Quality Assessment Report:

Forecast Icing Product (FIP)

Prepared By:
Quality Assessment Research Team
NOAA/ESRL/GSD

Sean Madine[3], Steven A. Lack[2], Stephen A. Early[2], Michael Chapman[2], Judy K. Henderson[1], Joan E. Hart[2], and Jennifer L. Mahoney[1]

March 17, 2008

Affiliations:
1 – National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD)
2- Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD
3 – Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

Corresponding Author:
J.L. Mahoney (NOAA/ESRL/GSD, 325 Broadway, Boulder, CO 80303; Jennifer.Mahoney@noaa.gov)

Table of Contents

List of Tables and Figures

EXECUTIVE SUMMARY

This report assesses the performance of the Forecast Icing Potential (FIP) product, which is under consideration for transition to operational status within the Federal Aviation Administration's Aviation Weather Technology Transfer (AWTT) process. Attributes of the forecast that were evaluated include icing probability, icing severity, and supercooled large drops (SLD). The FAA has designated that FIP, when certified for operational use, will be used as a supplemental product, which requires it to be used for flight planning purposes only in conjunction with the operational icing Airmen's Meteorological Information (AIRMET) issuances.

The primary objective of the report was to understand the value of FIP as a supplement to the icing AIRMET and secondarily as an independent forecast. First, agreement between the two forecasts was measured. Then, the skill of the supplemental product was examined in two ways: constraining the grid to within the boundaries of AIRMET polygons and constraining the grid to outside of the boundaries of the polygons. The performance of FIP was also assessed during the summer season, a time when icing AIRMET issuances substantially decrease. Finally, FIP was considered as an independent product and a reasonable attempt was made to compare its skill with that of the icing AIRMET.

Results from the study indicated:

- Qualitatively, FIP appears to effectively identify the structure of icing within the broad area outlined by the AIRMET polygons. Within an AIRMET polygon, FIP agrees with the operational forecast of icing in only about one-fifth of the forecast volume. However, FIP captures over 42% of the no-icing PIREPS while retaining a PODy of 0.66.
- FIP appears to identify some areas of icing that either weren't captured by the AIRMET or didn't meet minimum criteria for issuance. Outside of an AIRMET polygon, FIP agrees with the operational forecast of no icing in over 98% of the volume. FIP captures over 41% of the yes-icing PIREPS while retaining a PODn of 0.77.
- FIP provides significant support to the icing AIRMET forecast in the summer season by capturing 60% of the icing reports and 70% on the no-icing reports.
- In terminal areas, where rare events of SLD present a significant hazard, the FIP SLD forecast identifies over 50% of the observed freezing rain and freezing drizzle events.
- When FIP is considered as an independent product, its performance is similar to that of the icing AIRMET with respect to measures of PODy and PODn. However, FIP demonstrates an apparent improvement in the volume efficiency of capturing reports of icing.
- Overall, this study found FIP severity to perform best with no probability mask. The experimental Aviation Digital Display Service (ADDS) allows the icing

severity field to be “masked” by the icing probability field. Values include no mask, 0.25 probability, or 0.5 probability.

1. Introduction

This report summarizes a quality assessment performed on integrated icing forecasts (probability, severity, and supercooled large drops (SLD)) produced by the Forecast Icing Potential (FIP) product, which is under consideration for transition to operational status within the Federal Aviation Administration's Aviation Weather Technology Transfer (AWTT) process. Currently, FIP is designated as a supplemental product, which the FAA formally describes as:

> "An aviation weather product that may be used for enhanced situational awareness. If utilized, a supplementary weather product must only be used in conjunction with one or more primary weather product. In addition, the FAA may further restrict the use of supplementary aviation weather products through limitations described in the product label (FAA 2008)."

In the case of FIP, it is not to be considered independently in operations, but rather, as a supplement to issuances of icing Airmen's Meteorological Information (AIRMETs). AIRMETs are operational icing forecasts issued by the National Weather Service, National Centers for Environmental Prediction Aviation Weather Center (NWS/NCEP/AWC). In this study, the performance of FIP is primarily assessed in the context of the AIRMET issuances. Although difficult because of the difference in the product definitions, the performance of FIP as an independent product and based on verification data from 2006 and 2007 was also evaluated relative to that of AIRMETs.

