VE considers PODy relative to the volume covered by the diagnosis, and can be thought of as the POD per-unit volume. *The VE statistic must be used with some caution, however, and should not be used by itself as a measure of quality*. For example, sometimes it is easy to obtain a large VE value when PODy is very small. An appropriate use of VE is to compare the efficiencies of forecasting systems that have nearly equivalent values of PODy. In fact, *none of the statistics should be considered in isolation* – all should be examined in combination with the others to obtain a complete picture of the quality of the algorithm guidance.

Plots of PODy versus % Area (% Volume) show the relationship between PODy and % Area (% Volume) for various thresholds. For the PODy versus % Area (% Volume) plots, threshold values that result in optimal icing diagnoses are located along the part of the curve that is closer to the upper left corner of the diagram.

The relationship between PODy and 1-PODn for different algorithm thresholds is the basis for the verification approach known as "Signal Detection Theory" (SDT). For a given algorithm, this relationship can be represented by the curve joining the points (1-PODn, PODy) for different algorithm thresholds. The resulting curve is known as the "Relative Operating Characteristic" (ROC) curve in SDT. The area under this curve is a measure of overall forecast skill (e.g., Mason 1982), with ROC area values >0.5 indicating positive skill. As with the PODy versus % Area (%Volume) curves, the thresholds that represent the best algorithm performance are located along the part of the curve that approaches the upper left corner of the plot.

Table 1: Contingency table for evaluation of dichotomous (Yes/No) forecasts. Elements in the cells are the counts of forecast-observation pairs.

Forecast	**_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: Verification statistics used in this study.

Statistic	*Definition*	*Description*	*Interpretation*	*Range*
PODy	YY/(YY+NY)	Probability of Detection of Yes observations	Proportion of Yes observations that were correctly forecasted	0–1 Best: 1 Worst: 0
PODn	NN/(YN+NN)	Probability of Detection of No observations	Proportion of No observations that were correctly forecasted	0–1 Best: 1 Worst: 0
TSS	PODy + PODn – 1	True Skill Statistic; Hanssen-Kuipers discrimination	Level of discrimination between Yes and No observations	-1 to 1 Best: 1 No skill: 0
Curve Area	Area under the curve relating PODy and 1-PODn	Area under the curve relating PODy and 1-PODn (i.e., the ROC curve)	Overall skill (related to discrimination between Yes and No observations)	0–1 Best: 1 No skill: 0.5
% Area	[(Forecast Area) / (Total Area)] x 100	% of the total area (e.g., CONUS) that has a Yes forecast at some level above	% of the area that is impacted by a Yes forecast at one or more flight levels above	0–100 Smaller is better
% Volume	[(Forecast Vol) / (Total Vol)] x 100	% of the total air space volume that is impacted by the forecast	% of the total air space volume that is impacted by the forecast	0–100 Smaller is better
Volume Efficiency (VE)	(PODy x 100) / % Volume	PODy (x 100) per unit % Volume	PODy relative to airspace coverage	0–infinity Larger is better

As in previous icing algorithm verification analyses, emphasis in this report will be placed on PODy, PODn, and % Volume. Use of this combination of statistics implies that the underlying goal of the algorithm development is to include most Yes PIREPs in the "Yes icing" region, and most No PIREPs in the "No icing" region (i.e., to increase PODy and PODn), while minimizing the extent of the airspace volume diagnosed with icing by CIP-AK, as represented by % Volume. ROC curve areas also will be considered as a measure of the overall skill of CIP-AK at discriminating between Yes and No observations.

5. RESULTS

Basic results of the CIP-AK evaluations described here are organized in subsections that describe the overall performance and the performance by altitude.

5.1 Overall Results

Overall verification results for CIP-AK for summer 2002 and winter/spring 2003 combined are shown in the ROC curves (PODy versus 1-PODn) of Figs. 3–5. Figures 6–8 present plots of PODy versus % Area and Figs. 9–11 present plots of PODy versus % Volume. The results are based on "light or greater" PIREPs in which all CIP-AK diagnoses/observations pairs for the 3-h diagnoses are combined. The number of PIREPs available for evaluating the CIP-AK Land domain was 3,070, with 1,259 for the Zones domain, and 528 for the Anchorage domain. Each point on the curve represents a different threshold used to define Yes/No icing forecasts. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85. Forecast skill is indicated when the curve lies above the diagnonal. Greater CIP-AK skill is indicated as the curve approaches the upper left-hand corner.

These results suggest that CIP-AK has skill for distinguishing between Yes and No icing conditions for all three domains, as indicated by the location of the curve to the left of the diagonal line on Figs. 3-5. The tendency for the curves in Figs. 3–5 and Figs. 9-11 to approach the upper left is more pronounced for the CIP-AK Land and Zones domains than it is for the Anchorage domain. This implies that the ability for CIP-AK to distinguish between Yes/No icing conditions within a specific volume is better for the larger domains. This is likely related to the greater number of observations available in the larger domains. Furthermore, the "bumpy" nature of the plot in Fig. 5 is likely due to the small number of PIREPs in this region. Figures 6–8 (PODy vs. % Area) show that the CIP-AK has skill at some thresholds, but it is not as pronounced as the plots in Figs. 9–11 (PODy versus % Volume). One possible reason that the %Volume plots show more favorable results than the % Area plots is that CIP-AK is better at diagnosing airspace volumes (vertical and horizontal extent) than it is at diagnosing the combined horizontal extent used in the % Area. Furthermore, it is important to remember that the % Area is the diagnosed area at each level projected onto one plane. Therefore, small volumes diagnosed at different locations may add up to a relatively small % Volume but a very large % Area giving the impression of over-diagnosing the horizontal extent of icing conditions.

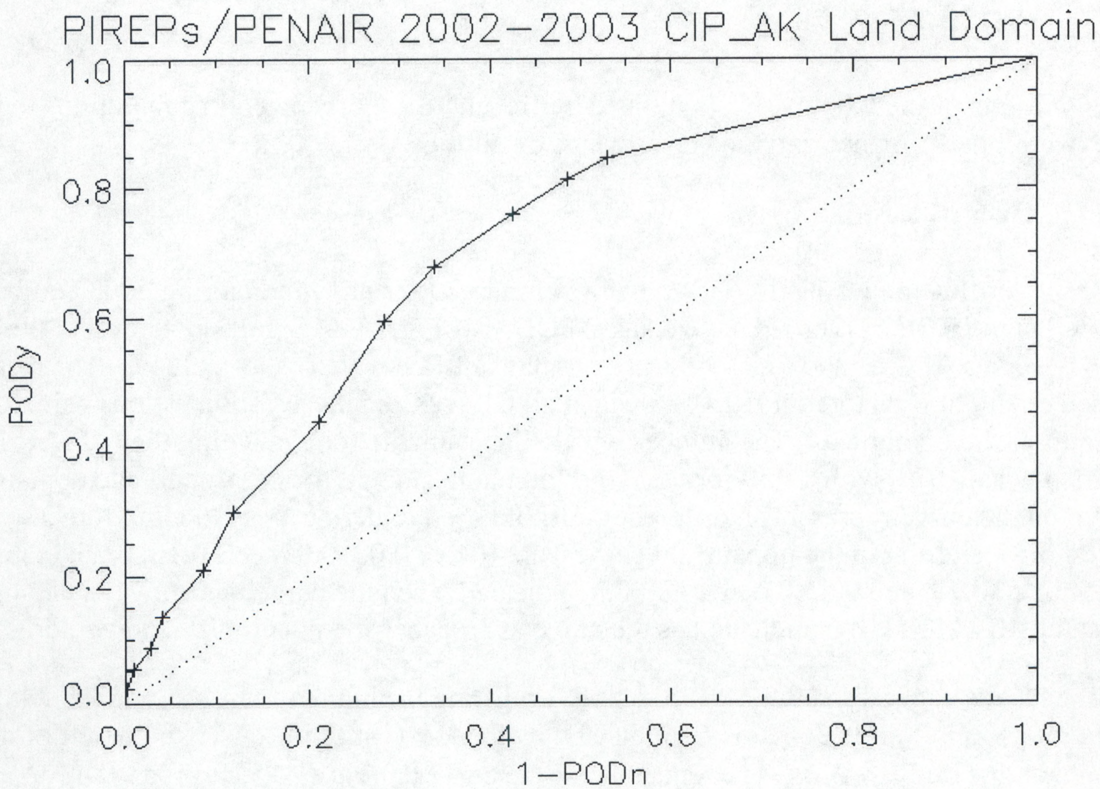

Figure 3. ROC curves (PODy versus 1-PODn) for CIP-AK during summer 2002 and winter/spring 2003, for all valid times combined, based on light and greater PIREPs for CIP-AK Land domain. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85. The (1,1) point is also included to complete the curve.

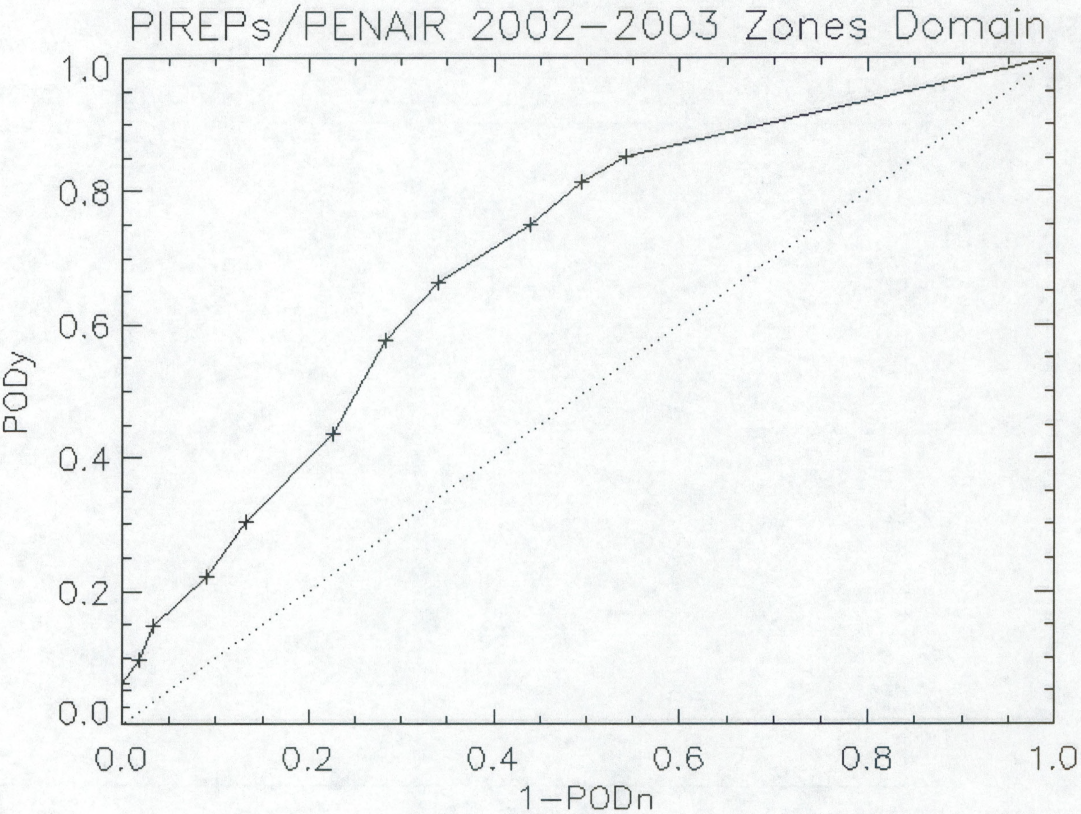

Figure 4. Same as Fig. 3, except for the Zones domain.

