Forecast Icing Potential for Alaska (FIP-AK): Quality Assessment Report

Aviation Weather Research Program
Quality Assessment Product Development Team

Jennifer Luppens Mahoney[3], Tressa Fowler [4], Judy Henderson[3], Chris Fischer[2,3], Stacey Seseske[2,3], Barbara G. Brown[4]

31 July 2004

Contents

Forecast Icing Potential for Alaska (FIP-AK): Quality Assessment Report

July 2004

Aviation Weather Research Program
Quality Assessment Product Development Team

Summary

This report summarizes the assessment of forecasts of icing conditions produced by the Forecast Icing Potential for Alaska (FIP-AK). FIP-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 FIP-AK computed for the period 1 October 2003−31 March 2004. The evaluation considers performance of the algorithm over four domains in Alaska where observation concentrations are highest. Voice pilot reports (PIREPs) were used to evaluate the FIP-AK performance. Because of the sparse nature of PIREP observations over Alaska, the verification was augmented with PIREPs over Canada, Washington, and Oregon. Overall results are presented in the report.

Forecasts for the FIP-AK were generated from the Eta numerical weather prediction model. The 3-, 6-, 9-, and 12-hour forecasts from the 0000, 0600, 1200, and 1800 UTC model initialization times were verified using Yes and No icing observations from PIREPs indicating either “light or greater” icing severity or explicitly stating No icing. FIP-AK forecasts were evaluated as a Yes/No forecast of icing 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 the FIP-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 forecast coverage. To provide a standard of comparison, 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. Though several thousand individual FIP-AK forecasts were considered in this evaluation, less than 200 AIRMETs were evaluated. The number of icing observations considered (Yes and No), including each vertical level, in the evaluation ranged from 26,272 for the FIP-AK domain (i.e., the largest of the domains) to 1,025 for the Anchorage-area subdomain.

While most of the analyses focus on the basic icing potential component of FIP-AK, a separate analysis was performed to evaluate the component of the algorithm that predicts the potential for supercooled large droplets (SLD). This analysis was based on SLD PIREPs (i.e., PIREPs reporting freezing precipitation and No-icing PIREPs over the large FIP-AK domain (including parts of Canada, Washington, and Oregon).

Results of the evaluations indicated that:

- FIP-AK is skillful at discriminating between Yes and No icing conditions in a fairly distinct volume.

- FIP-AK verified most favorably over the larger domains (FIP-AK, FIP-AK Land, and Zones). At least part of the reason for this result is the larger number of observations over the larger domains with the inclusion of observations covering the Northwest Pacific region of the CONUS.

- Even over rather small domains with limited numbers of observations, such as the Anchorage-area domain, the FIP-AK showed skill at forecasting icing potential.

- FIP-AK verified most favorably between the surface and 21,000 ft. The algorithm performed especially well at forecasting the No-icing events in the 18,000−24,000 ft layer, although Yes-icing events were not predicted very well at these levels. Above 24,000 ft, very few observations were available. However, the FIP-AK forecast skill for the No icing events decreased in the 3,000 to 6,000 ft layer over for the Alaska domains (FIP-AK Land and Zones).

- In areas with enough information to calculate SLD, the FIP-AK correctly identifies 43% of the reported SLD conditions and 86% of the reported No-icing conditions. This is excellent performance for a forecast designed to detect conditions that are relatively rare.

1. Introduction

This report summarizes basic results of an evaluation of the Forecast Icing Potential algorithm for Alaska (FIP-AK). This algorithm is under consideration for transition to experimental status through the Aviation Weather Technology Transfer (AWTT) process. FIP-AK is designed to forecast in-flight icing conditions over Alaska. Through real-time forecasting exercises the Inflight Icing Product Development Team has subjectively evaluated the FIP-AK over several seasons. This study includes an objective evaluation, performed by the Quality Assessment Product Development Team (QAPDT) of the Federal Aviation Administration's Aviation Weather Research Program (FAA/AWRP), which follows those verification techniques used to evaluate the quality of the CONUS Forecast Icing Potential (FIP) and the Current Icing Potential for Alaska (CIP-AK) algorithms (Quality Assessment Product Development Team 2003; Kelsch et al. 2003).

The analyses presented in this report focus primarily on evaluations of FIP-AK forecasts from 1 October 2003–31 March 2004. This period represents a combined total of roughly 24 weeks. The evaluation considers performance of the algorithm over four domains in Alaska where observation concentrations are highest. The FIP-AK domain was expanded to cover Alaska, portions of Canada, Washington, and Oregon, and much of the surrounding ocean area in order to increase the number of voice pilot reports (PIREPs) used to evaluate the algorithm, while the other three domains cover mainly Alaska. While most of the analyses focus on the basic icing potential forecasts, the component of the algorithm that predicts the existence of supercooled large droplets (SLD) was also considered in some analyses over the expanded FIP-AK domain.

Complementary results for the Airmens' Meteorological Advisories (AIRMETs, NWS 1991), 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 FIP-AK.

The report is organized as follows. The study approach is presented in Section 2. Section 3 briefly describes the FIP-AK 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 are provided in Section 6. Appendix A contains verification results for the AAWU AIRMETs.

2. Approach

The FIP-AK forecasts are based on output provided by the National Centers for Environmental Prediction (NCEP) Eta weather prediction numerical forecast model, which runs operationally over Alaska. Forecasts issued at 0000, 0600, 1200, and 1800 UTC with lengths of 3, 6, 9, and 12 hours were evaluated in the verification study. Statistics for FIP-AK were computed for four regions, which are shown in Fig. 1. These regions are, from largest to smallest, (1) the *FIP-AK domain,* (2) the *FIP-AK Land domain* where oceanic areas are removed due to lack of observations, (3) the *Zones domain* where three AAWU forecast zones are located, and (4) the *Anchorage* domain, which includes relatively high-traffic areas around Anchorage. In order to increase the number of PIREPs for verification, the largest domain used to verify the FIP-AK was extended to include parts of Canada, Washington, and Oregon. The icing regime over these areas is similar to the icing regime over southern Alaska thus, including these areas outside the Alaska domain is reasonable.

The verification approach is identical to the approach taken in previous studies associated with the CONUS versions of the CIP and FIP, as well as the CIP-AK (e.g., Brown et al. 2001; Quality Assessment Product Development Team 2003; Kelsch et al. 2003). In particular, the algorithm forecasts were verified using Yes and No PIREPs of icing. The algorithm forecasts 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 was utilized to examine the full range of performance of the algorithm. The Yes/No forecasts were evaluated using standard dichotomous verification techniques available for Yes/No forecasts, where the verification observations are based on PIREPs. In addition, the amount of airspace impacted by the forecasts was considered. For most of the analyses, PIREPs reporting at least “light to moderate” icing severity were included as Yes reports.

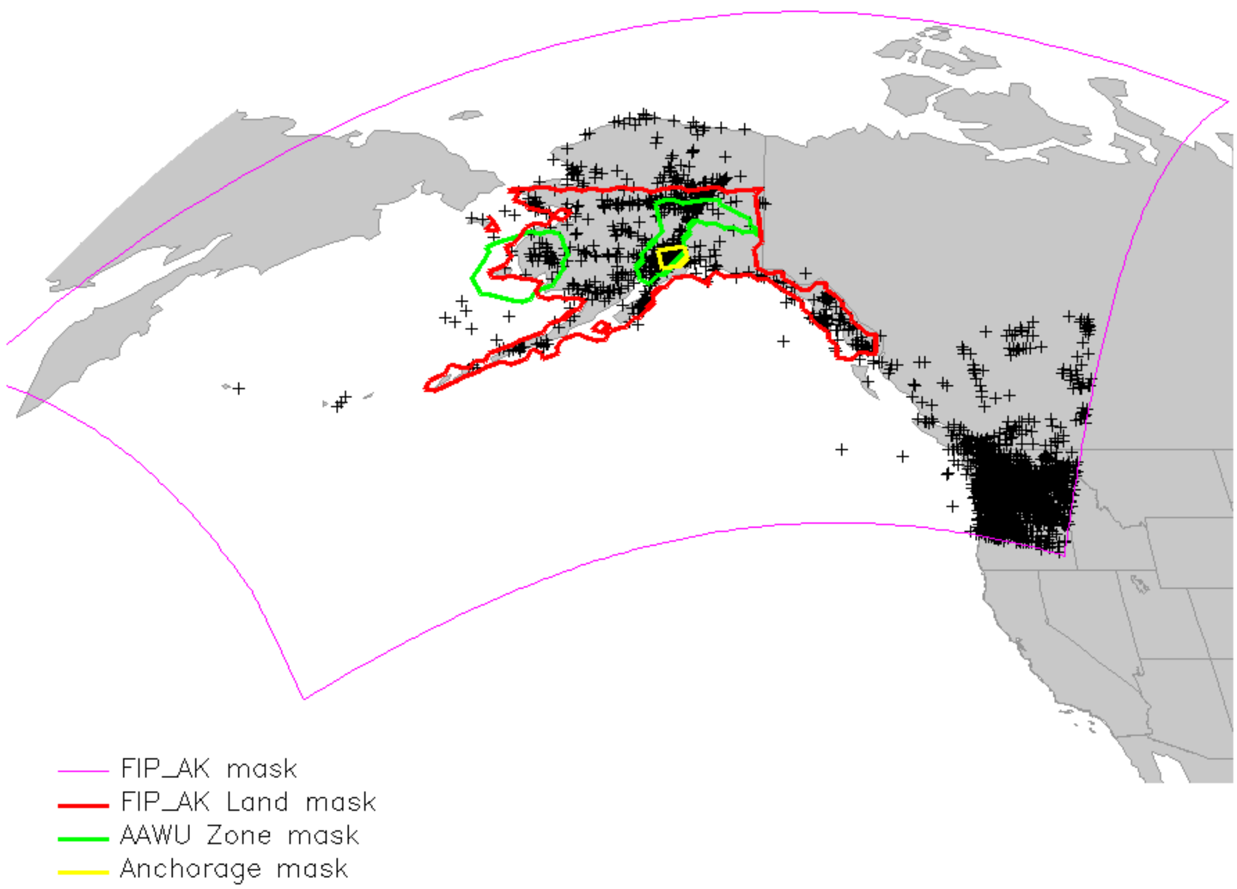

Figure 1. FIP-AK domains used in the evaluation: (1) FIP-AK domain (purple) (2) FIP-AK Land (red), (3) Zones (green), and (4) Anchorage (yellow). The PIREPs are shown as black “+”. PIREPs over Canada, Washington, and Oregon were included in the FIP-AK domain.

Analyses were stratified by forecast domain, height, and forecast lead time. However, to ensure an adequate number of observations for evaluating the FIP-AK, the results presented in this report incorporate the statistical information from the entire evaluation period into overall statistical results.

The evaluation of the SLD component of the algorithm, PIREPs that mention freezing drizzle (FZDZ) or freezing rain (FZRA) and report icing conditions of any intensity are considered Yes SLD observations, while PIREPs stating explicitly stating No icing are considered the No-icing observations for SLD evaluation. Due to the limited number of FZDZ/FZRA PIREPs, the SLD forecast from the FIP-AK is evaluated over the entire FIP-AK domain.