The report is organized into six sections. Section 2 outlines the study approach. Section 3 describes the different data types utilized in the evaluation, while the methods and techniques applied are detailed in Section 4. The results are presented in Section 5, and the conclusions are given in Section 6.

2. Approach

While borrowing some elements from the approach taken in previous studies of icing products (Brown et al. 2001, 2002; Chapman et al. 2007), the verification approach in this study represents a significant shift. To understand the value of FIP as a supplement to the icing AIRMETs, the product is evaluated in the context of the AIRMET forecast. First, agreement between the two forecasts is examined. Then, the value of the supplemental product is examined in two ways:

- Within an AIRMET polygon, does FIP effectively identify significant areas of no icing (complimentary disagreement) while strongly agreeing on areas of icing?

- Outside of an AIRMET polygon, does FIP effectively identify significant areas of icing (complimentary disagreement) while strongly agreeing on areas of no icing?

FIP is also evaluated as a supplement for the summer season, when AIRMET icing issuances decrease dramatically in response to a change in the frequency and nature of the icing threat to aviation. The performance and characteristics of the FIP product itself are then evaluated by analyzing discrimination, reliability, and skill across forecast increment (lead-time).

Finally, FIP is considered as an independent product and a reasonable attempt is made to compare its skill with that of the icing AIRMETs. The difficulties and consequences associated with this comparison are detailed in Section 4. The performance of FIP SLD, an attribute not present in the AIRMETs, is also included in the assessment.

The study utilizes several stratifications, including geographic domain, altitude, and FIP probability mask for the severity field. For FIP probability and severity and AIRMETs, Pilot Weather Reports (PIREPs) of icing were used as the verification data set. For FIP SLD, PIREPs were used in conjunction with a surface observation-based scheme. As done in previous studies (e.g. Chapman et al. 2007), many parts of the report focus on the forecast and observation of Moderate-Or-Greater (MOG) icing. In these cases, less-than-MOG (LTM) icing is treated as a negative report. This approach was chosen to ensure consistency with past reports and to treat all forecasts as equitably as possible during comparisons. The basis of the evaluation and comparisons is a suite of common (though restricted) comparative statistics, which is described to greater detail in Section 4.

3. Data

In this quality assessment, the complete FIP product (encompassing Probability, Severity, and Supercooled Large Drop (SLD) algorithms) is examined over the continental United States (CONUS) in conjunction with Airmen's Meteorological Information (AIRMET) issuances. Pilot weather reports (PIREPs) are used as the primary verification source. Additionally, METAR's are utilized to verify the SLD attribute of FIP.

3.1 Forecast Icing Potential (FIP)

The Forecast Icing Potential product (akin to the Current Icing Product, CIP) utilizes a physics-based conditional fuzzy logic technique (McDonough et al. 2003; Wolff et al. 2003) based on output from the Rapid Update Cycle (RUC) numerical weather prediction model (Benjamin et al. 1998, 2004). More specifically, the FIP algorithms process output model variables (e.g. temperature, relative humidity, vertical velocity) from the 13km RUC model (with fields down-sampled to 20km) to identify cloud layers, precipitation areas, and precipitation classification before creating interest maps using fuzzy logic membership functions (Wolff 2003). The FIP version used in this evaluation had a "freeze date" of July 13, 2007. FIP forecasts are provided on a three dimensional domain that covers the CONUS in the horizontal and 1kft vertical levels. Operationally, a three-hour FIP forecast with hourly output is generated and issued every hour, along with 6-, 9-, and 12-h forecasts that are issued every three hours starting at 00 UTC (NWS 2007). In this study, FIP output from 351 days over 2006 and 2007 were used. A total of 487,269 individual forecasts were used in this study, including the 1-,2-,3-,6-,9-, and 12-h lead times for every 3-h forecast. For this evaluation, FIP probability and severity forecasts are combined. The FIP probability algorithm uses many of the same RUC model output fields as the FIP severity algorithm, but with different combinations (Wolff 2003). FIP severity is thresholded into one of five icing categories, as detailed in Table 3.1, and combined with probability using different probability masks (e.g. 50%, 25%, 0%). FIP SLD forecasts are made on a scale from zero to 1 where 0 is no SLD and anything greater than 0.01 is a positive SLD forecast. An example of an FIP forecast from the Aviation Digital Data Service (ADDS) website (http://adds.aviationweather.noaa.gov) is given in Figure 3.1.