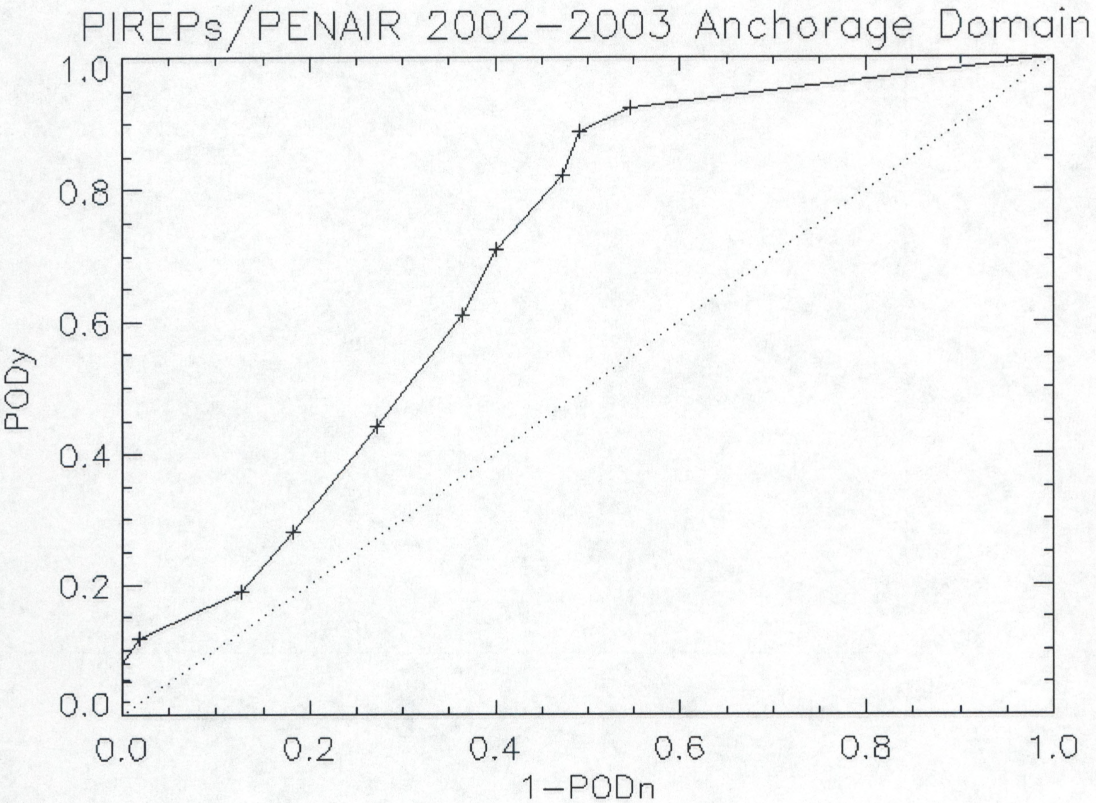

Figure 5. Same as Fig. 3, but for the Anchorage domain.

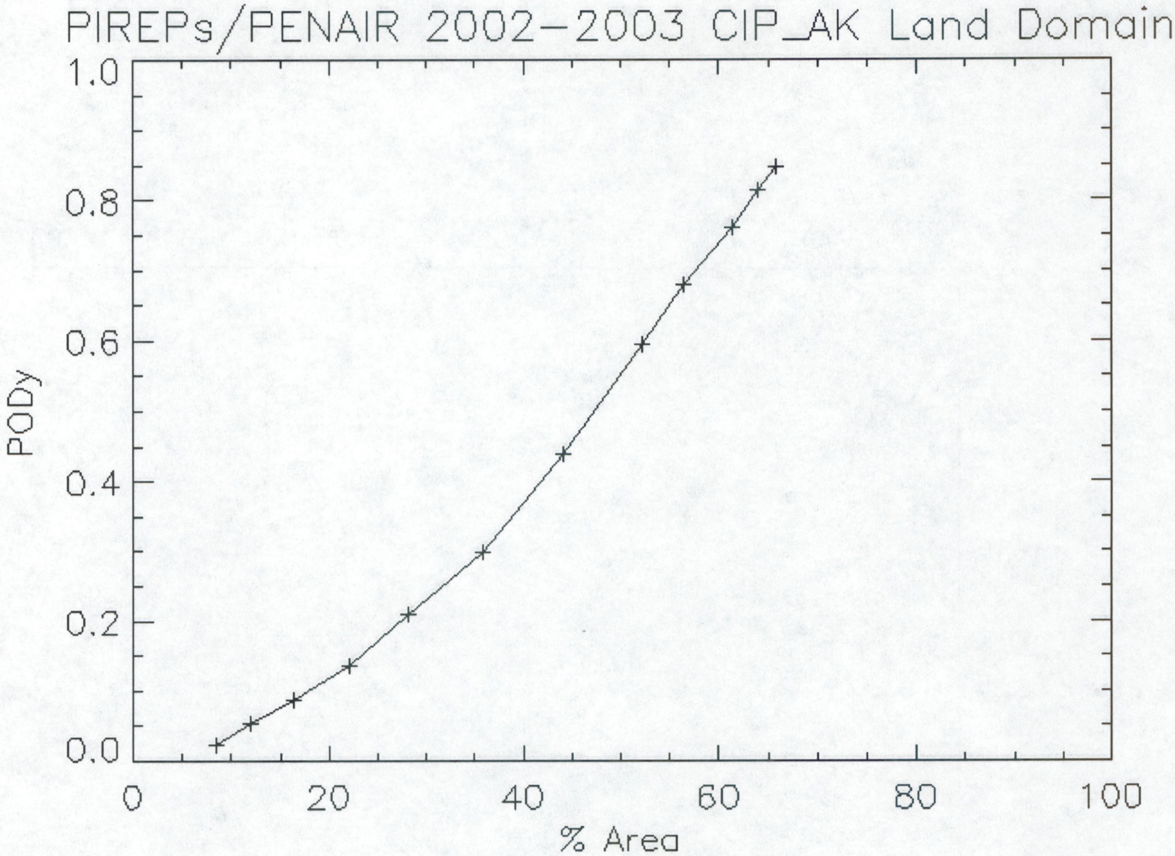

Figure 6. Plot of PODy versus % Area for summer 2002 and winter/spring 2003, for all valid times combined, based on light or greater PIREPs over the CIP-AK Land domain. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85.

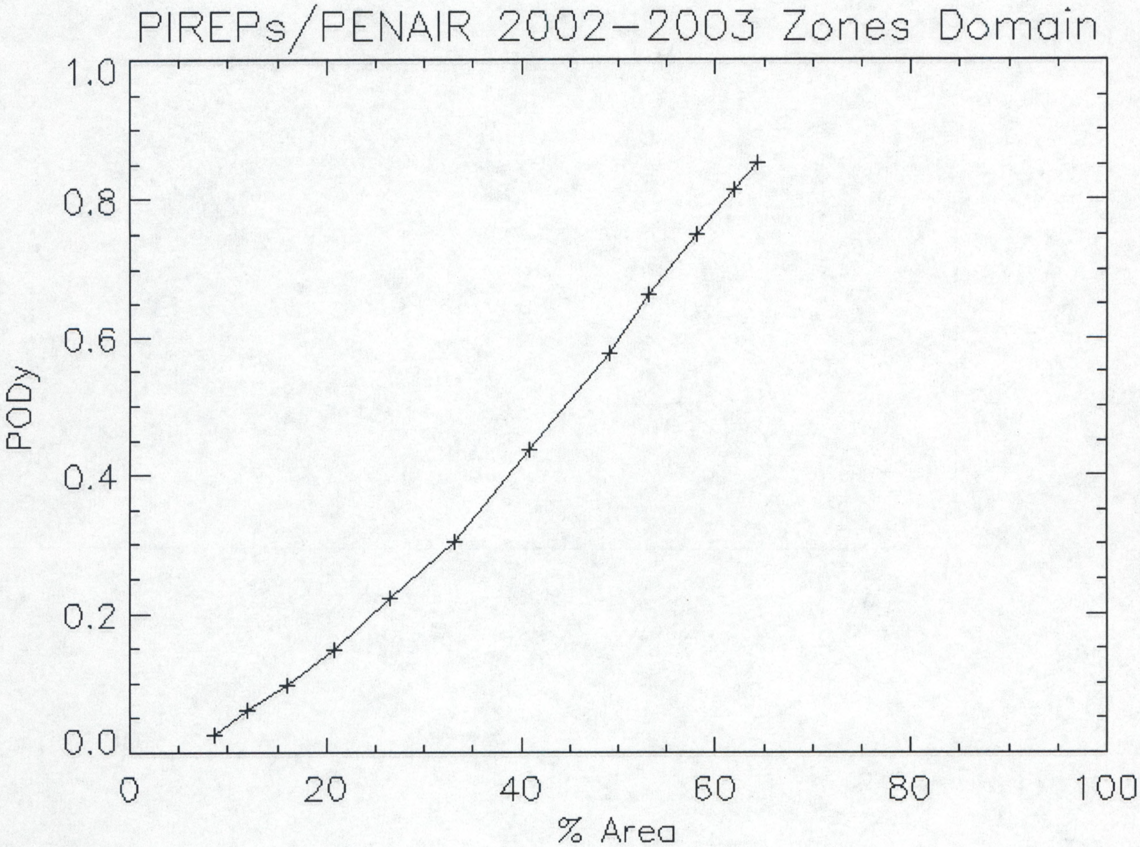

Figure 7. Same as Fig. 6, but for the Zones domain.

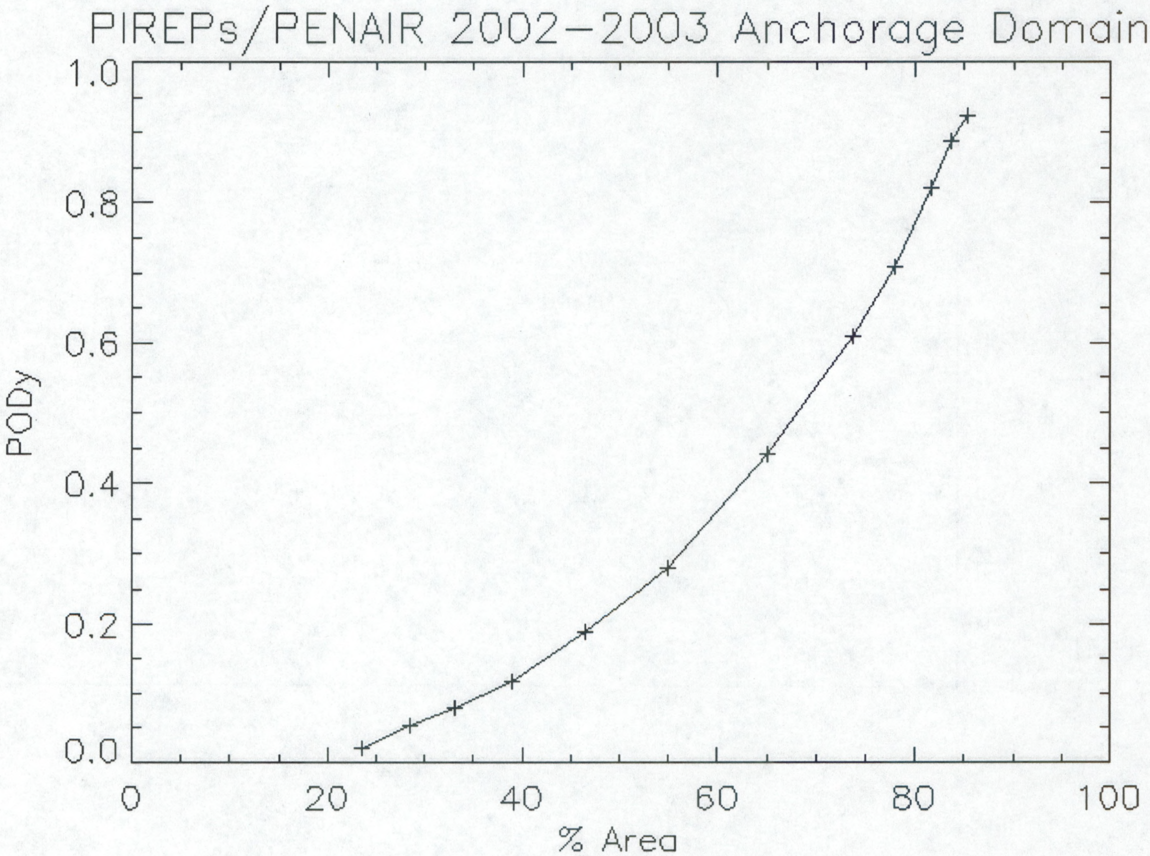

Figure 8. Same as Fig. 6, but for the Anchorage domain.