3. FIP-AK and PIREPs

FIP-AK**:** An in-flight icing forecast algorithm, FIP-AK, was developed for the Alaska air space (McDonough and Bernstein 1999; McDonough et al. 2004) and was modeled after the FIP for the CONUS. The algorithm produces icing potential forecasts in the range 0-1, which predict the likelihood of both general icing and SLD icing conditions. The FIP-AK algorithm uses output directly from the NCEP Eta numerical weather prediction model, which runs at 45-km horizontal resolution with 39 vertical pressure levels. The Eta provides atmospheric variables at 3-D grid points, including winds, temperature, and moisture fields, as well as cloud microphysics for predicting freezing and subfreezing liquid. The model is initialized every 6 hours and generates 3-hourly forecasts out to 12 hours, as well as 6-hourly forecasts out to 24 hours. During the evaluation, 2,219 FIP-AK forecasts for the 3, 6, 9, and 12 hour forecasts from the 0000, 0600, 1200, and 1800 UTC model initialization times were verified. Computation of the SLD component required specific information about the existence of freezing precipitation; unfortunately, it was frequently the case that sufficient information to calculate the SLD field was not available. In these cases, the FIP-AK SLD value was set to missing. The SLD field was verified separately from the icing potential field.

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 icing intensity of at least "light to moderate" were used to evaluate the FIP-AK. In addition, only reports explicitly stating No icing were used as No observations. The PIREP distribution used to assess the FIP-AK is shown in Fig. 1. More information on PIREPs in Alaska can be found in Kelsch et al. (2003). Overall, 26,272 PIREPs, including all vertical levels within a single PIREP either reporting icing or no-icing conditions, were used to verify the FIP-AK forecasts.

4. Verification methods

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

4.1. Matching methods

The methods used to match the PIREPs to the FIP-AK forecasts are the same as have been used in previous evaluations of FIP, CIP and other in-flight icing algorithms (e.g., Brown et al. 1997, 2001, 2002). In particular, each PIREP is connected to the FIP-AK forecasts at the nearest 8 grid points (four surrounding grid points, two levels vertically). A bi-linear interpolation is used to compute the appropriate FIP-AK value at each PIREP location. In order to increase the number of PIREPs for verification, a time window of $\pm$4 hours surrounding the algorithm valid time was used to evaluate the algorithm output.

The SLD PIREPs, reporting FZDZ or FZRA, were also matched to the eight surrounding gridpoints. However, the algorithm output was not interpolated to the PIREP location. Instead, in general, the maximum non-missing FIP-AK SLD value from the surrounding gridpoints was matched to the SLD PIREPs. Similar procedures were used for the No-icing PIREPs.

4.2. Statistical verification methods

The statistical verification methods used to evaluate the forecast skill for the FIP-AK 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.

FIP-AK icing forecasts and the PIREP observations are treated as dichotomous (i.e., Yes/No) values. The FIP-AK icing potential values are converted to a variety of Yes/No values by application of various thresholds for the occurrence of icing. For example, when the threshold of 0.10 is applied, FIP-AK values ≥ 0.10 are treated as "Yes" forecasts. Thus, the basic verification approach makes use of the dichotomous contingency table (Table 1) to represent the joint distribution of icing forecasts and observations where the forecasts are represented by the rows and the columns represent the 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 contingency 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 (CSI) 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 forecast by FIP-AK*.

Together, PODy and PODn measure the ability of the FIP-AK forecasts 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 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 forecast. The Area indicates the proportion of the surface area of the domain that is associated with a Yes icing forecast at some level above the surface. The VE considers PODy relative to the volume covered by the forecast, and can be thought of as the PODy 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, it can be 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 algorithm forecast quality.

Plots of PODy vs. Area (Volume) show the relationship between PODy and Area (Volume) for various thresholds. For the PODy vs. Area (Volume) plots, threshold values that result in optimal icing forecasts 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 vs. 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.

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 forecast with icing by FIP-AK. ROC curve areas, based on PODy and PODn, will also be considered as a measure of the overall skill of FIP-AK at discriminating between Yes and No observations.

Evaluation of the FIP-AK SLD forecasts was also based on the 2x2 contingency table (Table 1). However, only one threshold was applied to the FIP-AK SLD values, and the verification statistics were limited to PODy and PODn.

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

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

5. Results

Basic results of the FIP-AK evaluations described here are organized in subsections that describe the overall performance and the performance by altitude and forecast lead time. Basic results of the SLD forecast evaluation are also included.

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 to 1 Best: 1 Worst: 0
PODn	NN/(YN+NN)	Probability of Detection of No observations	Proportion of No observations that were correctly forecasted	0 to 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
ROC 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 to 1 Best: 1 No skill: 0.5
Area (%)	[(Forecast Area) / (Total Area)] x 100	Percent of the total area (e.g., Alaska) that has a Yes forecast at some level above	Percent of the area that is impacted by a Yes forecast at one or more flight levels above	0 to 100 Smaller is better
Volume (%)	[(Forecast Vol) / (Total Vol)] x 100	Percent of the total air space volume that is impacted by the forecast	Percent of the total air space volume that is impacted by the forecast	0 to 100 Smaller is better
Volume Efficiency (VE)	(PODy x 100) / % Volume	PODy (x 100) per unit % Volume	PODy relative to airspace coverage	0 to infinity Larger is better

5.1. Overall results

Overall verification results for the FIP-AK for 1 October 2003–31 March 2004 are shown in the curves (PODy vs 1-PODn, PODy vs Area, and PODy vs Volume) presented in Figs. 2−4. The results are based on light-to-moderate and greater

PIREPs, in which all pairs of FIP-AK forecasts issued at 0000, 0600, 1200, and 1800 UTC with lead times of 3, 6, 9, and 12 hours are combined with the corresponding observations. The results are presented individually for each domain. Each point on the curves 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. For all three plots, better performance is indicated by lines that are closer to the upper left-hand corner. For the ROC plots (Fig. 2; PODy vs. 1-PODn), positive forecast skill in discriminating between Yes and No observations is indicated when the ROC curve lies above the diagnonal line. The area under the ROC curve summarizes this skill, where areas greater than 0.5 indicate positive forecast skill.

The results in Fig. 2 suggest that FIP-AK has skill in distinguishing between Yes and No icing conditions for all four domains, as indicated by the location of the curves to the left of the no-skill diagonal line. The ROC area values, measuring the areas under the curves for the four domains shown in Fig. 2, are presented in Table 3. These statistics confirm that the FIP-AK has positive forecasting skill at correctly classifying Yes and No observations.

The tendency for the curves in Figs. 2−4 to approach the upper left is more pronounced for the FIP-AK, FIP-AK Land, and Zones domains than it is for the Anchorage domain. As was the case in the CIP-AK evaluation (Kelsch et al. 2003), the result is likely due, at least in part, to the small number of PIREPs available for verification of the FIP-AK over the Anchorage domain rather than a reflection of poor forecast quality.

Figures 3 and 4 show the relationships between PODy and Area and Volume, respectively. These plots measure the tradeoffs between values of increasing Area and Volume with increasing values of PODy. For all domains, the Volume plots are quite different from the Area plots, with larger Area tradeoffs associated with increases in PODy. This result is due to the fact that the Area is based on the forecast area at each level projected onto one plane. Therefore, small volumes forecast at different locations may add up to a relatively small Volume but a very large Area, giving the impression of overforecasting the horizontal extent of icing conditions. Both Figs. 3 and 4 indicate that the best performance, according to these measures, is associated with the FIP-AK domain and the worst is associated with the Anchorage domain.

Tables 4-7 present tabular results for the four domains. The rows indicate the 12 thresholds used in FIP-AK, and the columns provide the statistical measures as defined in Tables 1–2.

The results presented in Tables 4 for the FIP-AK domain show that for a threshold of 0.005 where 17.4% of the domain volume is considered to have a Yes forecast of icing, and the PODy, PODn and TSS values are nearly 0.85, 0.47, and 0.32, respectively. However, for thresholds of 0.85 when only 0.62% of the domain volume is covered by an icing prediction, the PODy, PODn, and TSS values are 0.05, 0.98, and 0.03, respectively. A rough comparison with statistics computed for CIP-AK at a threshold of 0.1 (Kelsch et al. 2003; PODy = 0.68, PODn = 0.66, TSS = 0.34, and Volume = 12.5%) for the FIP-AK Land domain (Table 5) indicate very similar results where skill for the FIP-AK at 0.1 threshold is PODy = 0.75, PODn = 0.60, TSS = 0.35, and Volume = 14%. These results indicate that the FIP-AK algorithm has good skill at forecasting icing conditions in Alaska, and that this skill is comparable to the skill of the CIP-AK algorithm.

Figure 5 shows a scatterplot of daily PODy vs. Volume per forecast lead time for observations and FIP-AK forecasts (threshold 0.1) that were available during the evaluation period. These results indicate very little correlation between forecast skill and the percentage of airspace volume covered by the icing potential from the FIP-AK algorithm. For instance, forecasts with PODy values of 0.2 and 0.8 often had similar Volumes of nearly 14%. These results suggest a large amount of variability in the daily statistics, thus supporting the decision to compute the verification statistics over longer time periods.

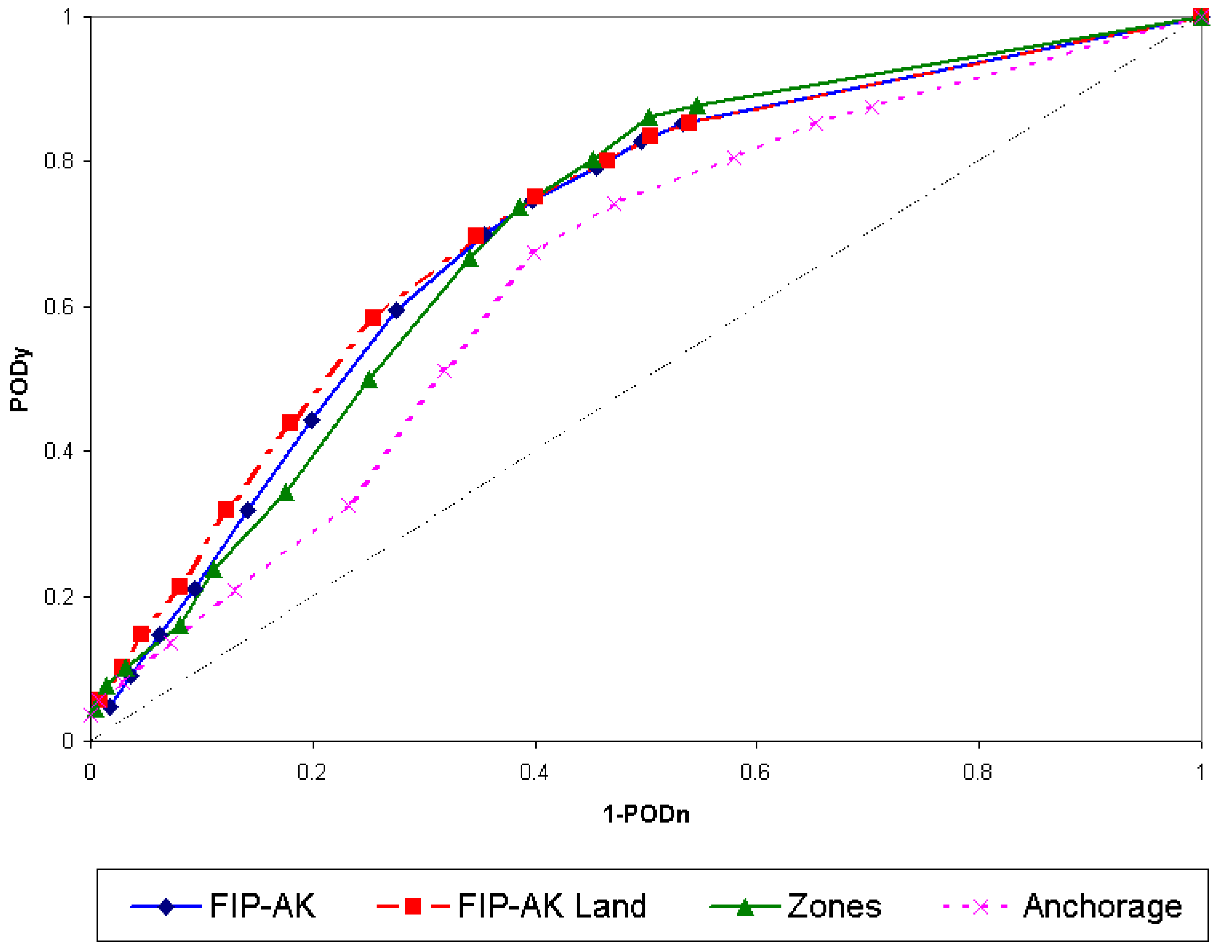

Figure 2. ROC curves (PODy vs. 1-PODn) for FIP-AK from 1 October 2003–31 March 2004, for all forecast issue and lead times combined, based on light-to-moderate and greater PIREPs for all four domains. Each point on the curve represents a different threshold used to define Yes/No icing. The thresholds (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.