Table 3.1. FIP Icing Category Classification.

FIP Icing Severity Category	FIP Severity Algorithm Value
None	FIPsev < 0.01
Trace	0.01 ≤ FIPsev ≤ 0.25
Light	0.25 < FIPsev ≤ 0.425
Moderate	0.425 < FIPsev ≤ 0.75
Heavy	FIPsev > 0.75

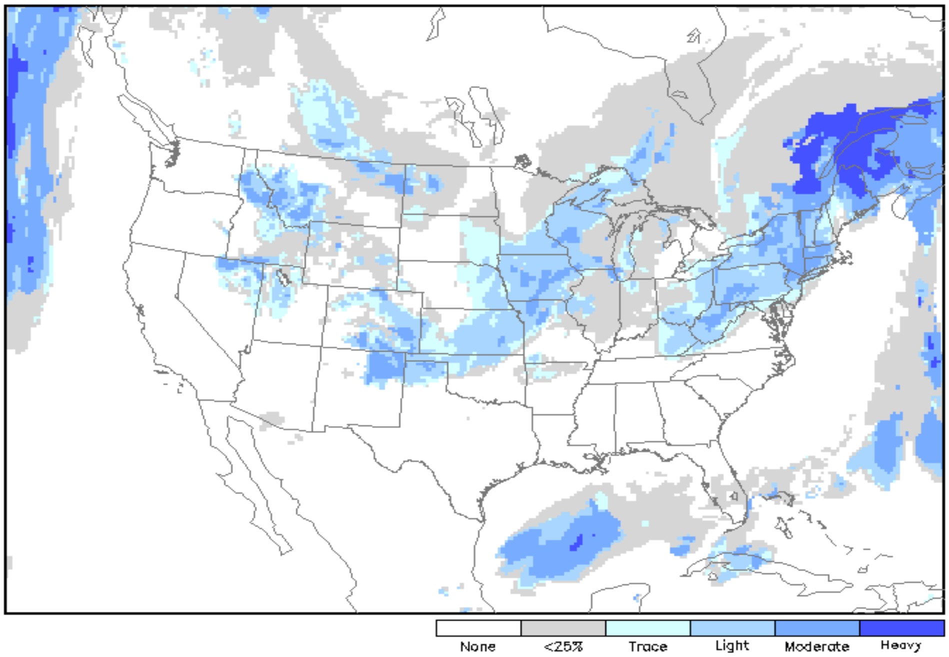


Figure 3.1. Example of a 6-hour FIP forecast. Severity categories are shaded, and the probability mask for the plot is 25%.

3.2 AIRMETs (Icing)

An Airmen's Meteorological Information (AIRMET) is a concise description of the occurrence of expected icing in time and space of specified en-route weather phenomena (NWS 2007). Thus, AIRMETs are operational forecasts of icing (and other significant weather phenomena), and are produced at the Aviation Weather Center (AWC) for CONUS. They are issued on a scheduled basis every six hours (at 0300, 0900, 1500, and 2100 UTC), but can be amended and updated if circumstances warrant (NWS, 2007). Icing AIRMETs (AIRMET-Zulu) are issued when moderate-or-greater (MOG) icing is occurring or expected to occur over an area of at least 3,000 square miles within the 6-h valid time of the AIRMET bulletin (NWS 2007). A sample icing AIRMET "forecast" is given in Figure 3.2.