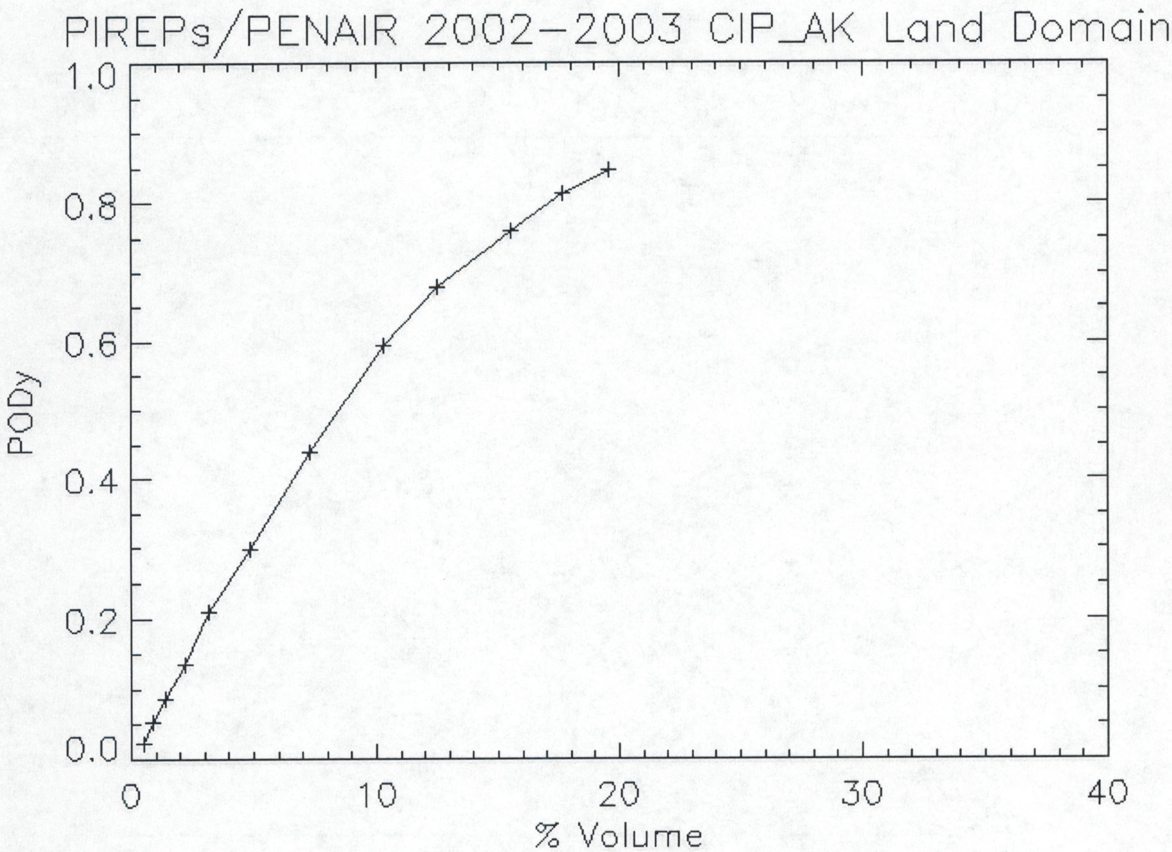

Figure 9. Plot of PODy versus % Volume for summer 2002 and winter/spring 2003 for all valid times combined, based on light or greater PIREPs over the CIP-AK Land domain. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85.

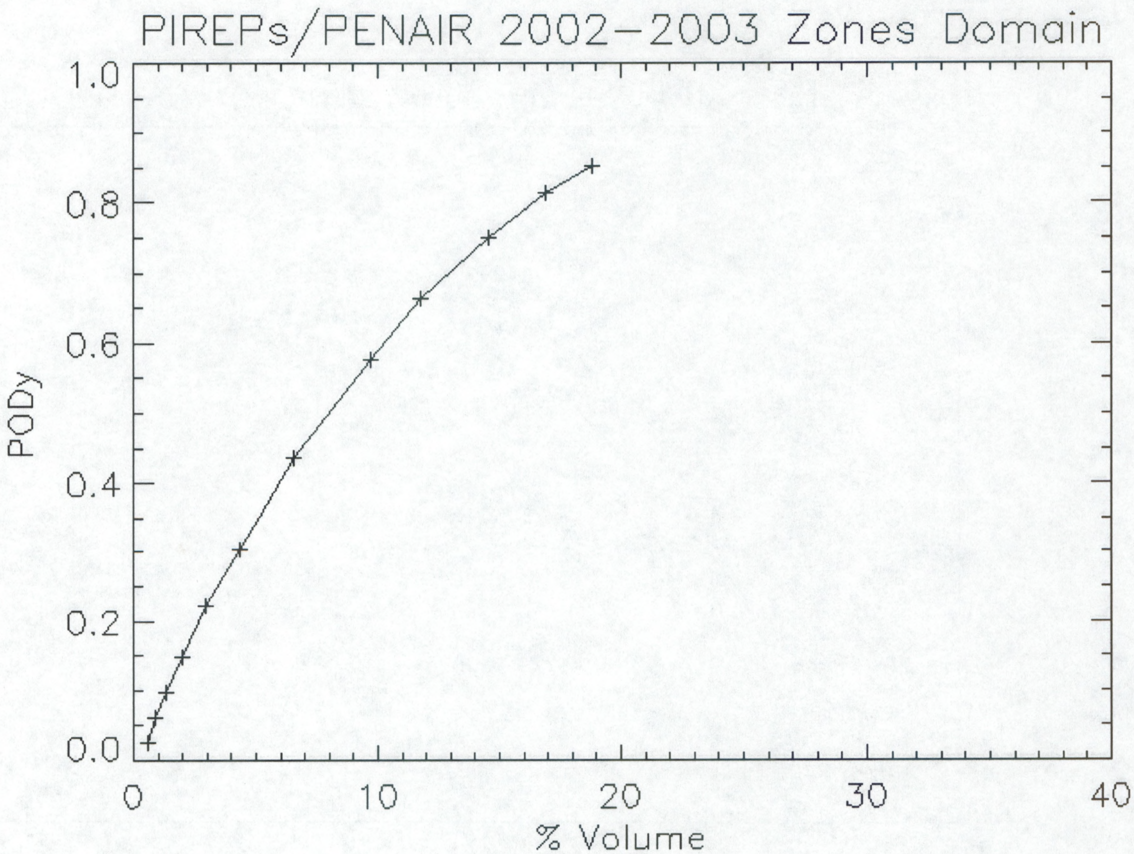

Figure 10. Same as Fig. 9, but for the Zones domain.

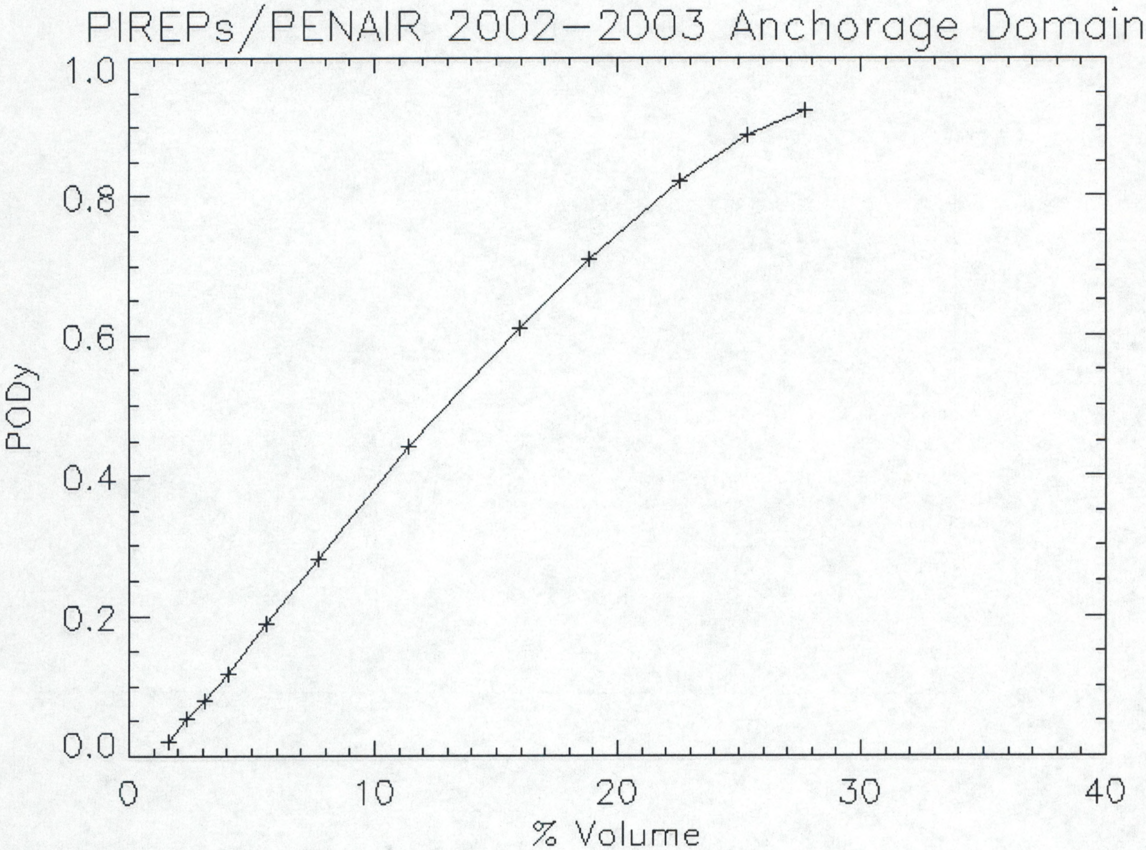

Figure 11. Same as Fig. 9, but for the Anchorage domain.

Tables 3–5 present results for the three domains. The rows provide a lookup for the 12 thresholds used in CIP-AK, and the columns provide the statistical measures as defined in Tables 1–2. Table 6 shows the area under the ROC curve for each domain and supports the plotted data in Figs. 3–5. That is, CIP-AK does show skill in its icing diagnoses, and this skill is most pronounced in the CIP-AK Land domain.

The results presented in Tables 3–5 show that for all CIP-AK thresholds, the PODy values range from 0.92 to 0.02, PODn from 1.00 to 0.45, and TSS from 0.40 to 0.02. The values presented in Tables 3–5 suggest that a threshold value of 0.10 may be a good choice for a detailed look at CIP-AK performance. This threshold results in a good balance between PODy and PODn with a reasonable % Volume . As compared to the quality of the national CIP algorithm at a threshold of 0.15 (Brown et al. 2001; PODy = 0.8, PODn = 0.62, and TSS = 0.45), the overall statistics for the CIP-AK at 0.10 indicate lower values for PODy, PODn, and TSS. For the three domains with a CIP-AK threshold of 0.10, PODy ranges from 0.66–0.71, the PODn from 0.68–0.60, and the TSS from 0.34–0.31. However, these lower values are expected because the observational dataset available for formulating and evaluating the CIP-AK was limited. Nevertheless, the results for the CIP-AK are comparable to the results for the CONUS CIP indicating some skill in diagnosing icing conditions in Alaska.