Table 3. The area under the ROC curve for each domain.

Domain	**Area Under the ROC**
FIP-AK	0.708
FIP-AK Land	0.717
Zones	0.706
Anchorage	0.645

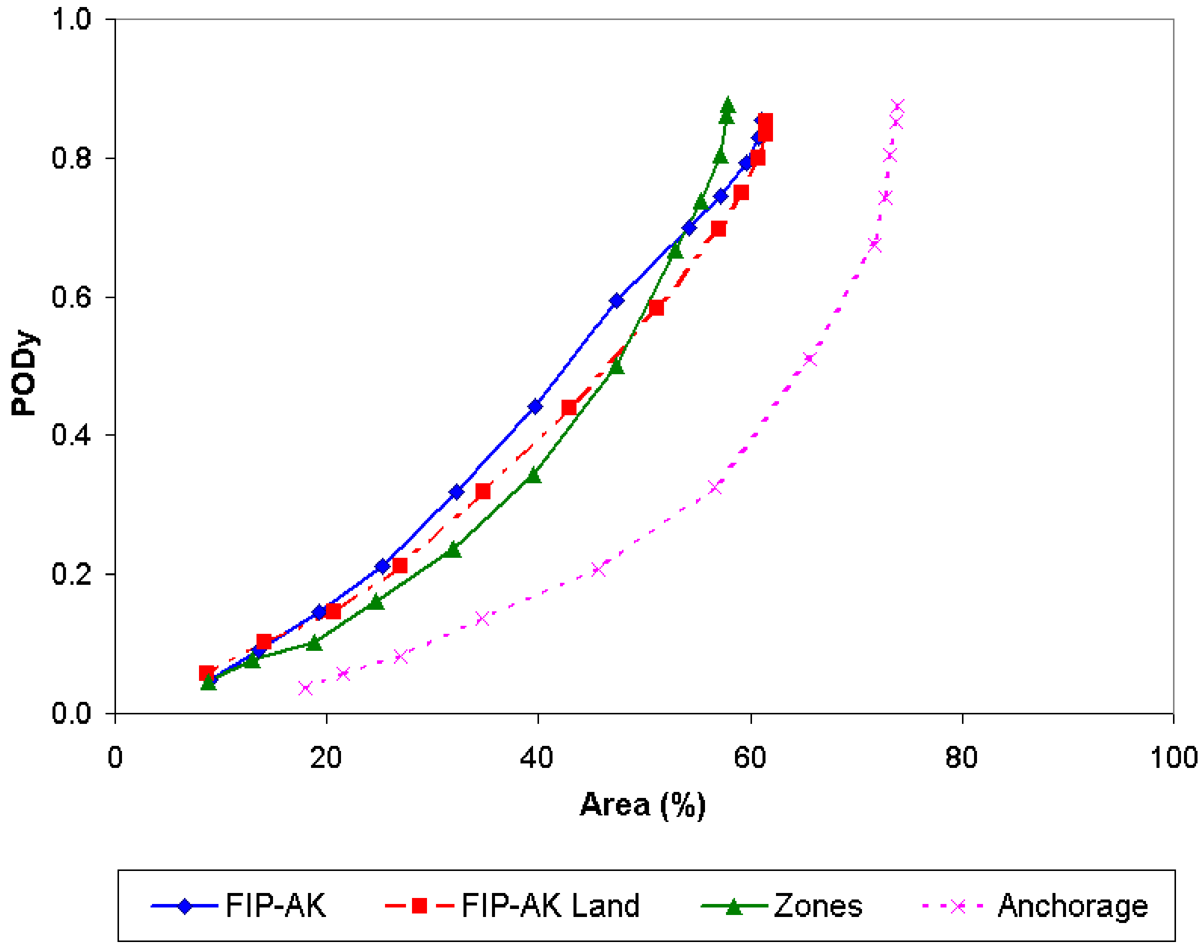

Figure 3. Same as Fig. 2, except for PODy vs Area (%).

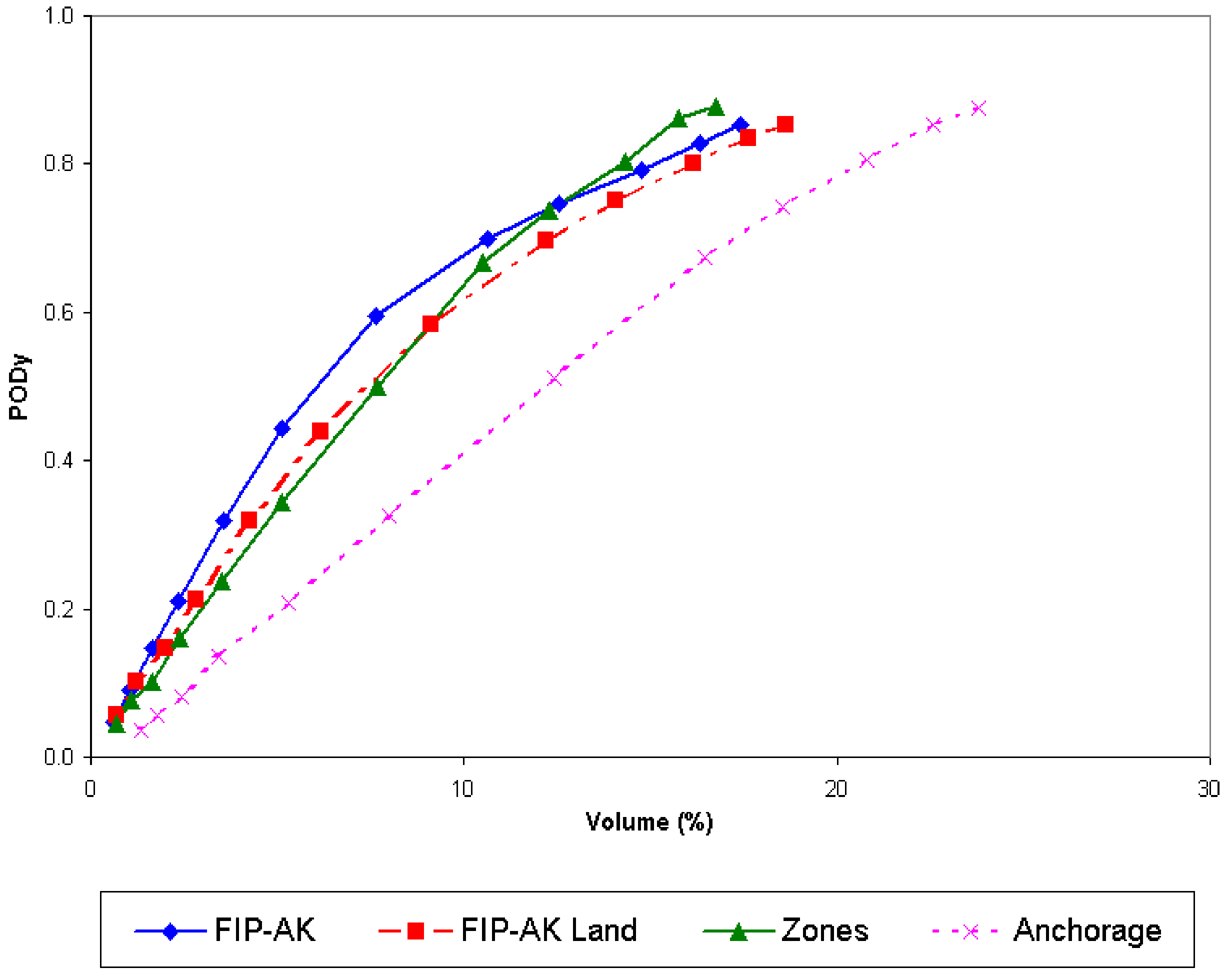

Figure 4. Same as Fig. 2, except for PODy vs Volume (%).

5.2. Comparisons by altitude

To assess the performance of FIP-AK forecasts at different altitudes, verification statistics were computed for the period 1 October 2003–31 March 2004 for each 3,000 ft interval from the surface to 30,000 ft. These statistics are presented in Figs. 6−9 for the FIP-AK, FIP-AK Land, and Zones domains. The figures show PODy and PODn (Fig. 6−8) and TSS (Fig. 9) as a function of altitude and represent the FIP-AK forecasts at the 0.10 threshold.

Figures 6-8 show that the greatest number of PIREPs reporting either light to moderate and greater icing or No icing, which occurred between 3,000 and 21,000 ft. Above and below these levels, the number of observations drop off dramatically, which has at least some impact on the skill of the forecast. The best forecast skill for all three domains occurs between the surface and 21,000 ft where the PODy values range from 0.9 to 0.7 (surface) and 0.6 to 0.5 (21,000 ft). Within this vertical range, the PODn values are typically between 0.8 and 0.6. However, at the 3,000-6,000 ft layer for the FIP-AK Land and Zones domains (which are limited to the Alaska region), the PODn drops off considerably to values less than 0.4. A similar, but less dramatic decrease is noted in the FIP-AK domain, but occurs within the 6,000–9,000 ft layer. The decrease in PODn between 3,000 and 6,000 ft was the result of 20% (FIP-AK Land domain) to 26% (Zones domain) of the forecasts indicating a potential for icing while No icing was reported in the PIREP observation. In the 18,000−21,000 and 21,000–24,000 ft layers, the FIP-AK algorithm appropriately classified almost all of the No-icing observations. For instance, out of 155 cases (including 43 No-icing cases) for the FIP-AK Land and 66 cases (including 10 No-icing cases) for the Zones domains for the 18,000–21,000 ft layer, the algorithm never produced a forecast for icing when No icing was observed. Similar results were noted for the 21,000−24,000 ft layer.

In summary, these results for all domains indicate that the FIP-AK is skillful at forecasting icing potential between the surface and 21,000 ft, and was particularly skillful at forecasting No icing within the 18,000−24,000 ft layer. However, within the FIP-AK Land and Zones domains, the FIP-AK had some difficulty adequately discriminating between Yes and No icing potential in the 3,000−6,000 ft layer.

Table 4. Overall statistics for the FIP-AK domain for 1 October – 31 March 2004 using light-to-moderate PIREPs for verification. The rows represent the specific FIP-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	16292	3827	2798	3355	0.85	0.47	0.32	60.99	17.40	4.90
0.02	15814	3558	3276	3624	0.83	0.50	0.33	60.71	16.33	5.07
0.05	15134	3272	3956	3910	0.79	0.54	0.34	59.69	14.74	5.38
0.1	14235	2857	4855	4325	0.75	0.60	0.35	57.24	12.53	5.95
0.15	13356	2546	5734	4636	0.70	0.65	0.35	54.22	10.64	6.58
0.25	11359	1978	7731	5204	0.60	0.72	0.32	47.42	7.65	7.78
0.35	8457	1425	10633	5757	0.44	0.80	0.24	39.71	5.14	8.61
0.45	6100	1019	12990	6163	0.32	0.86	0.18	32.26	3.54	9.02
0.55	4027	675	15063	6507	0.21	0.91	0.12	25.24	2.36	8.95
0.65	2795	442	16295	6740	0.15	0.94	0.08	19.32	1.65	8.89
0.75	1749	255	17341	6927	0.09	0.96	0.06	13.51	1.04	8.77
0.85	911	124	18179	7058	0.05	0.98	0.03	9.00	0.62	7.70

Table 5. Same as Table 4, but for FIP-AK Land domain.