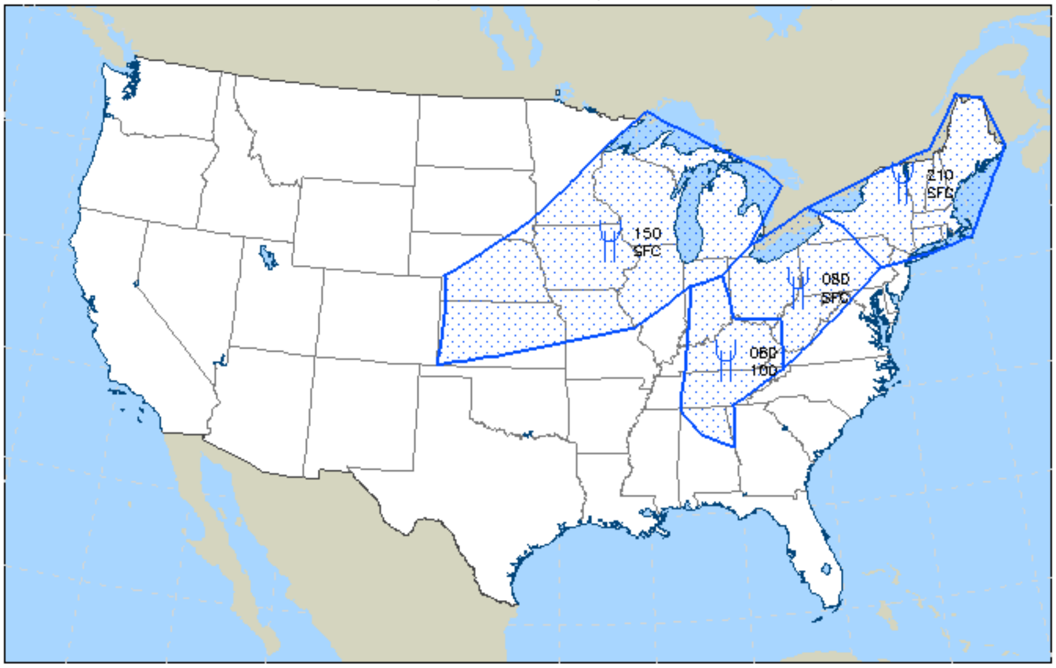


Figure 3.2. Sample graphical AIRMET depiction.

3.3 Pilot Weather Reports (PIREPs)

As the name suggests, Pilot Weather Reports are the method by which a pilot can relay a description of weather phenomena and related information. PIREPs provide a valuable observation source for weather-related flight hazards (icing, turbulence, sky condition, etc.). For in-flight icing, in the absence of special instrumentation, PIREPs are the only source of icing observations. As such, they are the primary verification observation source for this study.

3.3.1 PIREP Characteristics

It is important to point out that there are limitations involved when using PIREPs as a verification source. While undesirable in that respect, PIREPs do represent the most complete and tested verification data set available for this evaluation. PIREPs are subjective by nature and sporadic in space and time (Kane et al. 1998). Further, it has been shown that PIREPs are biased toward positive reports and not systematic (Brown et al. 1997). As a result, some standard verification methods are inappropriate when using PIREPs as the verification source (Brown et al. 1997). This is discussed in more detail in Section 4 or this report.

3.3.2 PIREP Data

PIREP icing observations include icing type and intensity, aircraft type, location, altitude (flight level), and time (NWS 2007). The intensity is categorized as none, trace, light, moderate, or severe, but ranges and variations in the intensities are allowed (NWS 2007). This means that there are 9 possible “native” PIREP intensity types. These intensity types were accumulated into 6 “intermediate” intensity bins. A description of the relationship between the original PIREP icing intensity types, the intermediate PIREP icing intensity types, and the FIP icing intensity types is presented in Figure 3.3.