Figure 12 shows a scatterplot of PODy versus % Volume for each day that observations and forecasts were available during the 30-week evaluation period. The points on the plot represent the average daily value of PODy and % Volume. The plot suggests that as % Volume increases, there is a tendency for PODy to also increase. However, as shown by the scatter in Fig. 12, there is a large amount of variability in the daily statistics, thus supporting the decision to focus on longer time periods for generating the results.

Table 3. Overall statistics for the CIP-AK Land domain for summer 2002 and winter/spring 2003 combined using PIREPs and PenAir observations for verification. The rows represent the specific CIP-AK thresholds, and the columns represent the different statistical measures.

Thresh	Forecast/Observation Pairs				Statistic					
	YY	YN	NY	NN	PODY	PODN	TSS	%Area	%Vol	Vol Eff
0.005	2221	238	400	211	0.84	0.47	0.32	65.8	19.6	4.3
0.02	2135	218	486	231	0.81	0.51	0.33	64.0	17.7	4.6
0.05	1994	191	627	258	0.76	0.57	0.34	61.4	15.5	4.9
0.1	1777	152	844	297	0.68	0.66	0.34	56.5	12.5	5.4
0.15	1560	127	1061	322	0.60	0.72	0.31	52.2	10.3	5.8
0.25	1147	95	1474	354	0.44	0.79	0.23	44.2	7.3	6.0
0.35	783	53	1838	396	0.30	0.88	0.18	35.9	4.8	6.2
0.45	547	38	2074	411	0.21	0.92	0.12	28.1	3.2	6.6
0.55	356	18	2265	431	0.14	0.96	0.10	22.0	2.2	6.2
0.65	229	12	2392	437	0.09	0.97	0.06	16.3	1.4	6.2
0.75	142	4	2479	445	0.05	0.99	0.05	12.0	0.9	6.0
0.85	58	1	2563	448	0.02	1.00	0.02	8.4	0.6	4.0

Table 4. Same as Table 3, but for Zones domain.

Thresh	Forecast/Observation Pairs				Statistic					
	YY	YN	NY	NN	PODY	PODN	TSS	%Area	%Vol	Vol Eff
0.005	891	115	156	97	0.85	0.46	0.31	64.2	18.8	4.5
0.02	852	105	195	107	0.81	0.50	0.32	62.0	16.9	4.8
0.05	786	93	261	119	0.75	0.56	0.31	58.1	14.5	5.2
0.1	695	72	352	140	0.66	0.66	0.32	53.0	11.8	5.6
0.15	603	60	444	152	0.58	0.71	0.29	49.0	9.7	6.0
0.25	456	48	591	164	0.44	0.77	0.21	41.0	6.6	6.6
0.35	319	28	728	184	0.30	0.87	0.17	33.0	4.3	7.1
0.45	232	19	815	193	0.22	0.91	0.13	27.0	2.9	7.6
0.55	154	7	893	205	0.15	0.97	0.11	21.0	2.0	7.4
0.65	101	4	946	208	0.10	0.98	0.08	16.0	1.4	7.1
0.75	65	0	982	212	0.06	1.00	0.06	12.0	0.9	7.1
0.85	28	0	1019	212	0.03	1.00	0.03	8.7	0.6	4.8

Table 5. Same as Table 3, but for Anchorage domain.

Thresh	**Forecast/Observation Pairs**				**Statistic**					
	YY	**YN**	**NY**	**NN**	**PODY**	**PODN**	**TSS**	**%Area**	**%Vol**	**Vol Eff**
0.005	437	30	36	25	0.92	0.45	0.38	85.2	27.7	3.3
0.02	420	27	53	28	0.89	0.51	0.40	83.7	25.3	3.5
0.05	388	26	85	29	0.82	0.53	0.35	81.6	22.6	3.6
0.1	336	22	137	33	0.71	0.60	0.31	78.0	18.9	3.8
0.15	288	20	185	35	0.61	0.64	0.25	73.7	15.9	3.8
0.25	209	15	264	40	0.44	0.73	0.17	65.1	11.4	3.9
0.35	133	10	340	45	0.28	0.82	0.10	55.0	7.7	3.6
0.45	89	7	384	48	0.19	0.87	0.06	47.0	5.6	3.5
0.55	56	1	417	54	0.12	0.98	0.10	39.0	4.0	2.9
0.65	38	0	435	55	0.08	1.00	0.08	33.1	3.1	2.6
0.75	25	0	448	55	0.05	1.00	0.05	28.6	2.4	2.2
0.85	10	0	463	55	0.02	1.00	0.02	243.6	1.7	1.3

Table 6. The area under the ROC curves for each domain.

	Area under ROC
CIP-AK Land	0.707
Zones	0.701
Anchorage	0.694

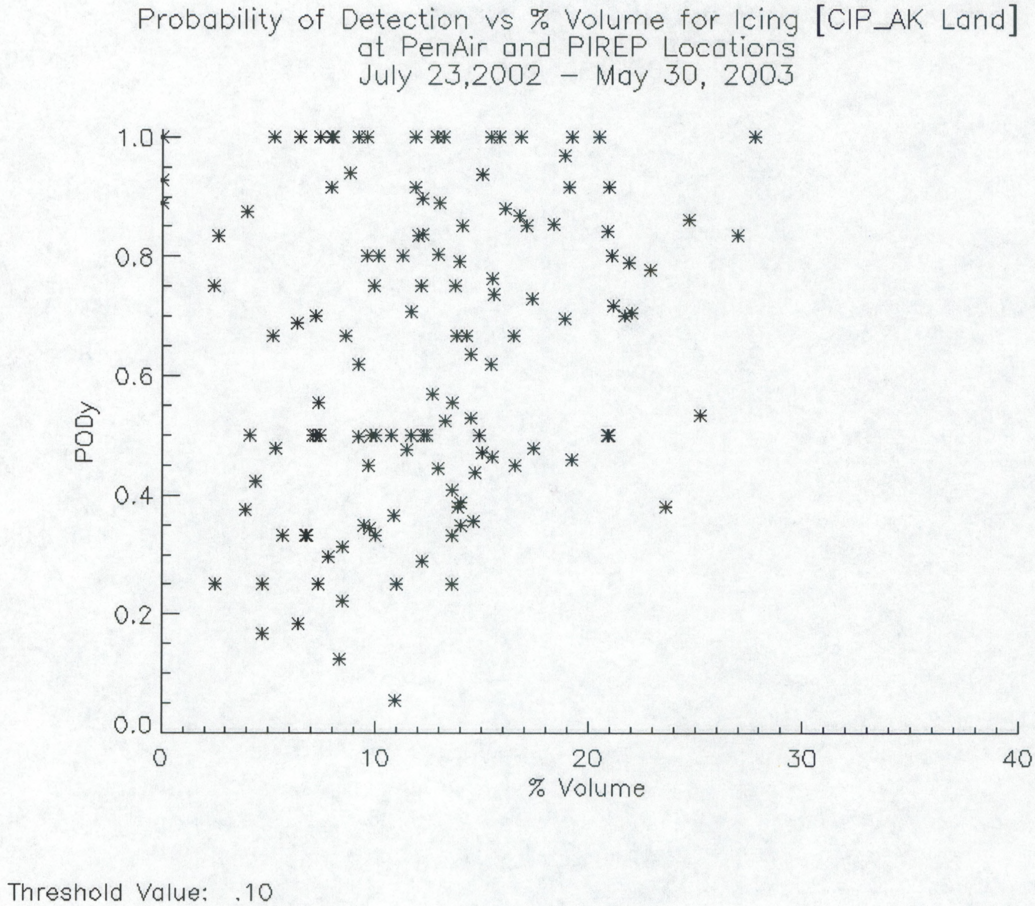

Figure 12. Scatterplot of PODy versus % Volume. Points represent the average daily PODy and % Volume values for each day during the 30-week period when forecasts and observations were available.

5.2 Comparisons by Altitude

To assess the performance of CIP-AK forecasts at different altitudes, verification statistics were computed separately for each 3,000 ft interval from the surface to 30,000 ft. Plots of statistics at all altitudes are presented for CIP-AK in Figs. 13–15 for each of the three domains using the combined summer 2002 and winter/spring 2003 time periods. The figures show PODy and PODn as a function of altitude and represent the CIP-AK diagnoses with a 0.10 threshold.

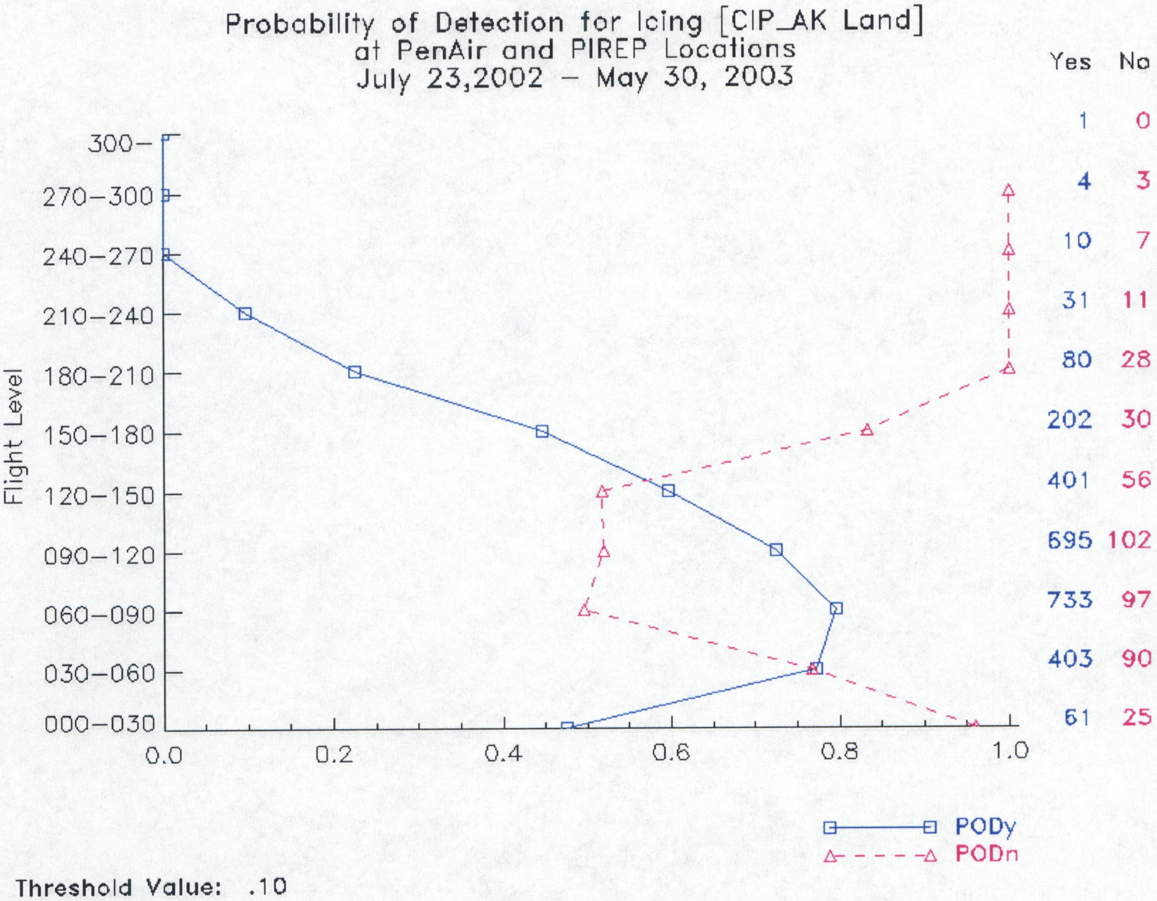

Figure 13. Height series of PODy (solid) and PODn (dashed) as a function of altitude for CIP-AK diagnoses over the CIP-AK Land domain using a 0.10 threshold. The figure includes all observations combining the summer 2002 and winter/spring 2003 period. The two columns on the right show the number of "Yes" and "No" icing observations from each 3,000-ft layer.