Thresh	Forecast/Observation Pairs				Statistic					
	YY	YN	NY	NN	PODy	PODn	TSS	Area(%)	Vol(%)	Vol Eff
0.005	2838	695	490	596	0.85	0.46	0.31	61.53	18.62	4.58
0.02	2775	651	553	640	0.83	0.50	0.33	61.42	17.62	4.73
0.05	2664	600	664	691	0.80	0.54	0.34	60.84	16.14	4.96
0.1	2497	517	831	774	0.75	0.60	0.35	59.25	14.06	5.34
0.15	2320	448	1008	843	0.70	0.65	0.35	57.00	12.22	5.71
0.25	1944	330	1384	961	0.58	0.74	0.33	51.16	9.11	6.41
0.35	1463	233	1865	1058	0.44	0.82	0.26	42.95	6.15	7.15
0.45	1060	158	2268	1133	0.32	0.88	0.20	34.84	4.26	7.48
0.55	707	105	2621	1186	0.21	0.92	0.13	26.90	2.82	7.54
0.65	487	60	2841	1231	0.15	0.95	0.10	20.70	1.98	7.41
0.75	341	37	2987	1254	0.10	0.97	0.07	14.12	1.22	8.40
0.85	188	12	3140	1279	0.06	0.99	0.05	8.68	0.68	8.31

Table 6. Same as Table 4, but for Zones domain.

Thresh	Forecast/Observation Pairs				Statistic					
	YY	YN	NY	NN	PODy	PODn	TSS	Area(%)	Vol(%)	Vol Eff
0.005	1194	270	167	225	0.88	0.45	0.33	57.88	16.74	5.24
0.02	1173	249	188	246	0.86	0.50	0.36	57.77	15.78	5.46
0.05	1094	224	267	271	0.80	0.55	0.35	57.16	14.33	5.61
0.1	1003	191	358	304	0.74	0.61	0.35	55.38	12.29	6.00
0.15	907	169	454	326	0.67	0.66	0.33	52.93	10.50	6.35
0.25	682	124	679	371	0.50	0.75	0.25	47.39	7.67	6.54
0.35	468	87	893	408	0.34	0.82	0.17	39.52	5.13	6.71
0.45	322	55	1039	440	0.24	0.89	0.13	31.90	3.53	6.70
0.55	220	40	1141	455	0.16	0.92	0.08	24.66	2.37	6.81
0.65	138	16	1223	479	0.10	0.97	0.07	18.77	1.67	6.09
0.75	105	7	1256	488	0.08	0.99	0.06	12.98	1.07	7.20
0.85	62	2	1299	493	0.05	1.00	0.04	8.90	0.68	6.70

Table 7. **S**ame as Table 4, but for Anchorage domain.

Thresh	Forecast/Observation Pair				Statistic					
	YY	YN	NY	NN	PODy	PODn	TSS	Area(%)	Vol(%)	Vol Eff
0.005	776	97	111	41	0.87	0.30	0.17	73.83	23.78	3.68
0.02	756	90	131	48	0.85	0.35	0.20	73.76	22.57	3.78
0.05	714	80	173	58	0.81	0.42	0.23	73.25	20.81	3.87
0.1	658	65	229	73	0.74	0.53	0.27	72.75	18.54	4.00
0.15	598	55	289	83	0.67	0.60	0.28	71.75	16.45	4.10
0.25	453	44	434	94	0.51	0.68	0.19	65.55	12.40	4.12
0.35	289	32	598	106	0.33	0.77	0.09	56.59	7.97	4.09
0.45	184	18	703	120	0.21	0.87	0.08	45.70	5.31	3.91
0.55	121	10	766	128	0.14	0.93	0.06	34.71	3.45	3.95
0.65	73	4	814	134	0.08	0.97	0.05	26.90	2.41	3.41
0.75	51	1	836	137	0.06	0.99	0.05	21.58	1.76	3.27
0.85	32	0	855	138	0.04	1.00	0.04	18.02	1.33	2.72

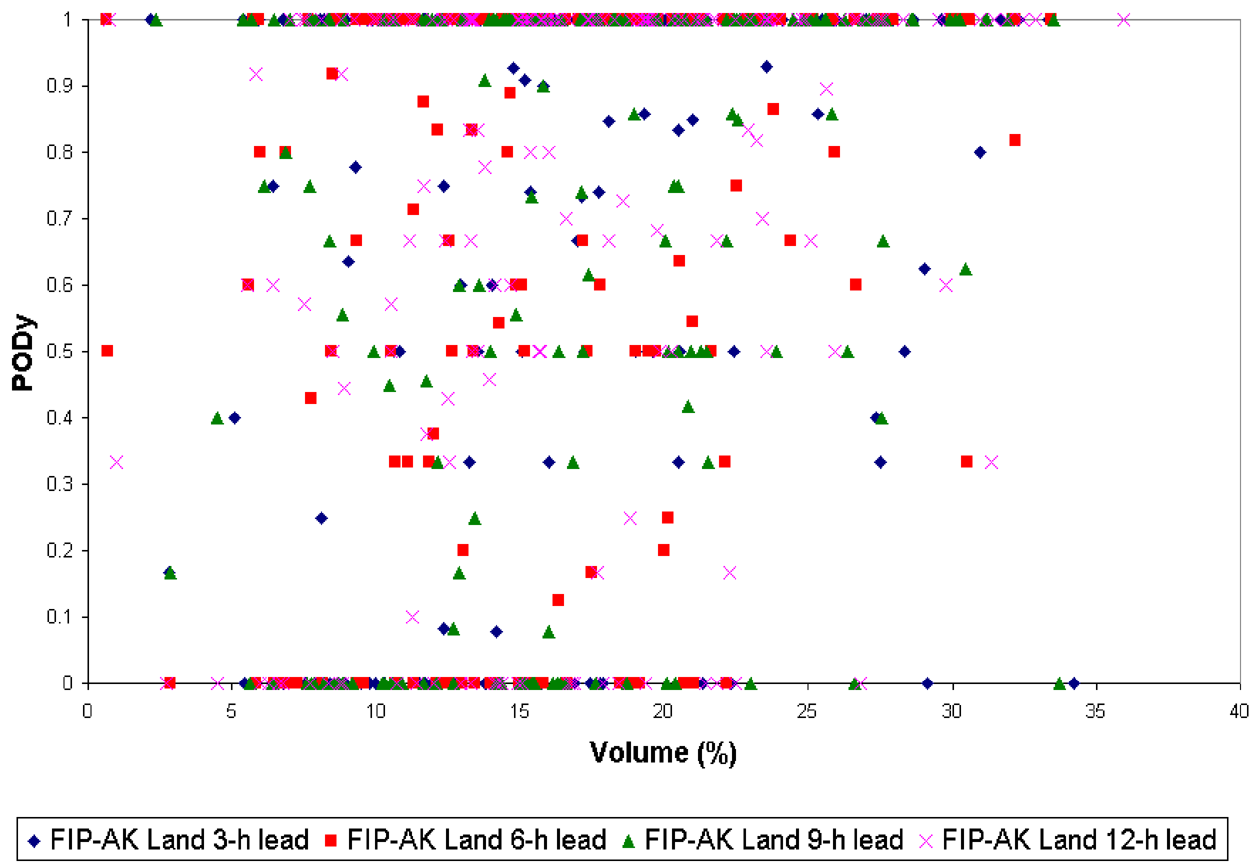

Figure 5. Scatterplot of daily PODy vs. Volume (%) for FIP-AK threshold 0.1. Points represent the daily PODy and Volume (%) value for each lead time for each day during the evaluation period when forecasts and observations were available.

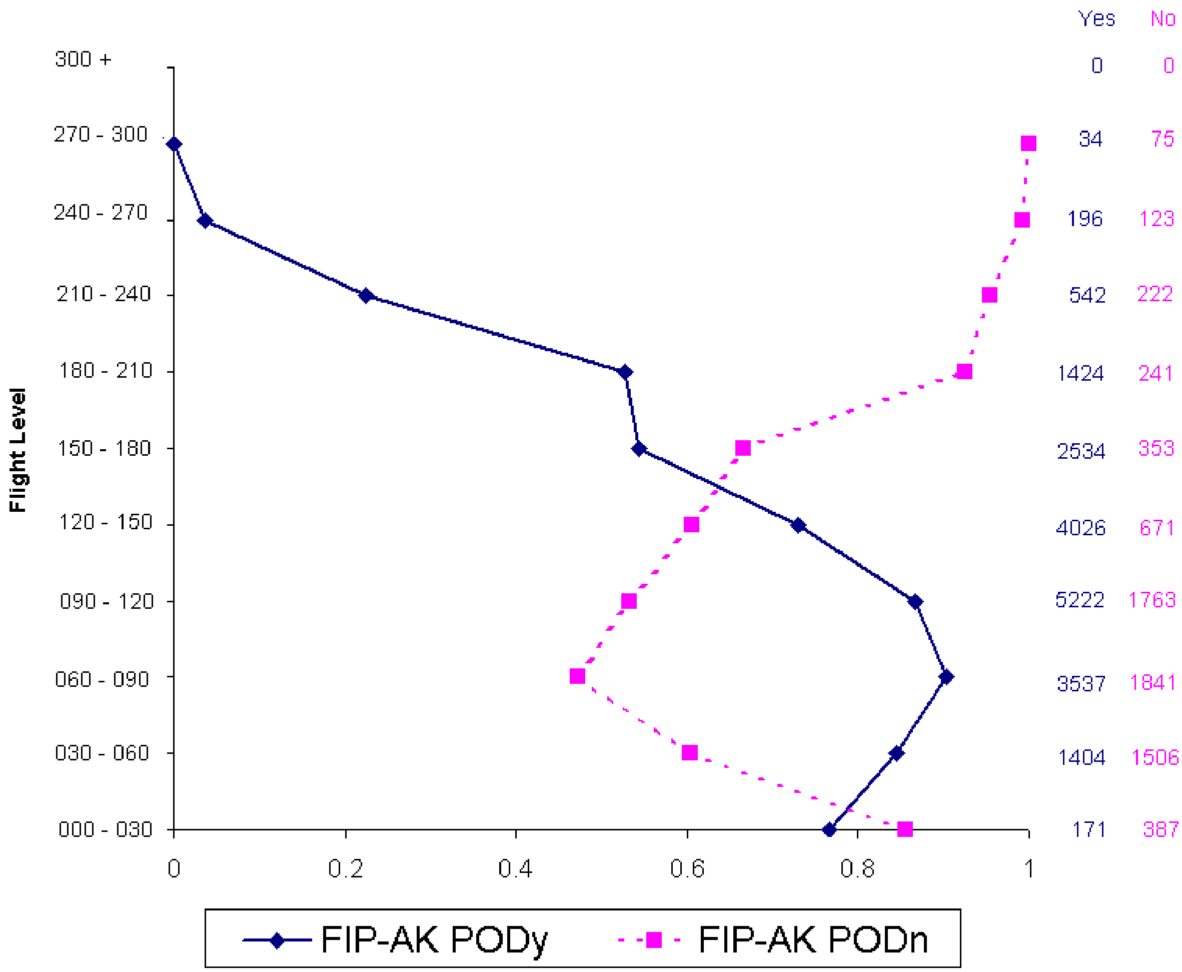

Figure 6. Height series of PODy (diamond) and PODn (square) as a function of altitude for FIP-AK over the FIP-AK domain using a 0.10 threshold. The figure includes all observations from 1 October 2003−31 March 2004. The two columns on the right show the number of “Yes” and “No” icing observations from each 3,000-ft layer.