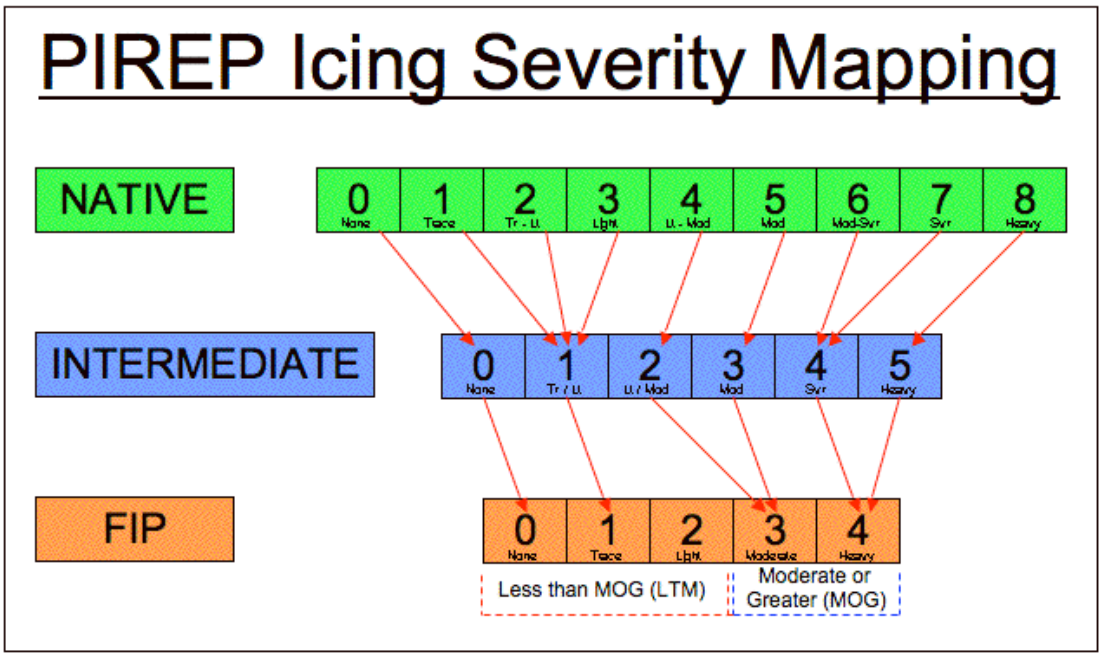


Figure 3.3. Mapping relationship between PIREP icing intensity scales

A total of 155,099 PIREPs were used in this study. The distribution of PIREPs by intensity category (on the GSD paradigm scale) and season is given in Figure 3.4.