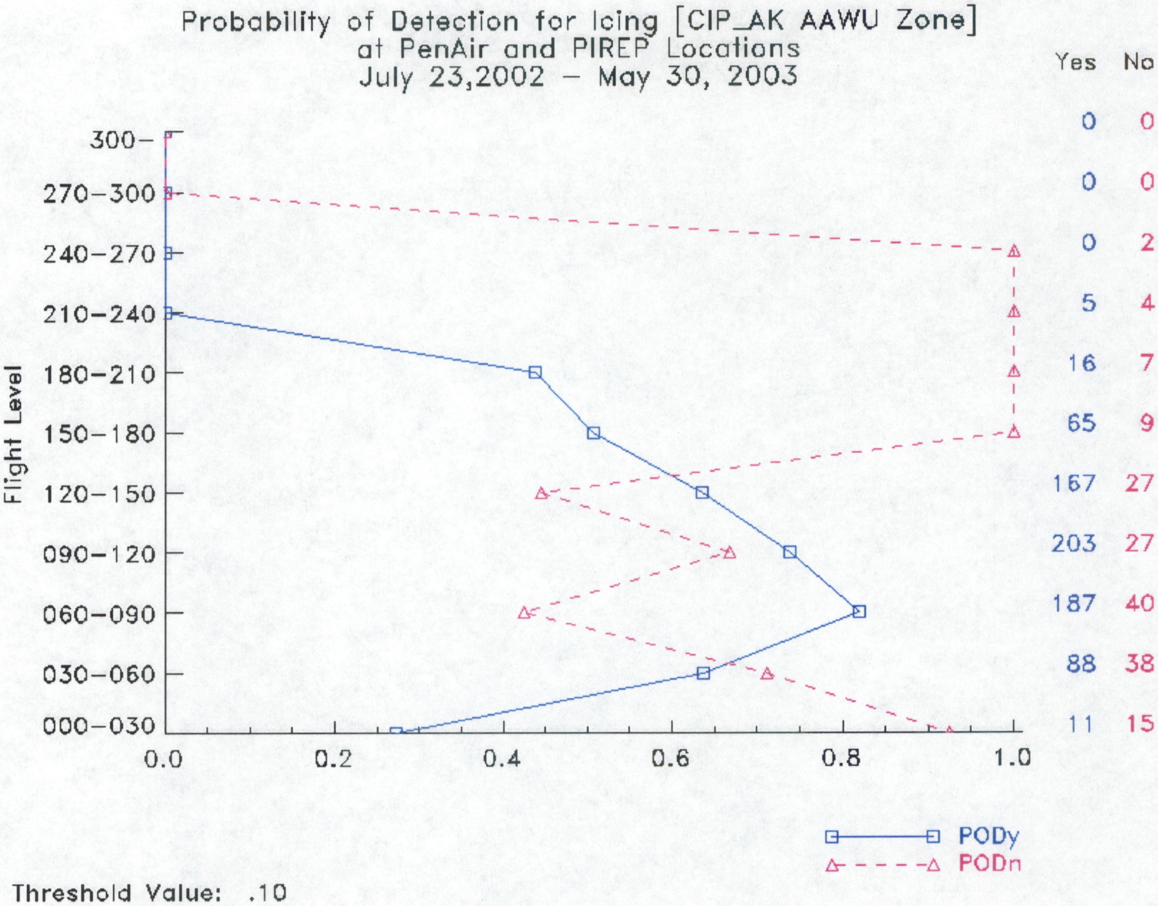

Figure 14. Same as Fig. 13, but for Zones domain.

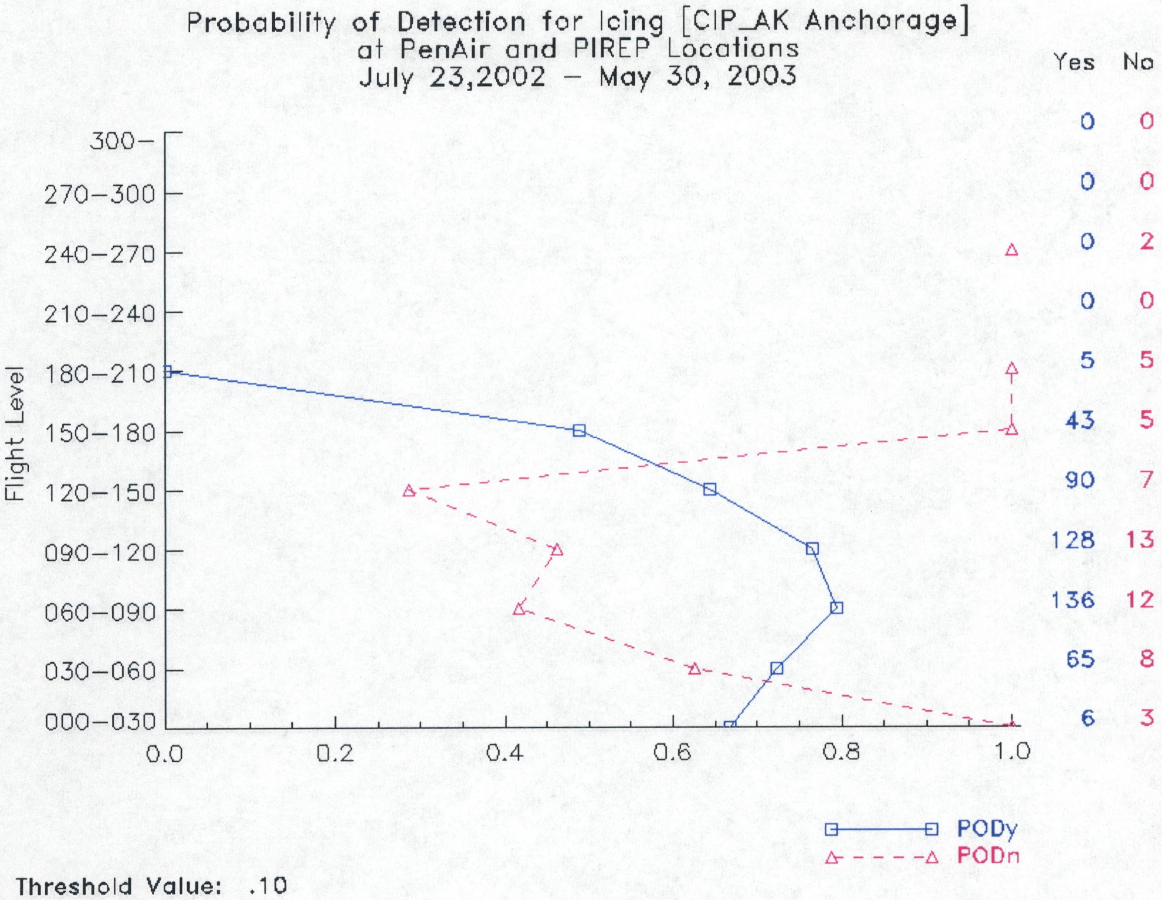

Figure 15. Same as Fig.13, but for Anchorage domain.

Figures 13–15 show that the most favorable values of PODy and PODn are in the 6,000 to 12,000 ft altitude range for all three domains. This is also the altitude range where the most observations are located. In fact, 51–55% of the observations (depending on which domain is chosen) can be found in this 6,000-ft layer between 6,000 and 12,000 ft. Above 15,000 ft there is a rapid dropoff in the number of observations and thus the PODy values go to 0.0 while the PODn values go to 1.0. Below 6,000 ft there also is a decrease in the number of observations. This suggests that CIP-AK has its best performance in the 6,000–12,000 ft layer, where observations are relatively frequent and supercooled liquid water more commonly exists and can lead to icing conditions.

6. SUMMARY

This report has summarized evaluations of icing potential diagnoses provided by CIP-AK. The evaluation covers approximately 30 weeks of time during two times periods: 23 July–11 September 2002 (summer 2002) and 1 January–30 May 2003 (winter/spring 2003). This evaluation built off of previous work concerning the evaluation of CIP over the CONUS (Brown et al. 2001, 2002). The results here suggest that CIP-AK, like CIP, may be a useful tool for providing icing diagnostics. In particular:

- CIP-AK diagnoses are skillful, as measured by their ability to discriminate between Yes and No PIREPs of icing in a fairly distinct volume.
- CIP-AK verified most favorably over the larger domains (CIP-AK Land and Zones). At least part of the reason for the difference is the greater number of observations over the larger domains.
- Even over rather small domains with limited numbers of observations, such as the Anchorage-area domain, the CIP-AK showed skill at diagnosing icing potential.
- CIP-AK verified most favorably between 3,000 and 15,000 ft, and performed especially well in the 6,000 to 12,000 ft layer. Above 15,000 very few observations were available.

The results described in this report are a small fraction of the verification results that are available. For example, a wide variety of verification information for CIP, CIP-AK, other algorithms, and the AIRMETs is available at the RTVS Website (http://www-ad.fsl.noaa.gov/fvb/rtvs/; link FSL-RTVS).

7. ACKNOWLEDGEMENTS

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 and position of the U.S. Government.

We would like to thank the members of the Icing Product Development Team for their support of the independent verification effort over the last several years. The verification groups at FSL and NCAR/RAP provided outstanding support and many important contributions that made this study and report possible. Finally, we express thanks to the AWRP Leadership Team for their support of independent verification activities.

8. REFERENCES

Bernstein, B.C., F. McDonough, M.K. Politovich and C.A. Wolff, 2003: Current Icing Potential – Alaska Version (CIP-AK) Technical Description. *Report to the Aviation Weather Research Program*, FAA. (Available from B. Bernstein, bernstei@ucar.edu)

Bernstein, B.C., F. McDonough, and M.K. Politovich, 2001: Integrated Icing Diagnosis Algorithm – Technical Description. *Report to the FAA Aviation Weather Research Program* (Available from M. Politovich, NCAR, P.O. Box 3000, Boulder CO 80307-3000).

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., J.L. Mahoney, R. Bullock, T. L. Fowler, J. Henderson, and A. Loughe, 2001: Quality Assessment Report: Integrated Icing Diagnostic Algorithm (IIDA). *Report to the FAA Aviation Weather Research Program and the FAA Aviation Weather Technology Transfer Board.* Available from B.G. Brown (bgb@ucar.edu), 36 pp.

Brown, B.G., J.L. Mahoney, and T.L. Fowler, 2002: Verification of the in-flight icing diagnostic algorithm (IIDA). *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR, American Meteorological Society (Boston), 311–314.

Brown, B.G., and G.S. Young, 2000: Verification of icing and icing forecasts: Why some verification statistics can't be computed using PIREPs. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, American Meteorological Society (Boston), 393–398.

Mason., I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30**, 291–303.

McDonough, F. and B. C. Bernstein 1999: Combining satellite, radar, and surface observations with model data to create a better icing diagnosis. *Preprints 8th Conf. On Aviation, Range, and Aerospace Met.*, Dallas TX, 467–471.

NWS, 1991: *National Weather Service Operations Manual*, D-22. National Weather Service. (Available at Website http://www.nws.noaa.gov).

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

APPENDIX A: AIRMET VERIFICATION STUDIES

The icing Airmens' Meteorological Advisories (AIRMETs), which are the operational icing forecasts issued by the AAWU, were included for comparison purposes (i.e., the results presented in this report are not intended to be an evaluation of the icing AIRMETs).

In evaluating an algorithm diagnostic, it is important to compare the quality of the algorithm output to the quality of one or more standards of reference. Thus, the quality of the CIP-AK diagnoses is compared to the quality of the operational forecasts (i.e., AIRMETs). However, it is important to emphasize that the CIP-AK diagnoses is not a forecast like the AIRMETs. Furthermore, CIP-AK and AIRMETs provide different types of information, for different time periods, and have different objectives. CIP-AK diagnoses are generally understood to be valid at a particular time. The AIRMETs, on the other hand, are valid over a 6-h period and are designed to capture icing conditions as they move through the AIRMET area over the period. Because of the differences between AIRMET and CIP-AK information, it is difficult to clearly compare their performance. However, in order to understand the quality of CIP-AK, it is necessary that CIP-AK diagnoses be compared to the operational standard, especially since both types of information will be available to users. The comparisons are made in such a way as to be as fair as possible to both the AIRMETs and CIP-AK while still obtaining the information needed. Nevertheless, users of these statistics should keep these assumptions in mind when evaluating the strengths and weaknesses of each type of product.