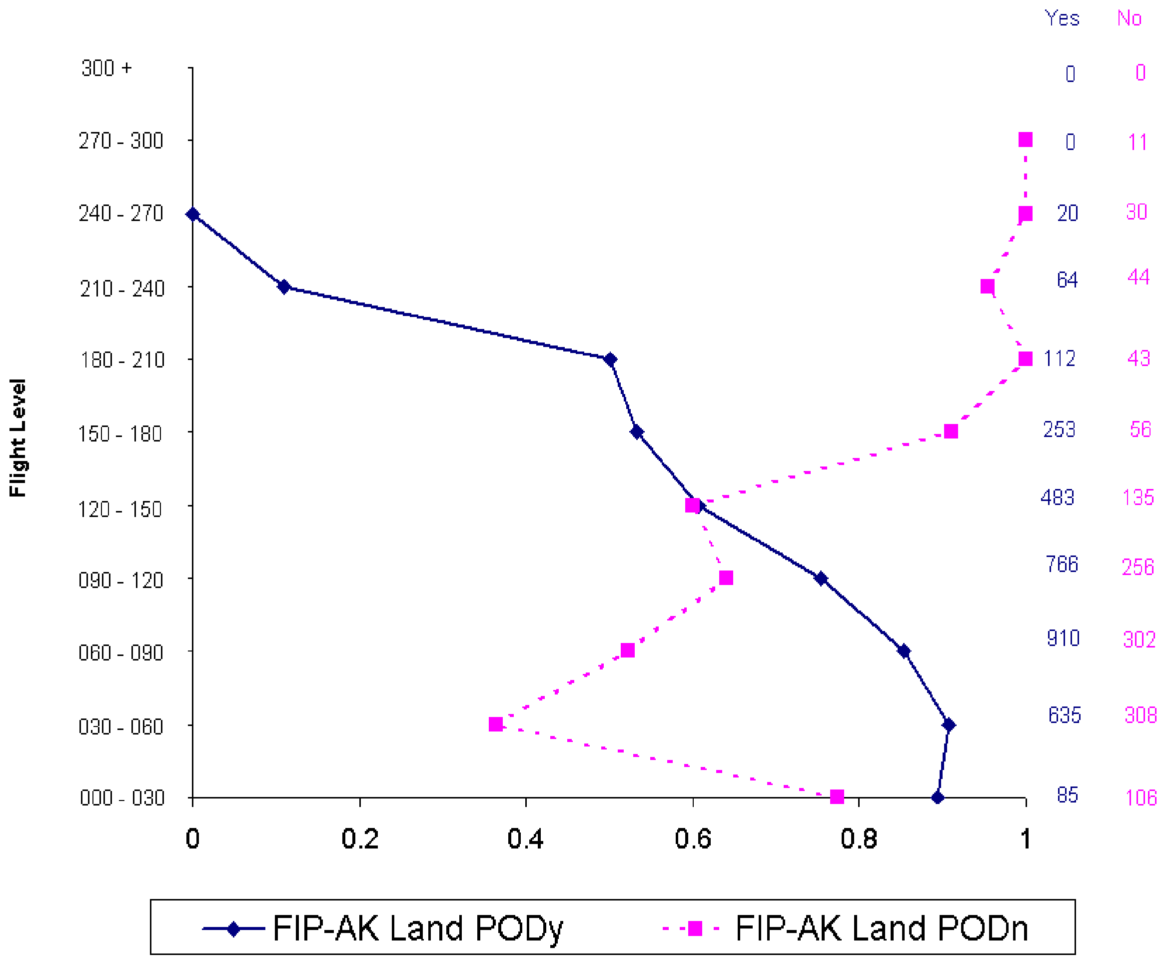

Figure 7. Same as Fig. 6, but for the FIP-AK Land domain.

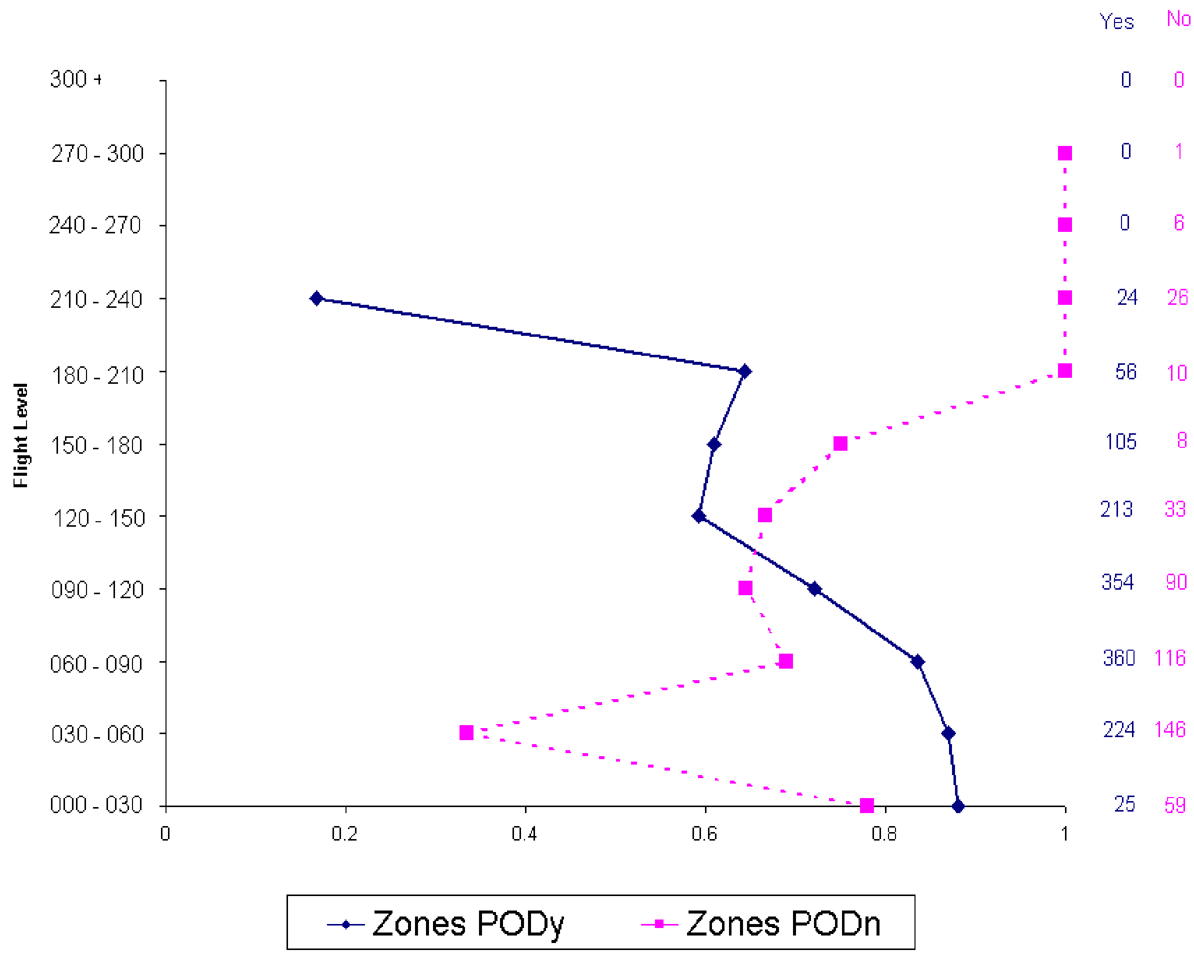

Figure 8. Same as Fig. 6, but for the Zones domain.

Figure 9 shows that the trend in the TSS values between the four domains is quite consistent, however, variability in the TSS score is somewhat greater for the domains where the number of observations is small (i.e., Anchorage and Zones). Using the FIP-AK domain as a baseline, the TSS score decreases with height from a value of 0.6 at the surface to nearly 0.2 at 18,000 ft. The increase in TSS at the 18,000−21,000 ft layer is the result of the algorithm appropriately forecasting No icing conditions.

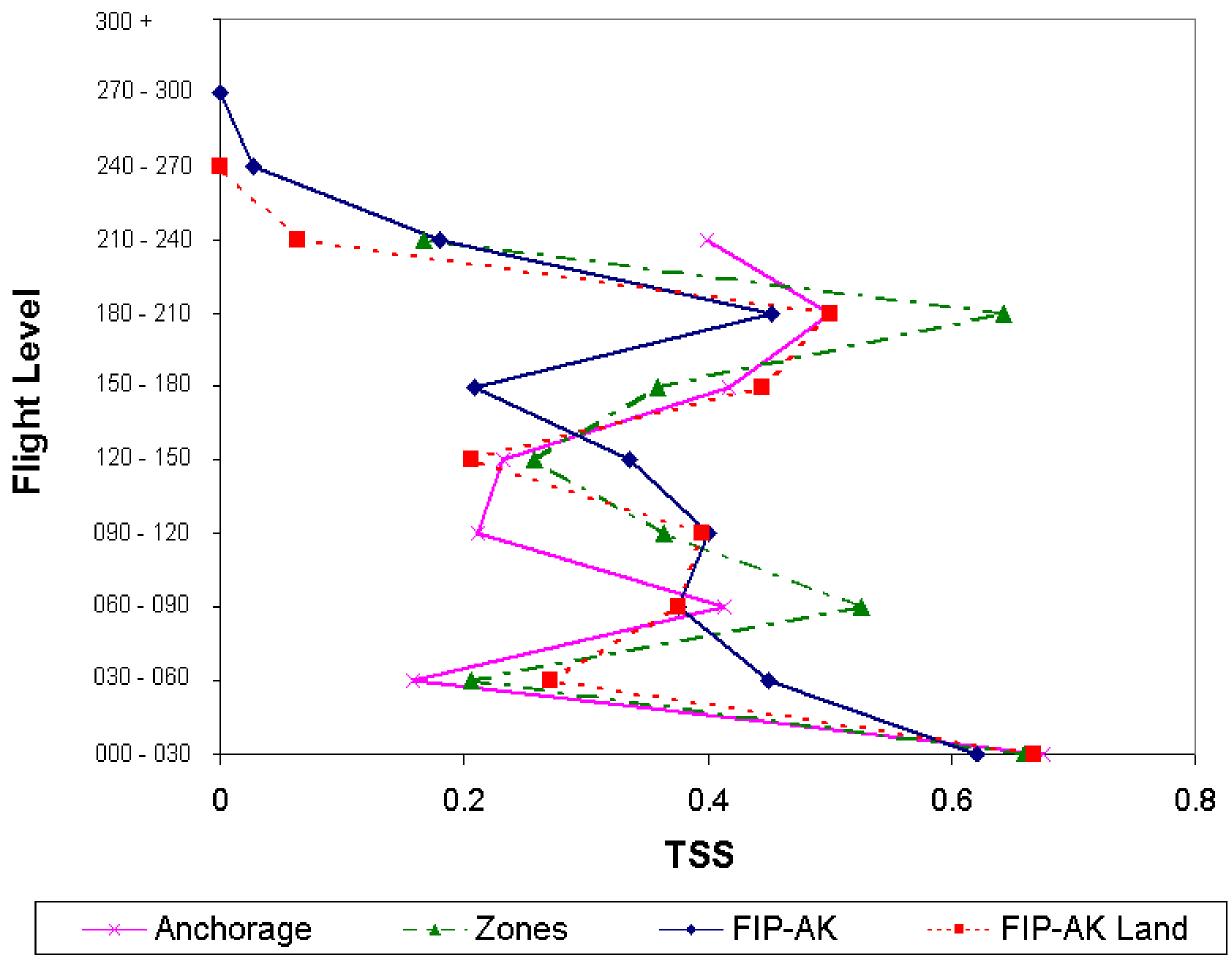

Figure 9. Same as Fig. 6, except for True Skill Statistic (TSS) as a function of altitude for FIP-AK (diamond), FIP-AK Land (square), Zones (triangle), and Anchorage ('x') domains using a 0.10 threshold.