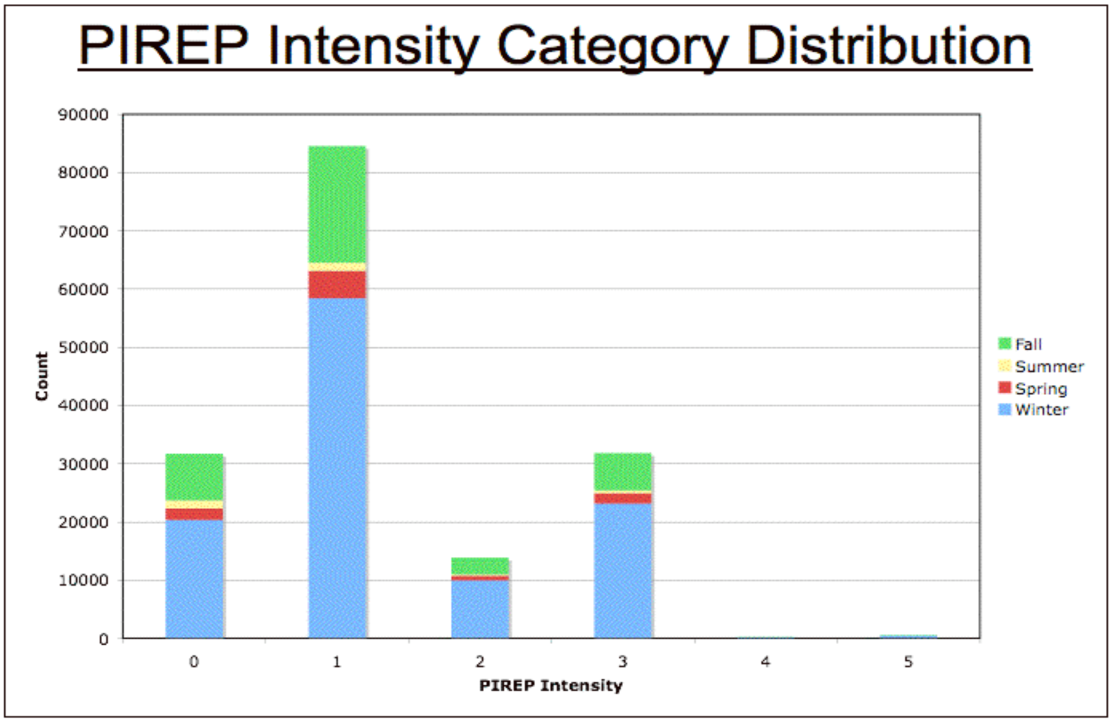


Figure 3.4. PIREP Distribution by Intensity Category and Season.

3.3.3 Observations of Super cooled Large Drops (SLD)

Supercooled Large Drops (SLD) are defined as super cooled water droplets larger than 50 micrometers in diameter, and include freezing drizzle and/or freezing rain aloft (NWS 2007). In-situ observation of SLD is problematical due to the danger associated with flying in such conditions and the fact that pilots actively try to avoid flying in areas with suspected SLD. In this study, there are two sources of observations of SLD. The first is PIREPs, where SLD is interpreted when "SLD" or freezing rain or freezing drizzle is reported in the remarks, or where icing intensity of 5 is reported with "clear" icing type. The second is through the use of aviation routine weather report (METAR) data. SLD events are inferred between the surface and ceiling (lowest cloud layer of at least "broken" coverage) when freezing rain or freezing drizzle is reported in the observation.

4. Methods and Techniques

This section defines terminology and describes the statistical and verification methods and techniques used in the report.

4.1 Definitions

MOG	= Moderate or Greater Icing
LTM	= Less than MOG Icing
FIP-CON	= FIP Forecast Constrained to be Inside of AIRMET Boundary
FIP-NA	= FIP Forecast Constrained to be Not Inside of AIRMET
FIP-NC	= FIP Forecast Not Constrained (Independent)
PODy	= Probability ("Proportion") of Detecting MOG Icing
PODn	= Probability ("Proportion") of Detecting Less than MOG Icing

4.2 Matching Methods

In order to enable comparison and evaluation, forecasts and observations were matched together spatially and temporally.

4.2.1 FIP (Probability and Severity) to PIREP (Severity)

Spatial Matching – As in previous evaluations, PIREPs are matched to FIP values at the nearest 12 grid points (the nearest 4 points at flight level of the PIREP, the nearest 4 points at the grid level above the PIREP, and the nearest 4 points at the grid level below the PIREP).

Temporal Matching – The temporal matching of PIREPs to FIP is a multi-step process. The process is conditioned on the PIREP. First, each PIREP is matched to every FIP forecast in a day (each separate issue time). Next, an initial mask is applied whereby the PIREP is only associated with a particular FIP forecast if it falls within a certain time window from the FIP forecast valid time. The time window is variable depending on the forecast lead time (+-.5/-6.5 hours for a 1-h forecast; +/- 0.5 hours for a 2-h forecast; +1.5/-0.5 hours for a 3-h forecast; +1.5/-1.5 hours for forecasts of 6-h and 9-h forecast; and +6.5/-1.5 hours for a 12-h forecast). Finally, a secondary filter is applied that only retains PIREP/FIP pairs if the PIREP time is within 2 hours (+/- 1 hour) of the FIP valid time. A graphic description of this process is given in Figure 4.1.