AIRMETs: AIRMETs are the operational forecasts of icing conditions. These forecasts are produced by AAWU forecasters every 6 h and are valid for up to 6 h (NWS 1991). AIRMETs may be amended, as needed, between the standard issue times. The forecasts cover prespecified Alaskan zones, with tops and bottoms of the icing regions defined in terms of altitude. Unfortunately, some other more descriptive elements of the AIRMETs cannot be decoded and thus are excluded in AIRMET verification analyses. For comparison with the diagnoses from CIP-AK, the AIRMETs are evaluated over the same time window as the model-based algorithms.

The AIRMETs were decoded to extract the relevant location, altitude range, and other specific information. AIRMETs essentially are dichotomous (that is, icing exists inside the AIRMET region and does not exist outside the AIRMET region).

Table A1 shows the verifcation scores of the AIRMETs for the summer 2002 and winter/spring 2003 time periods. Figures A1–A3 show verification plots for CIP-AK diagnoses with an AIRMET points indicated by the asterisk in each figure. These show the ROC curve (PODy versus 1-PODn), the PODy versus % Area cruve, and the PODy versus % Volume curve. Thus, these plots show where forecast from AIRMETs verify with respect to the CIP-AK diagnoses. Figure A4 is a height series plot for AIRMETs which shows the PODy and PODn distribution with respect to altitude for AIRMETs.

It was noted that the number of CIP-AK diagnoses during the study was greater than the number of AIRMET forecasts for icing during the same period. The smaller number of AIRMETs is reflected in the lower PODy values, and the greater PODn values. Again, this section provides just a basic comparison. CIP-AK diagnoses and AIRMET forecasts provide different types of icing information to the user.

Table A1. The various verification scores for AIRMETs in the Zones domain for both the ummer 2002 and winter/spring 2003 time period.

	PODy	PODn	TSS	% Area	% Volume	Vol. Eff.
AIRMETs	0.15	0.93	0.08	31.9	10.2	1.48

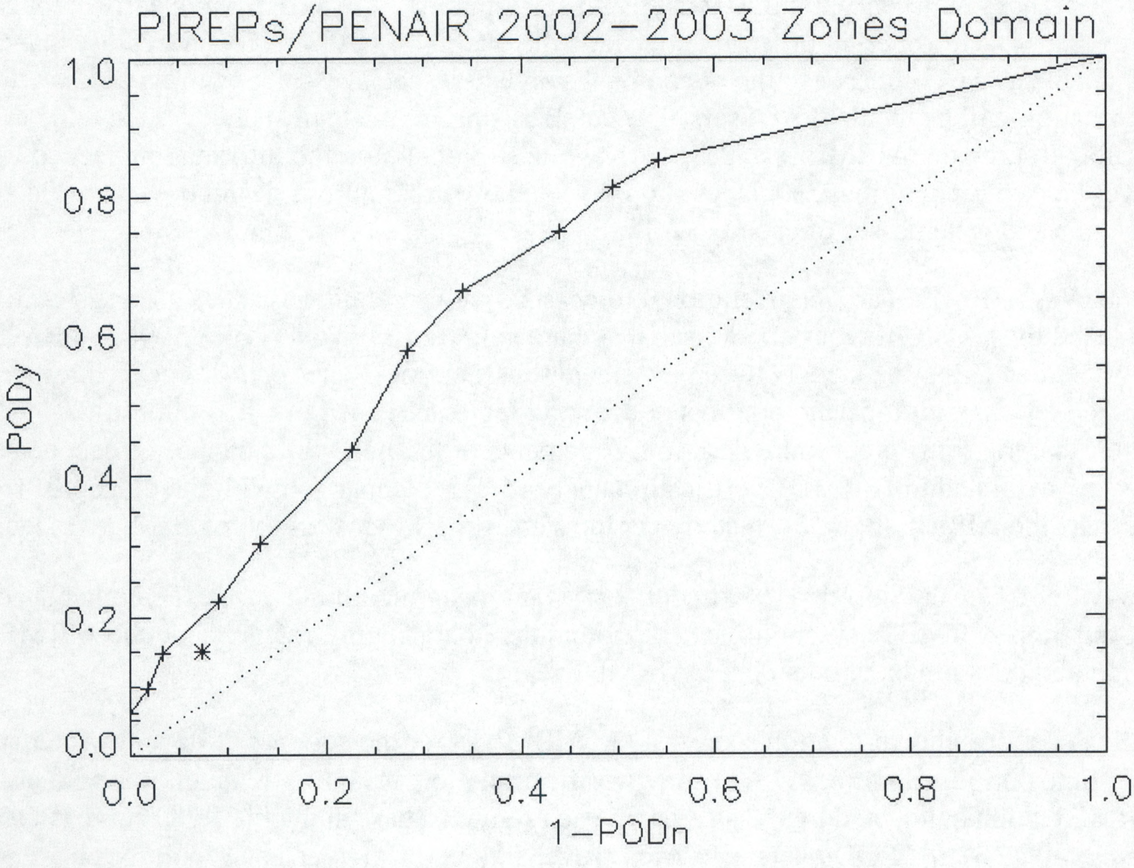

Figure A1. ROC curves (PODy versus 1-PODn) for CIP-AK during summer 2002 and winter/spring 2003, for all valid times combined, based on light and greater PIREPs for CIP-AK Land domain. The AIRMETs point is indicated by the asterisk. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85. The (1,1) point is also included to complete the curve.

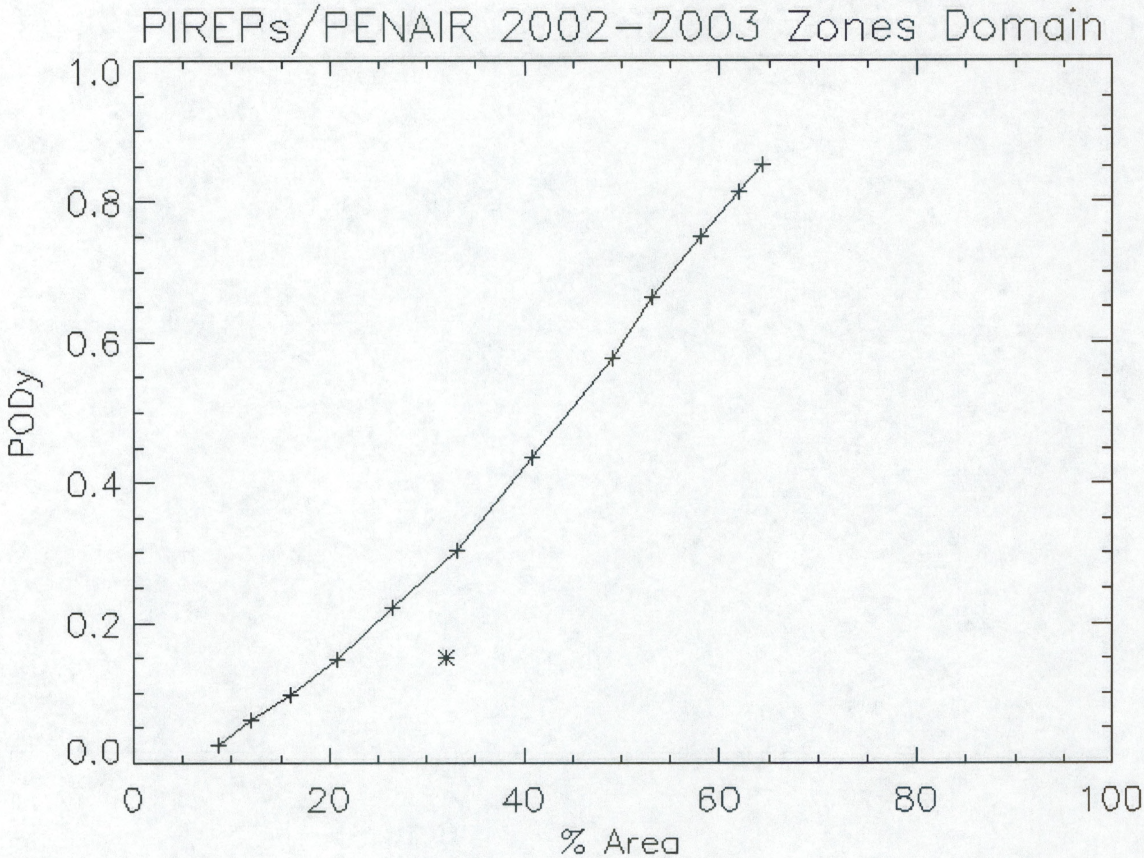

Figure A2. Plot of PODy versus % Area for summer 2002 and winter/spring 2003, for all valid times combined, based on light or greater PIREPs over the CIP-AK Land domain. The AIRMETs point is indicated by the asterisk. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85.

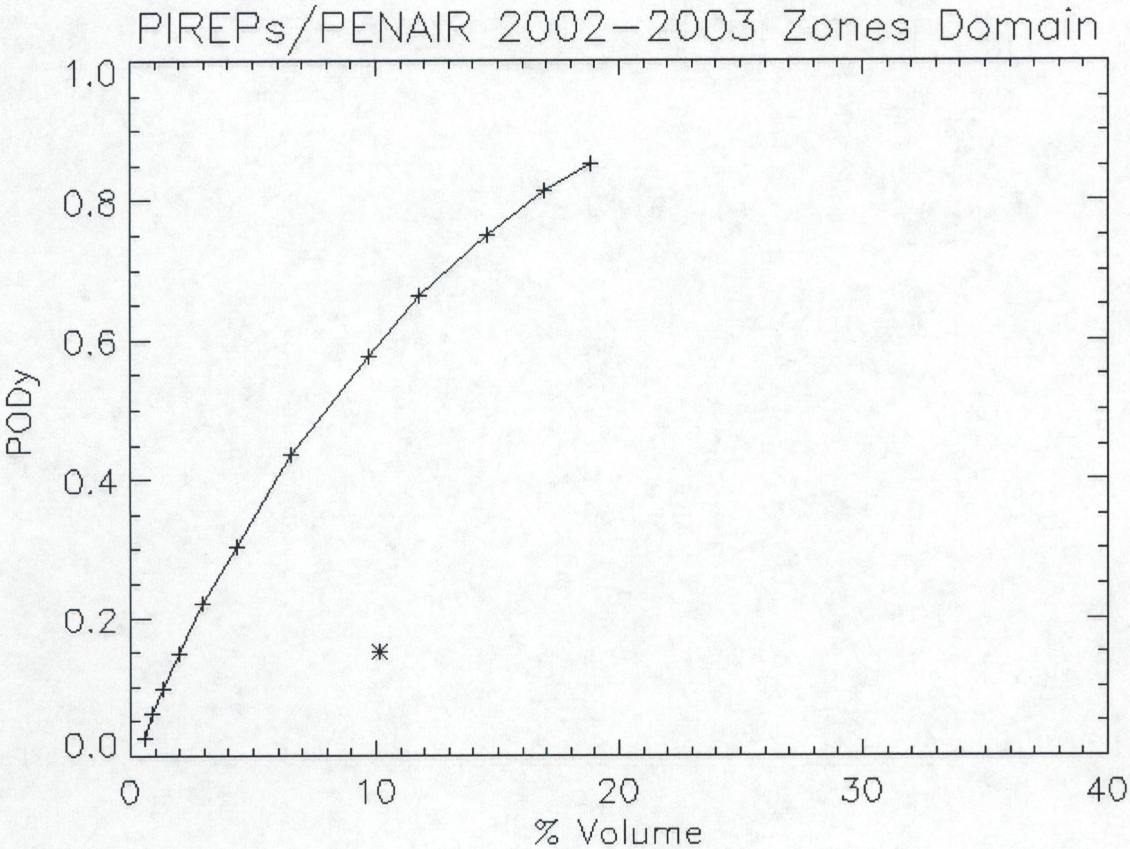

Figure A3. Plot of PODy versus % Volume for summer 2002 and winter/spring 2003 for all valid times combined, based on light or greater PIREPs over the CIP-AK Land domain. The AIRMETs point is indicated by the asterisk. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (starting in the upper right corner) are 0.005, 0.02, 0.05, 0.1, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, and 0.85.