5.3. Comparisons by forecast lead time

Overall verification results by forecast lead time for FIP-AK are shown in the curves of Figs. 10 and 11 and the height series plots of Figs. 12 and 13. As in Fig. 2, these results are based on PIREPs reporting at least light-to-moderate icing severity where all forecast issue times are combined. Only the results for the FIP-AK domain are shown. Each curve represents a specific forecast lead time for every 3 hours out to 12 hours.

The results in Figs. 10 and 11 indicate very little difference in forecast skill at the various lead times. In particular, the curves for PODy vs. 1-PODn and PODy vs. Volume for different lead times lie nearly on top of each other. A slight degradation in skill is noted in the ROC plot for the 9-hour forecasts, but this result is not statistically significant. Similar to the results for all lead times combined, the forecasts for specific forecast lead times show skill at correctly classifying Yes and No PIREPs. These figures also indicate that the relationship between PODy and Volume is very consistent for all lead times.

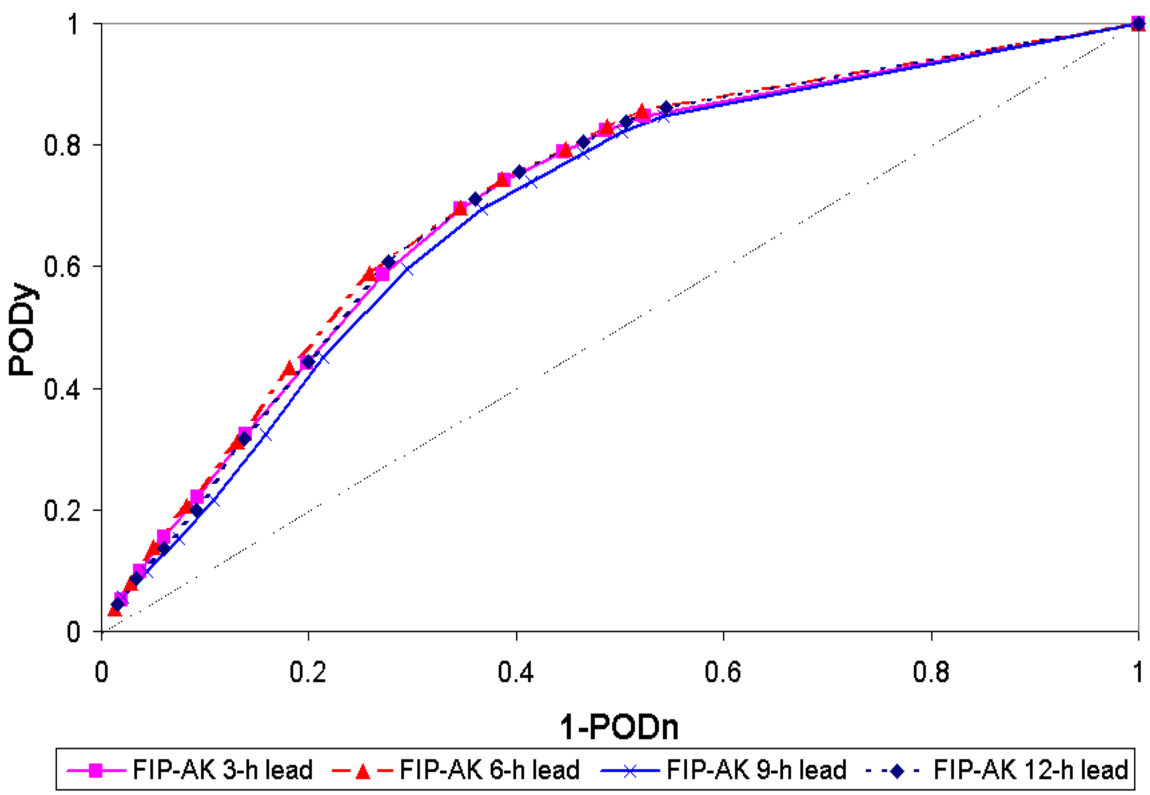

Figure 10. ROC curves (PODy vs. 1-PODn) for FIP-AK from 1 October 2003−31 March 2004 for all forecast issue times combined for the FIP-AK domain, based on light-to-moderate and greater PIREPs and separated by forecast lead time: 3 hours (square), 6 hours (triangle), 9 hours ('x'), and 12 hours (diamond). 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.

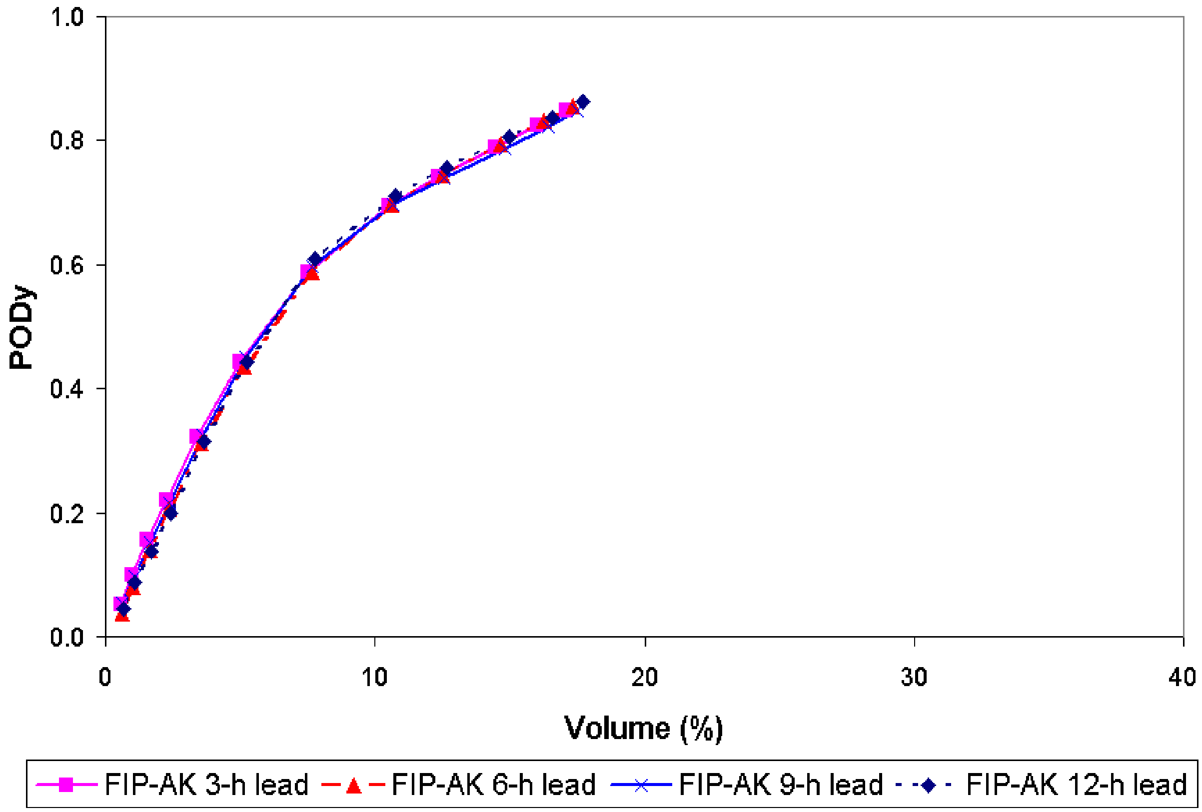

Figure 11. Same as Fig. 10, except for PODy vs Volume (%).

Results shown in Figs. 12 and 13 indicate slight variations in skill with height among the various forecast lead times. For instance, the largest variations in PODy among the forecast lead times occurs from the surface to 12,000 ft and above 24,000 ft. For values of PODn, the largest variations among lead times occurrs in the mid-layers from 6,000−18,000 ft.

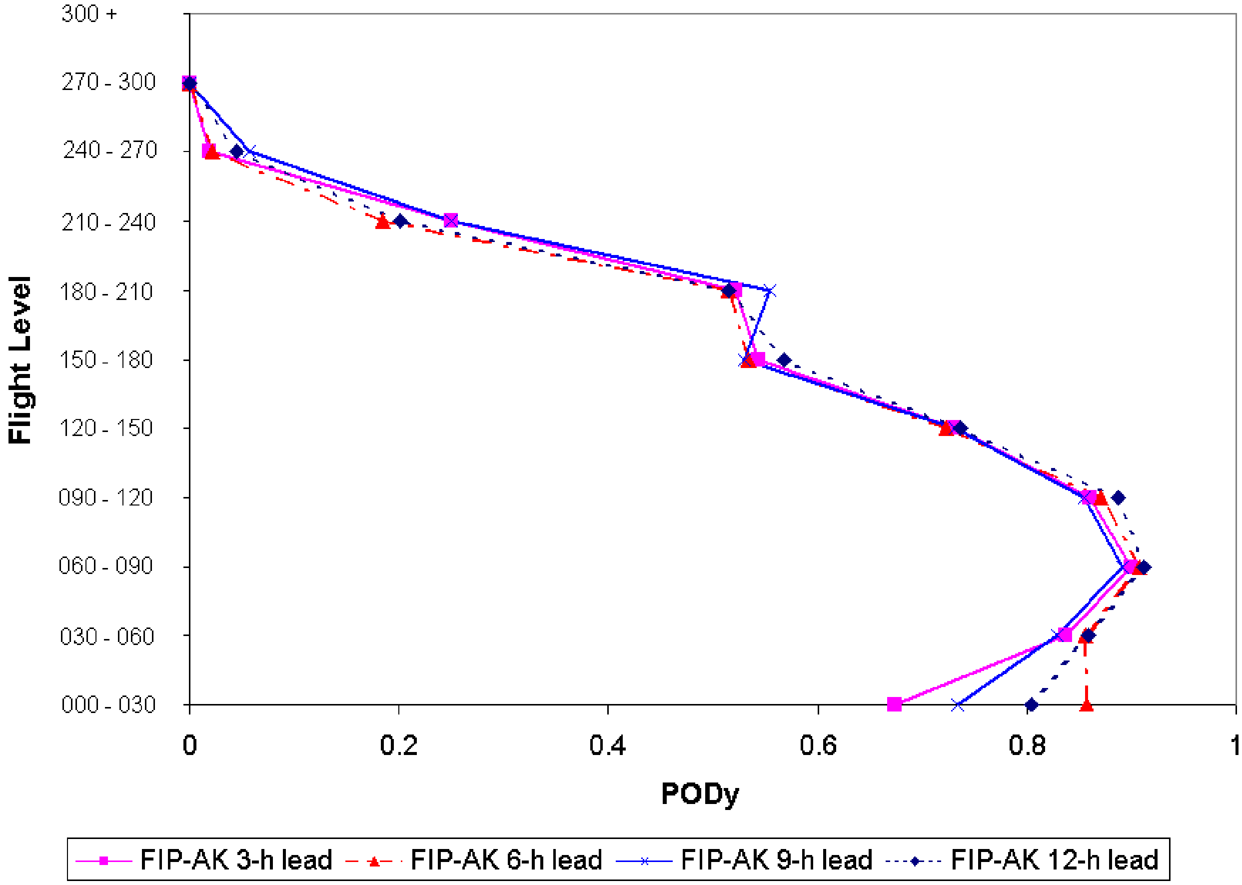

Figure 12. Height series of PODy as a function of altitude over the FIP-AK domain using a 0.10 threshold for 3-hour (square), 6-hour (triangle), 9-our ('x'), and 12-hour (diamond) forecast leads. The figure includes all observations from 1 October 2003−31 March 2004.

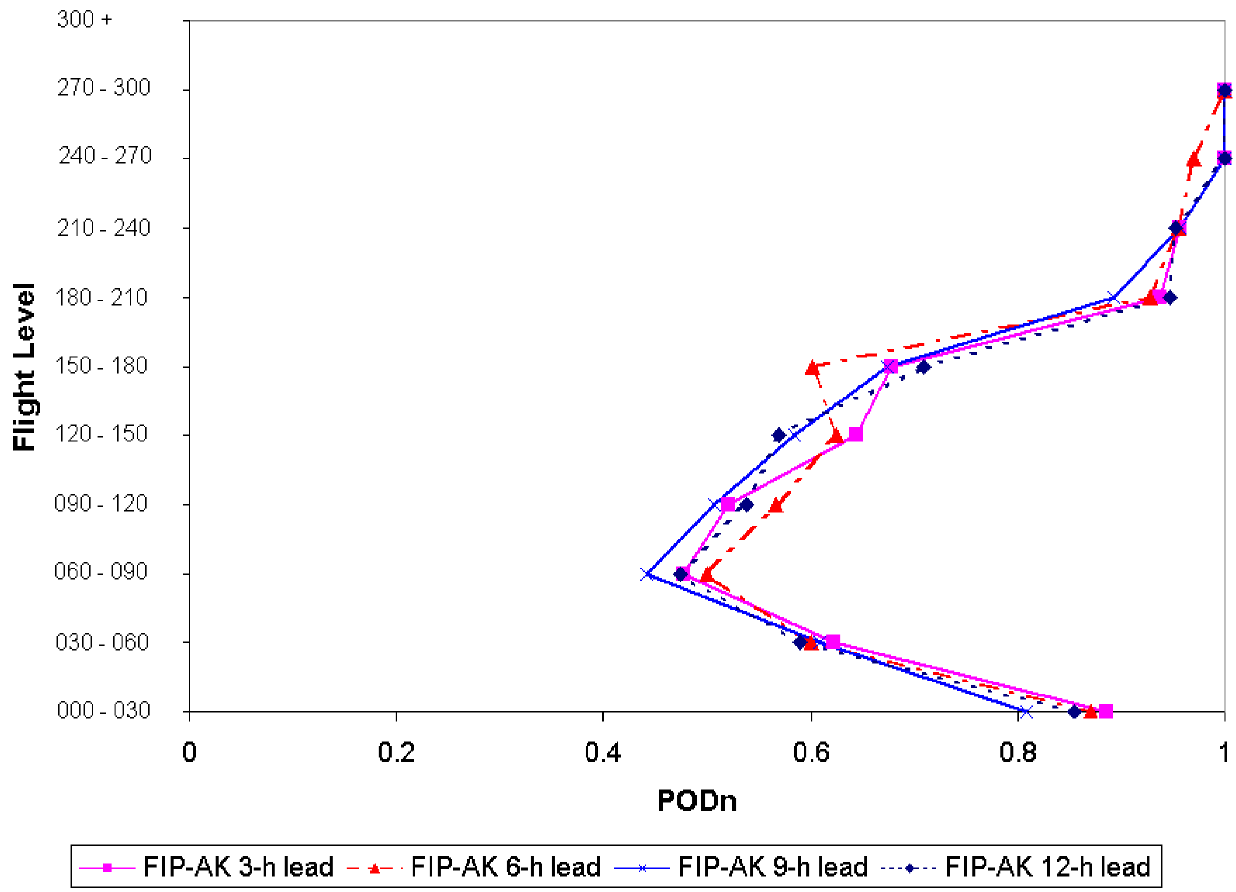

Figure 13. Same as Fig. 12, except for PODn.

5.4. Evaluation of FIP-AK SLD forecasts

PIREPs over Alaska, Washington, and Oregon for 1 October 2003−31 March 2004 that mentioned freezing drizzle (FZDZ) or freezing rain (FZRA) were selected and matched to FIP-AK SLD values. Although Alaska is the location of primary interest, few PIREPs from the Alaska region are available. Oregon and Washington are in the domain of the FIP-AK product, and have quite a few PIREPs. Thus, PIREPs from those states are used to supplement the ones available in Alaska.

The period from 1 October 2003−March 2004 had 584 PIREPs that mentioned FZDZ or FZRA. Of those, 64 reported No icing. The remaining 520 PIREPs are used to assess the quality of the FIP-AK SLD forecasts. Additionally, the 7,458 reports of No icing from the same area and time period are matched to the FIP-AK SLD values. Table 1 shows the percent of SLD and No PIREPs matched to FIP-AK SLD forecasts for no SLD (FIP-AK SLD = 0), for any SLD (FIP-AK SLD > 0), and for unknown SLD (FIP-AK SLD missing). The FIP-AK SLD value matched to the No (SLD) PIREPs was missing in 1,820 (254) cases, or in about 24% (49%) of the cases.

Table 8. Percent of SLD and No PIREPs matched to FSL-AK SLD values.

	SLD PIREPs (520)	**No PIREPs** (7,458)
FIP-AK SLD = 0	29%	65%
FIP-AK SLD > 0	22%	11%
FIP-AK SLD missing value	49%	24%

The percentages in the table are equivalent to PODy and PODn statistics for a FIP-AK SLD threshold value greater than zero. Thus, the PODy for this threshold is 22% while the PODn is 65%.

Table 9 shows the percentages when the missing FIP-AK SLD values are excluded from the calculations. The resulting PODy for a FIP-AK SLD value greater than zero is 43% while the PODn is 86%.

Table 9. Statistics for the non-missing FIP-AK SLD values.

	SLD PIREPs (266)	**No PIREPs** (5,638)
FIP SLD = 0	57%	86%
FIP SLD > 0	43%	14%

Because the primary focus of the FIP-AK SLD product is the Alaska airspace, the PIREPs from above 50 degrees latitude were analyzed separately. These PIREPs account for about half of the SLD PIREPs (276) from the FIP-AK domain, but only about 18% (1338) of the No-icing PIREPs. Few of the No PIREPs had a valid FIP-AK SLD value (only 5%) so no conclusions can be drawn about the ability of FIP-AK SLD to correctly classify PIREPs of No icing in Alaska. However, the SLD PIREPs had a corresponding FIP-AK SLD value above zero in (35%) 23% of the (non-missing) cases.

Table 10. Results for FIP-AK SLD using PIREPs above 50 degrees latitude.

	Alaska SLD PIREPs (276)	**Alaska SLD PIREPs matched to non-missing FIP-AK SLD values** (276)	**Alaska No PIREPs** (1,338)
FIP-AK SLD = 0	41%	65%	3%
FIP-AK SLD > 0	23%	35%	2%
FIP-AK SLD missing value	36%	NA	95%

6. Conclusions

This report summarizes the evaluation of icing potential forecasts provided by the FIP-AK algorithm from 1 October 2003−31 March 2004. The evaluation technique was built from previous work used to verify the CIP and FIP over the CONUS (Brown et al. 2001, 2002). The results presented here suggest that FIP-AK may be a useful tool for providing icing forecasts over Alaska. In particular:

- FIP-AK diagnoses are skillful, as measured by their ability to discriminate between Yes and No PIREPs of icing in a distinct volume.
- FIP-AK verified most favorably over the larger domains (FIP-AK, FIP-AK Land, and Zones). At least part of the reason for the difference is the greater number of

observations over the larger domains, especially with the inclusion of observations over the Pacific Northwest region of the CONUS.

- Over rather small domains with limited numbers of observations, such as the Anchorage domain, the FIP-AK showed skill at forecasting icing potential.
- FIP-AK verified most favorably between the surface and 21,000 ft, and performed especially well at forecasting the No icing events in the 18,000−24,000 ft layer. Above 24,000 ft, very few observations were available. However, FIP-AK forecast skill for the No-icing events decreased in the 3,000−6,000 ft layer over for the Alaska domains (FIP-AK Land and the Zones).
- In locations with sufficient information to calculate FIP-AK SLD, this forecast correctly identifies 43% (86%) of the FZDZ/FZRA (No) PIREPs. Furthermore, in Alaska, it correctly classified 35% of the SLD PIREPs in areas with a non-missing FIP-AK SLD value.

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

Acknowledgments

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. We express thanks to the AWRP Leadership Team for their support of independent verification activities. Finally, we would like to thank Nita Fullerton for her helpful review of this work.

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., 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.

Kelsch, M., J.L. Mahoney, T. Fowler, B.G. Brown, J. Henderson, C. Fischer, 2003: Current Icing Potential for Alaska (CIP-AK): Quality assessment report. Submitted to the Aviation Weather Technology Transfer Technical Review Panel, 3 April 2003. Available from J. Mahoney (Jennifer.Mahoney@noaa.gov) 35pp.

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.

McDonough, F. B.C. Bernstein, C.A. Wolff, and M.K. Politovich, 2004: The Alaska Forecast Icing Potential technical description. Submitted to the Aviation Weather Technology Transfer Technical Review Panel, 31 July 2004. Available from F. McDonough (Mcdono@ucar.edu).

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

Quality Assessment Product Development Team, 2003: Forecast Icing Potential (FIP): Quality assessment report. Submitted to the Aviation Weather Technology Transfer Technical Review Panel, 3 April 2003. Available from B.G. Brown (bgb@ucar.edu), 35 pp.

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; NWS 1991), 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 forecast, 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 FIP-AK forecast is compared to the quality of the operational forecasts (i.e., AIRMETs). FIP-AK and AIRMETs provide different types of information, for different time periods, and have different objectives. FIP-AK forecasts are generally understood to be valid at a particular time. The AIRMETs, on the other hand, are valid over a 6-hour 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 FIP-AK information, it is difficult to clearly compare their performance. However, in order to understand the quality of FIP-AK, it is necessary that FIP-AK forecasts 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 FIP-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 are the operational forecasts of icing conditions. These forecasts are produced by AAWU forecasters every 6 hours and are valid for up to 6 hours (NWS 1991). AIRMETs may be amended, as needed, between the standard issue times. The AIRMETs 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, only the AIRMETs corresponding to the FIP-AK valid times were evaluated.

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 forecast/observation pairs along with the area and volume scores for the AIRMETs for the period from 1 October 2003−31 March 2004. Due to the small sample size, the forecast/observation pair information for the AIRMETs is more meaningful than the verification statistics since the number of observations and forecasts were limited. However, for the interested reader, the statistics can be obtained from Figs. A1−A3. During the evaluation period, only 193 AIRMETs were verified. Figures A1–A3 show verification plots for the 3- and 6-hour FIP-AK forecasts with the corresponding AIRMETs (as indicated by the asterisks in each figure) that are valid at the same time. These figures show the ROC curve (PODy vs. 1-PODn; Fig. A1), the PODy vs. Area curve (Fig. A2), and the PODy vs. Volume curve (Fig. A3) for AIRMETs and FIP-AK forecasts. It was noted that the number of FIP-AK forecasts during the study was greater than the number of AIRMET forecasts of icing during the same period. In addition, a verification window of 4 hours was used to evaluate the FIP-AK, while no verification window was applied to the verification of the AIRMETs. The AIRMETs were only verified for the Zones domain since some specific geographic information was applied to the AIRMET decoder allowing for a somewhat more precise AIRMET forecast.

Based on the results presented in Table A1, which are based on the number of PIREP observations, 31% of the AIRMET forecasts correctly classified both the Yes and No icing events. On the other hand, 69% of the time, AIRMETs were either issued when No icing was observed or, icing was observed and an AIRMET was not issued.

Table A1. Verification forecast/observation pairs and Area and Volunes for the AIRMETs in the Zones domain.