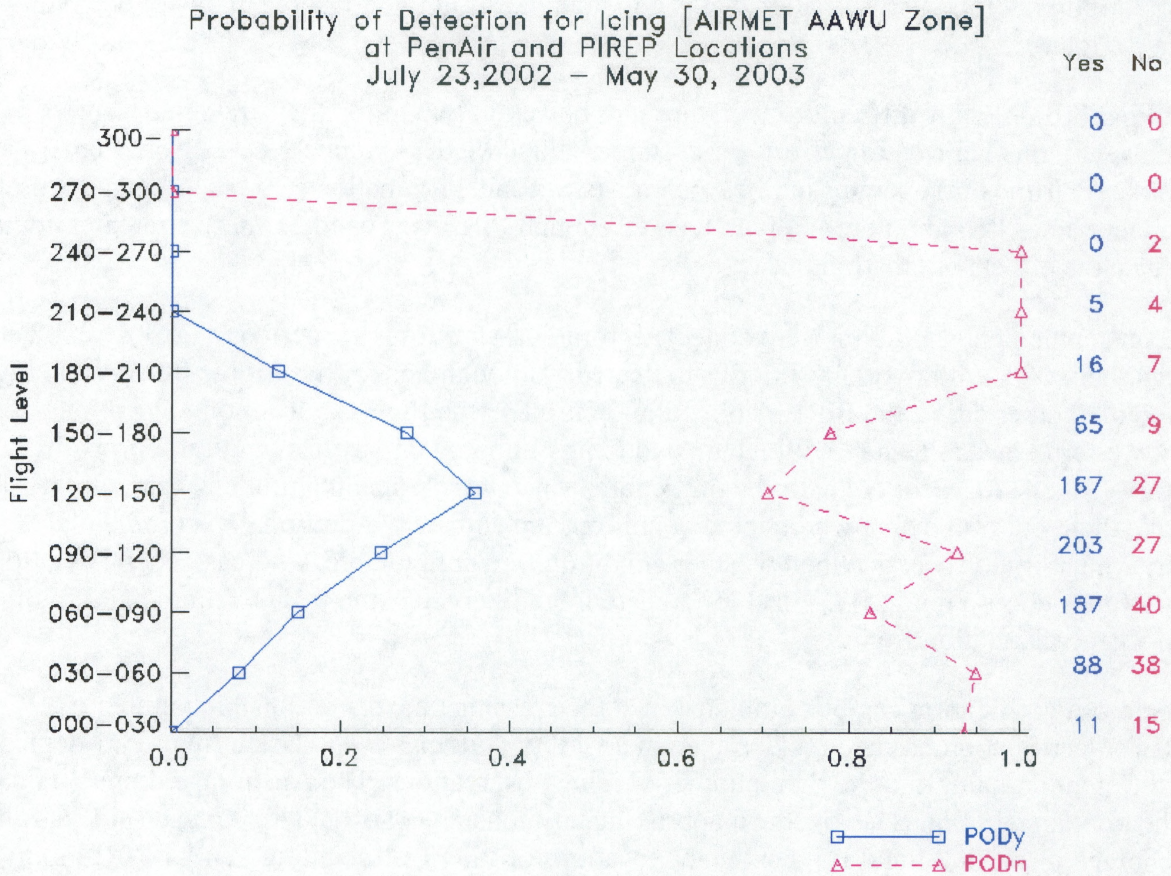

Figure A4. Height series of PODy (solid) and PODn (dashed) for the AIRMETs in the Zones domain for the combined time period of summer 2002 and winter/spring 2003.

APPENDIX B: PIREPs IN ALASKA

PIREPs have traditionally been used for verification of icing forecasts and diagnoses. Unfortunately, PIREPs are relatively scarce in Alaska, but when they are received they tend to be concentrated around a few cities, such as Anchorage, Juneau, and Fairbanks. Lack of PIREPs makes verification of the CIP-AK icing diagnoses difficult and complex. Forecasts and diagnoses can only be verified where observations exist. Consequently, PIREPs have an inherent location bias. Large cities and commonly traveled routes receive a lot of PIREPs, but remote locations have very few. As illustrated in Fig. B1, this location bias is particularly severe in Alaska. The total number of icing reports (including reports of "no icing") and the total number of positive icing reports (not including reports of "no icing") received in each 1 degree by 1 degree latitude-longitude box from 23 July–11 September 2002 are presented in Figs. B2 and B3, respectively. Many places have no PIREPs at all, and only the Anchorage and Fairbanks areas display a reasonable density of PIREPs. Thus, the results of the verification may be biased because diagnoses in remote locations with no observations will never be verified whereas diagnoses near major airports will be verified often. Thus good diagnoses in two locations may not be given equal weight.

The PIREP problem is difficult to work around; however, two mitigative strategies were employed in this verification effort. First, supplemental PIREPs were collected from PenAir, a commuter airline that operates out of Anchorage. Second, verification efforts are being focused in smaller areas that are more likely to receive enough PIREPs. Together, these strategies could mitigate the effects of location bias.

Special supplemental PIREPs were collected from PenAir during the period of 23 July 2002–11 September 2002, with over 340 PIREPs collected. Although PenAir has flights from Anchorage to about 10 other cities, the PIREP collection focused on the flights to/from three locations in southwestern Alaska: Kenai, Dillingham and King Salmon. PenAir flies twin turboprop aircraft on these flights, ideal for collecting icing reports, since their cruising altitude is generally too low to take them out of the area where icing conditions are common. Additionally, commercial aircraft must try to fly even when weather conditions are unfavorable, whereas general aviation aircraft may choose not to fly. Thus, reports are more likely to come from a commercial flight than a general aviation flight.

A localized verification can be completed in some areas that have a reasonable number of PIREPs. Verification has been undertaken in a variety of domains in Alaska. Figure B1 depicts the different domains and the total number of icing observations. The Anchorage domain (small dashed-dot area in Fig. B1) covers an approximately square area of 100 km on each side around Anchorage. Figures B4 and B5 display the locations of PIREPs (including PenAir PIREPs) in the Anchorage verification region for the two time periods of the study. For the winter/spring 2003 time period, there were no special PenAir PIREPs. This region has relatively dense coverage of PIREPs compared with the rest of Alaska. Verification is also undertaken in three of the Alaska AIRMET zones which can be seen as two areas enclosed by the dashed lines on Fig. B1: the Yukon-Kuskokwim Delta region (includes Bethel), the Tanana Valley region (includes Fairbanks), and the Cook Inlet and Susitna Valley region (includes Anchorage).

A graphical representation of the CIP-AK over a limited geographic area is available at http://www.rap.ucar.edu/iida/alaska/ice.html.

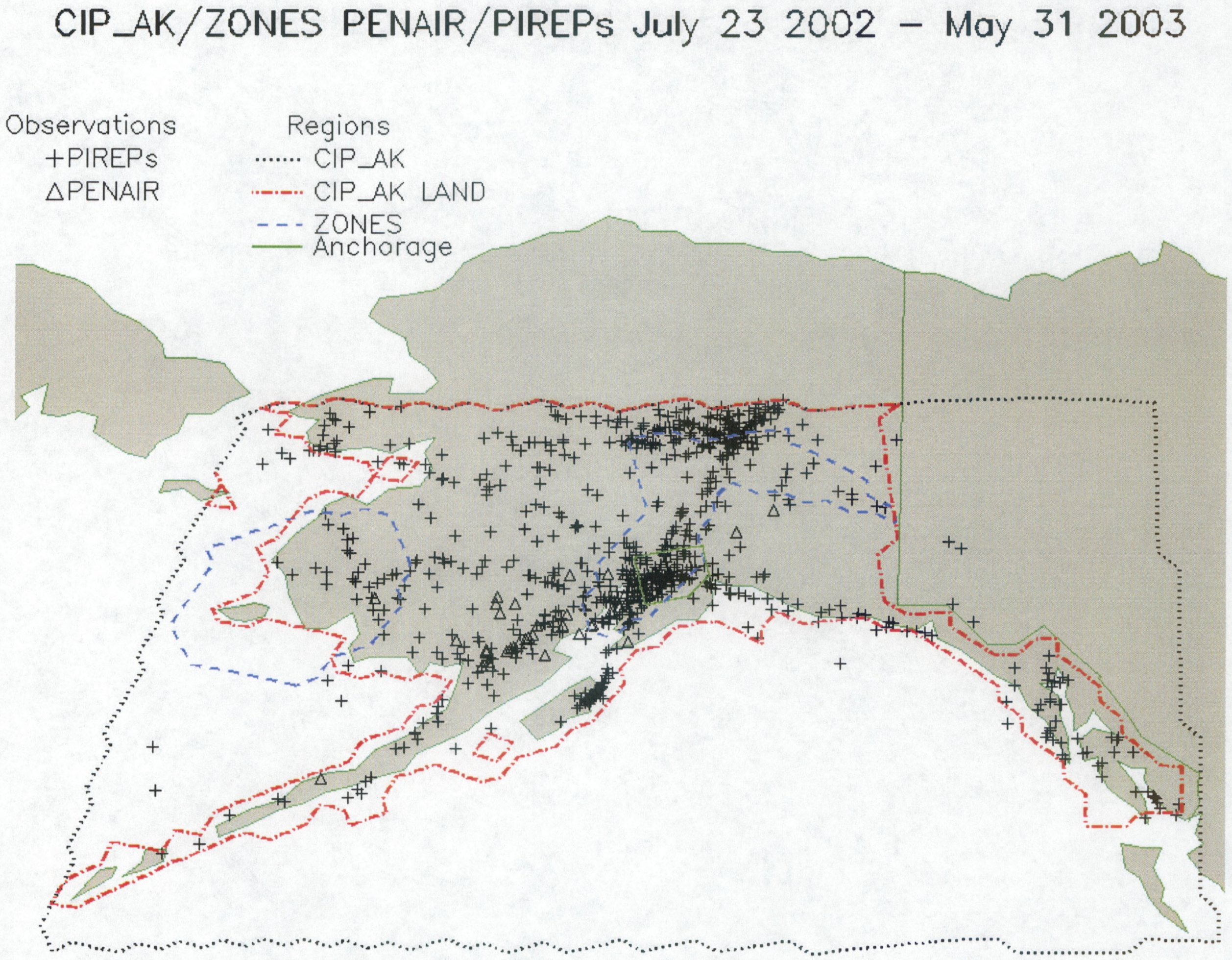

Figure B1. Distribution of observations used in the CIP-AK evaluation. The legend provides information about the type of PIREPs. PenAir PIREPs were only available during the summer 2002 period.

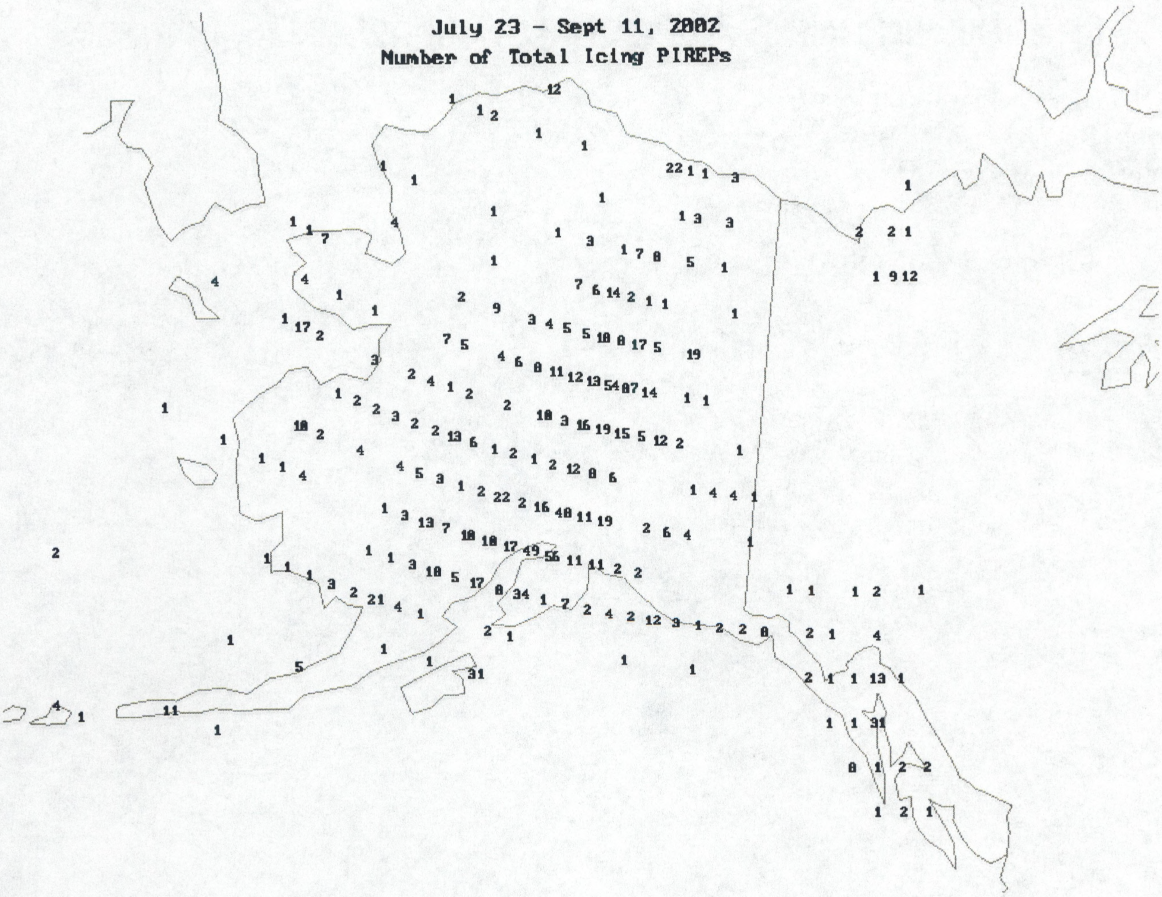

Figure B2. Total number of icing reports (Yes and No) during the summer 2002 period.

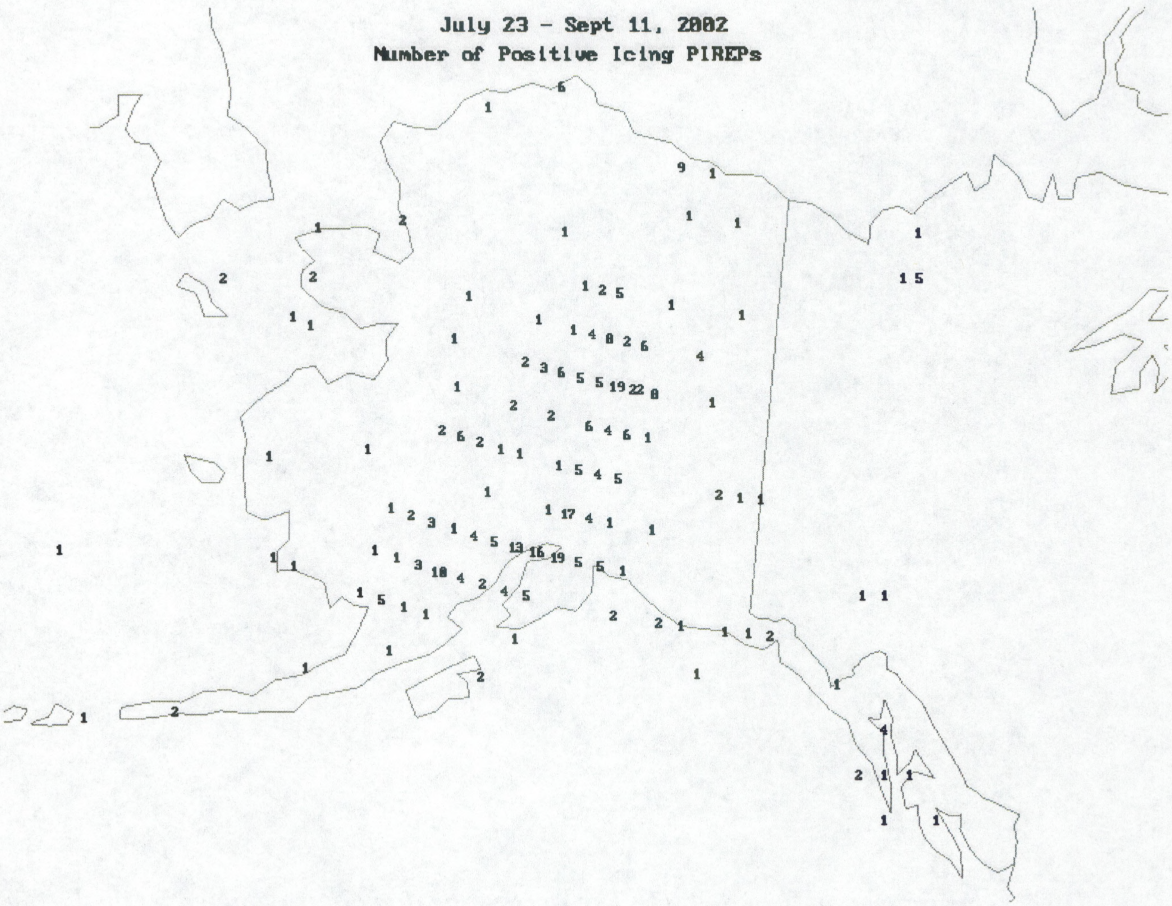

Figure B3. Total number of positive (light or greater intensity) icing observations during the summer 2002 period.

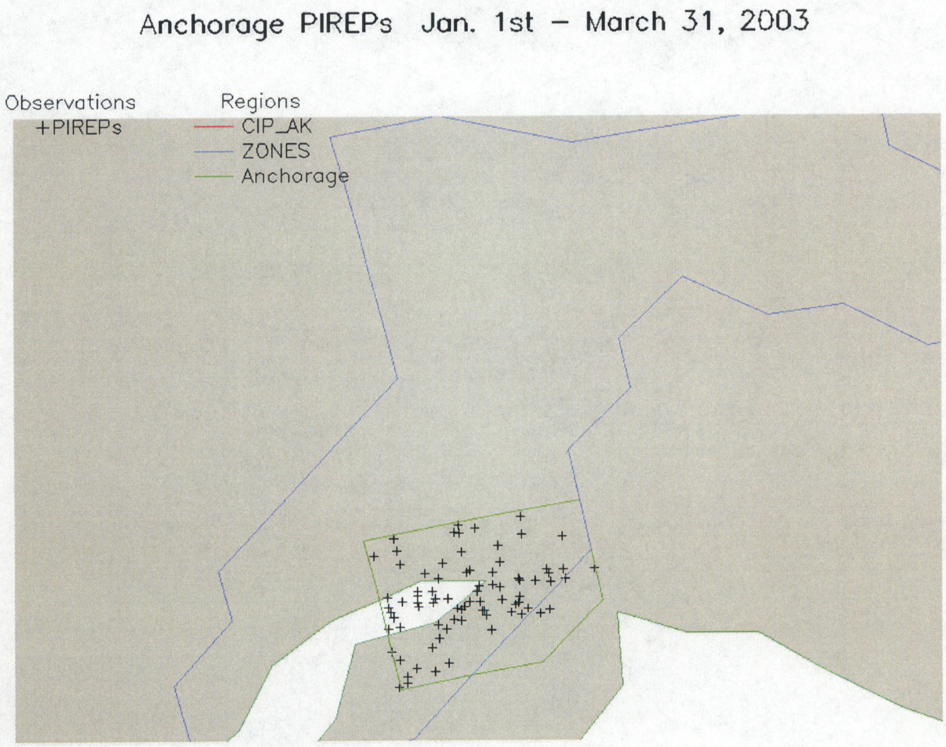

Figure B4. PIREPS during winter/spring 2003. The Anchorage domain is depicted by the box in the lower-middle part of the graphic.

Anchorage PENAIR/PIREPs July 23 – Sept. 11, 2002

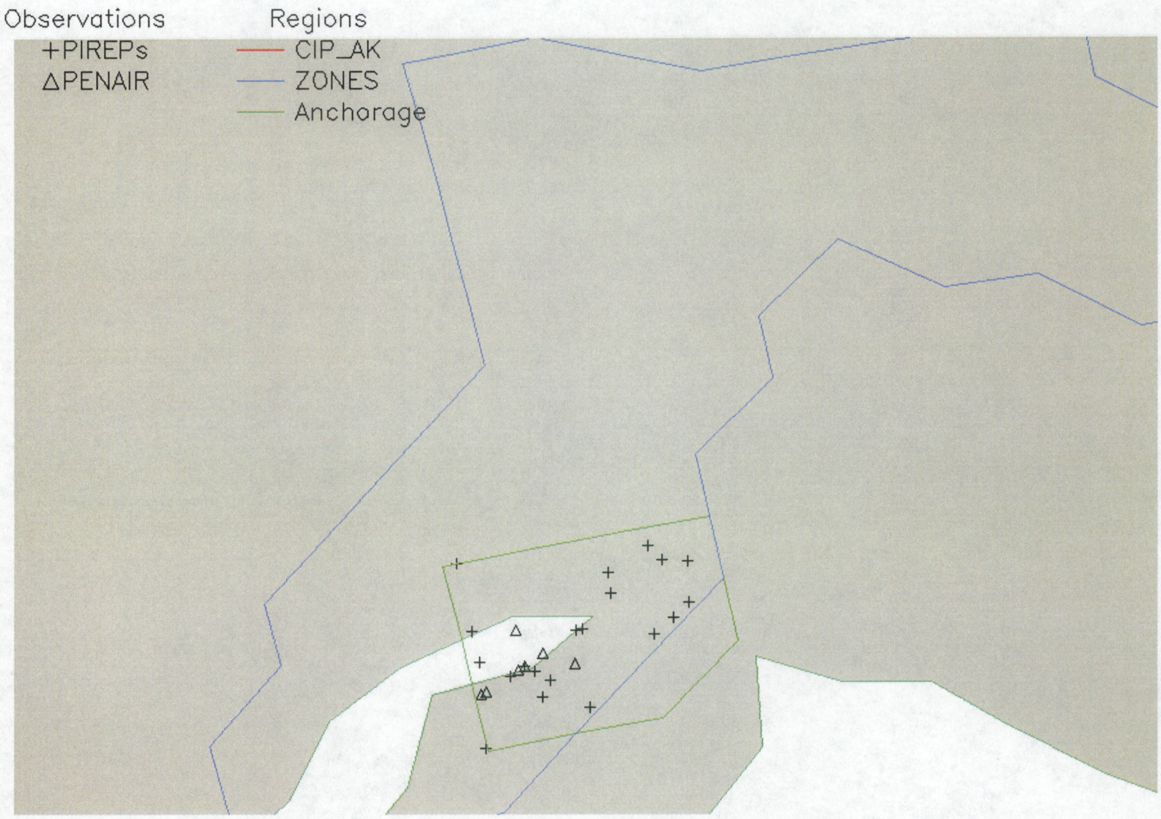

Figure B5. Same as Fig. B4, but for the summer 2002 time period. The observations include special PenAir PIREPs.