	YY	YN	NY	NN	Area (%)	Volume (%)
AIRMETs	27	12	481	199	6.6	1.42

When FIP-AK results for the 3- and 6-hour forecasts for the Zones domain are compared to the AIRMETs valid at the same time, it is apparent that the statistics for the AIRMETs are similar in skill and area and volume coverage to the FIP-AK at the highest thresholds, with FIP-AK performance slightly better than the AIRMET performance.

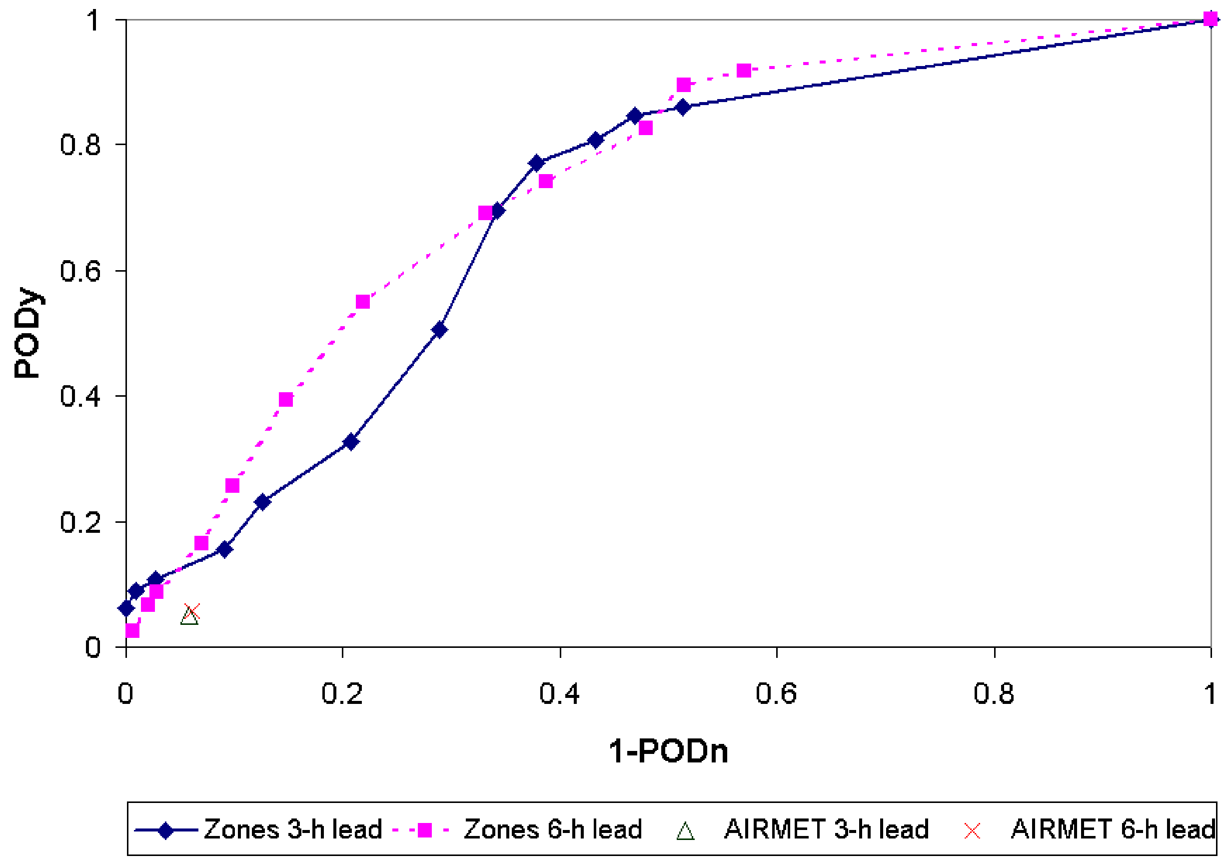

Figure A1. ROC curves (PODy vs. 1-PODn) for FIP-AK during from 1 October 2003−31 March 2004, for all issue times combined for the 3- and 6-hour forecasts, based on light and greater PIREPs for the Zones domain. The 3-hour (triangle) and 6-hour ("x") AIRMETs are plotted. Each point on the curve represents a different threshold used to define Yes/No icing for FIP-AK. 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.

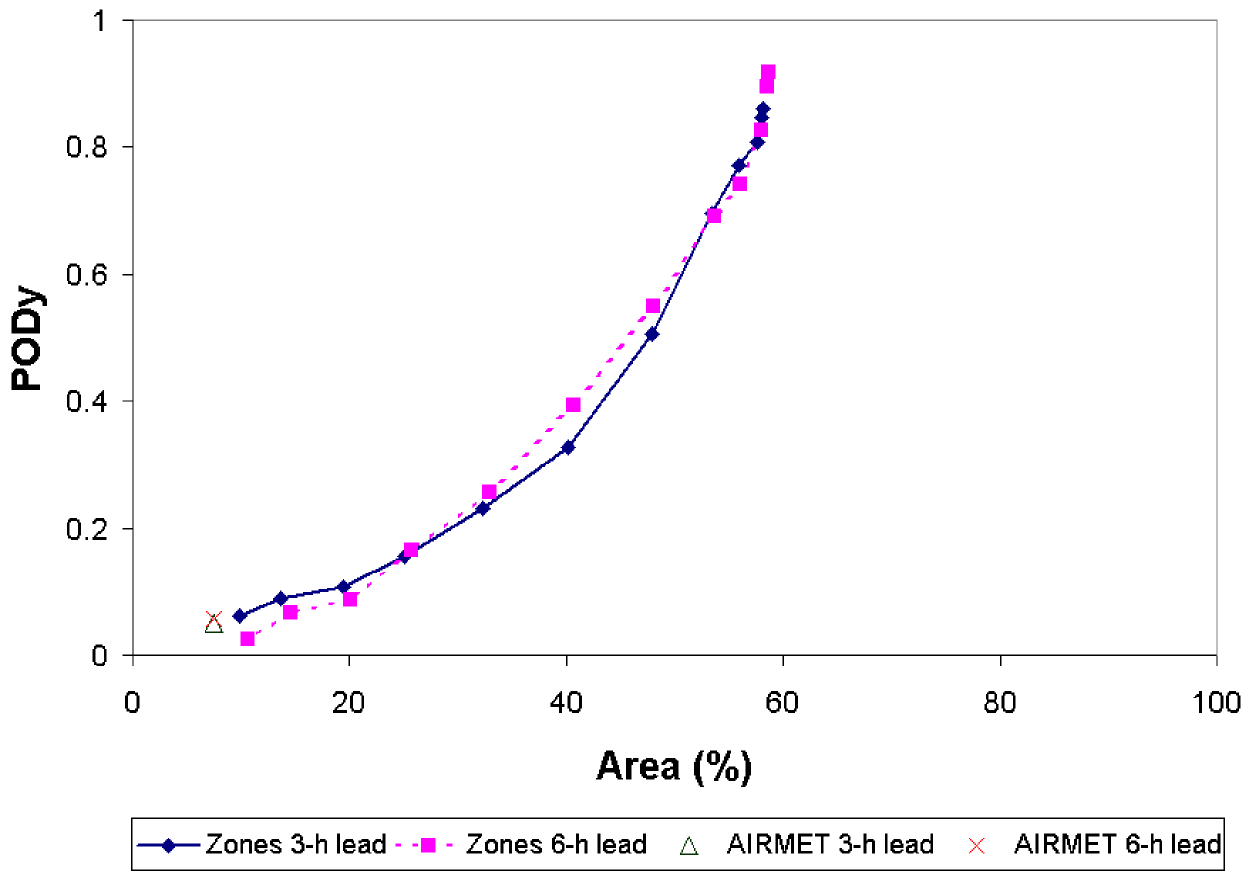

Figure A2. Same as Fig. A1, except for PODy vs. Area (%).

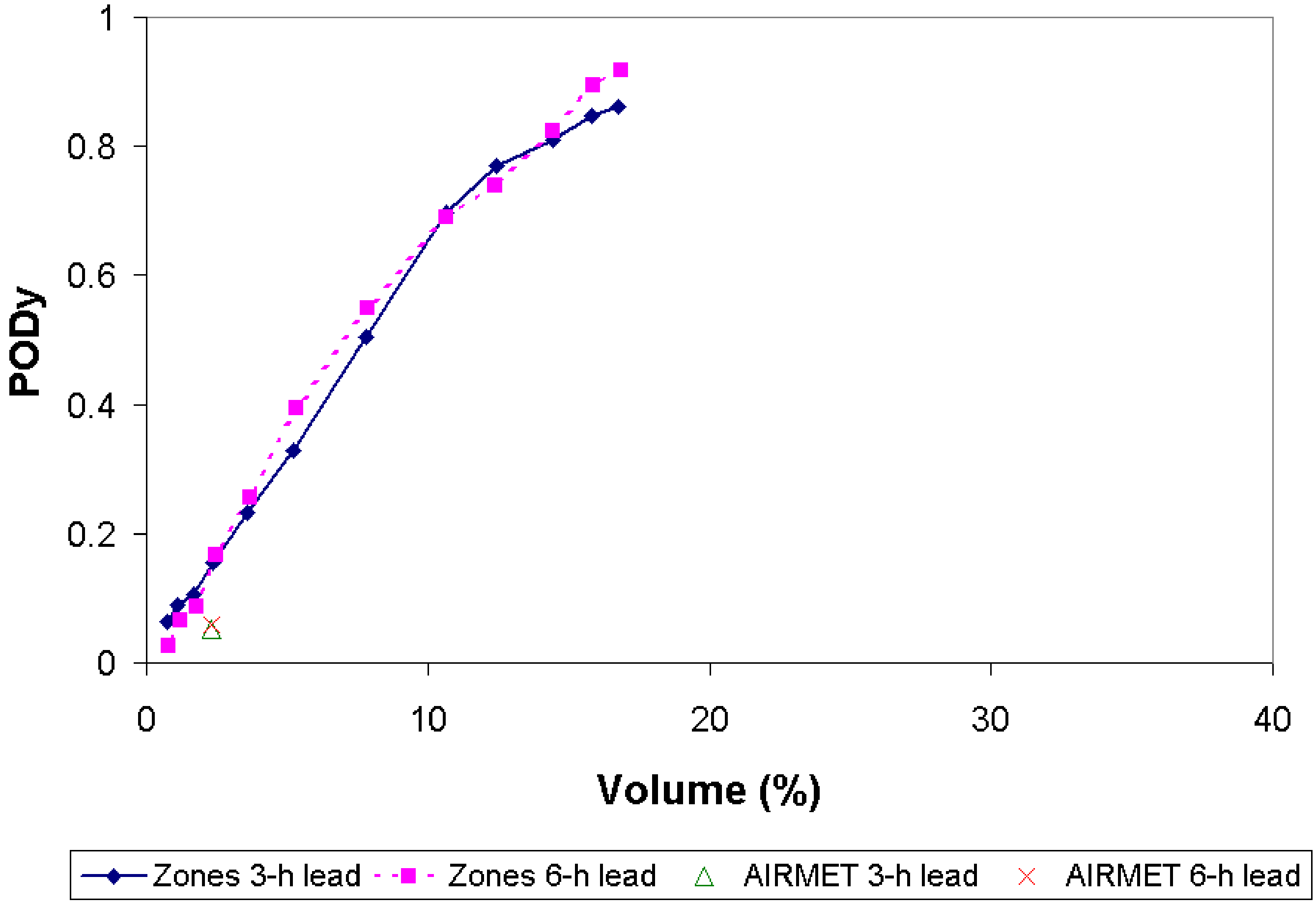

Figure A3. Same as plot Fig. A1, except for PODy vs. Volume (%).