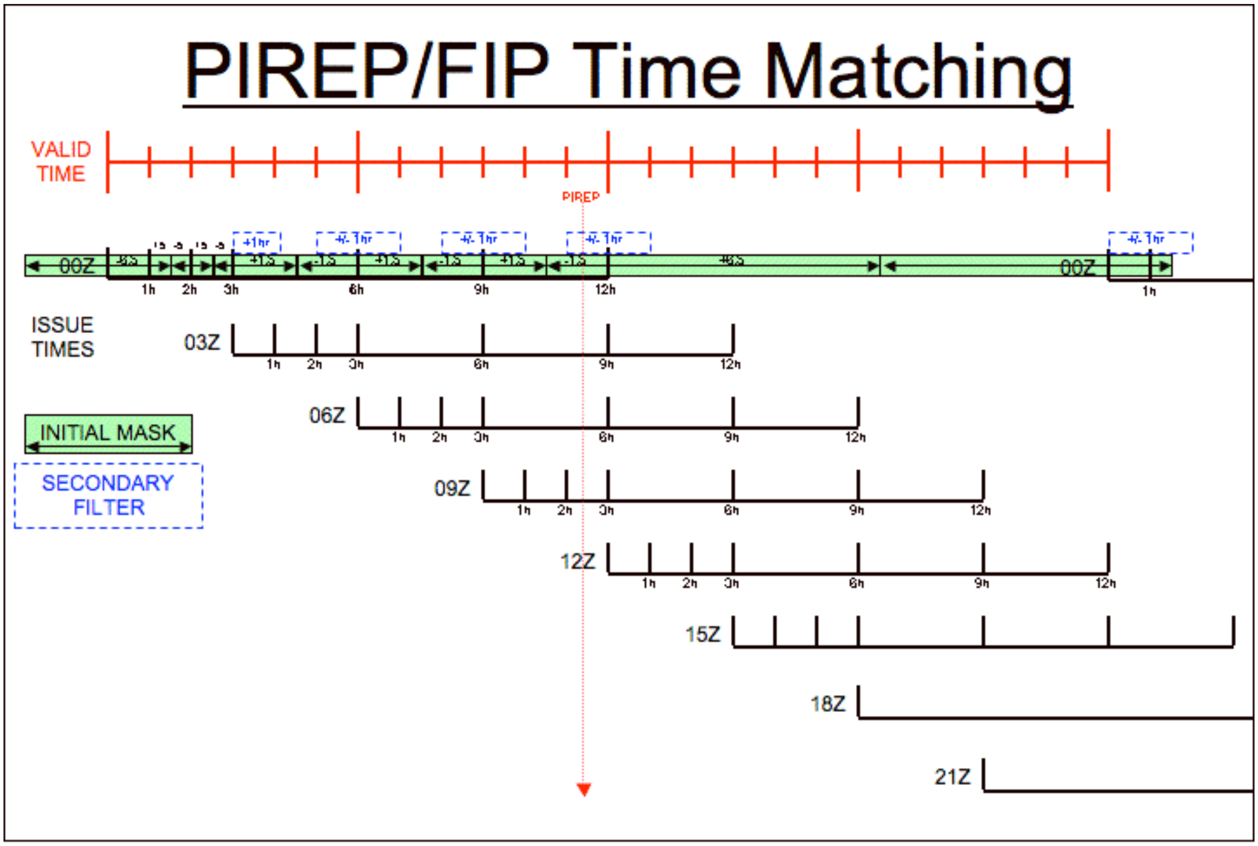


Figure 4.1. PIREP/FIP Time Matching Scheme.

4.2.2 <u>AIRMET to PIREP (Severity)</u>

Spatial Matching – PIREPs are matched to an AIRMET if they fall within the AIRMETs boundaries.

Temporal Matching – PIREPs are matched to an AIRMET if the AIRMET valid range encompasses the PIREP time +/- 30 minutes.

4.2.3 <u>FIP (Probability and Severity) to AIRMET</u>

Spatial Matching – A "relative bias" spatial matching structure was created that allows for comparison of the forecasts made by FIP and AIRMETs. Part of this structure involved the construction of a 2x2 contingency table to analyze the frequency of agreement and disagreement between FIP and AIRMET forecasts of MOG icing. A diagram detailing this approach is found in Figure 4.2.

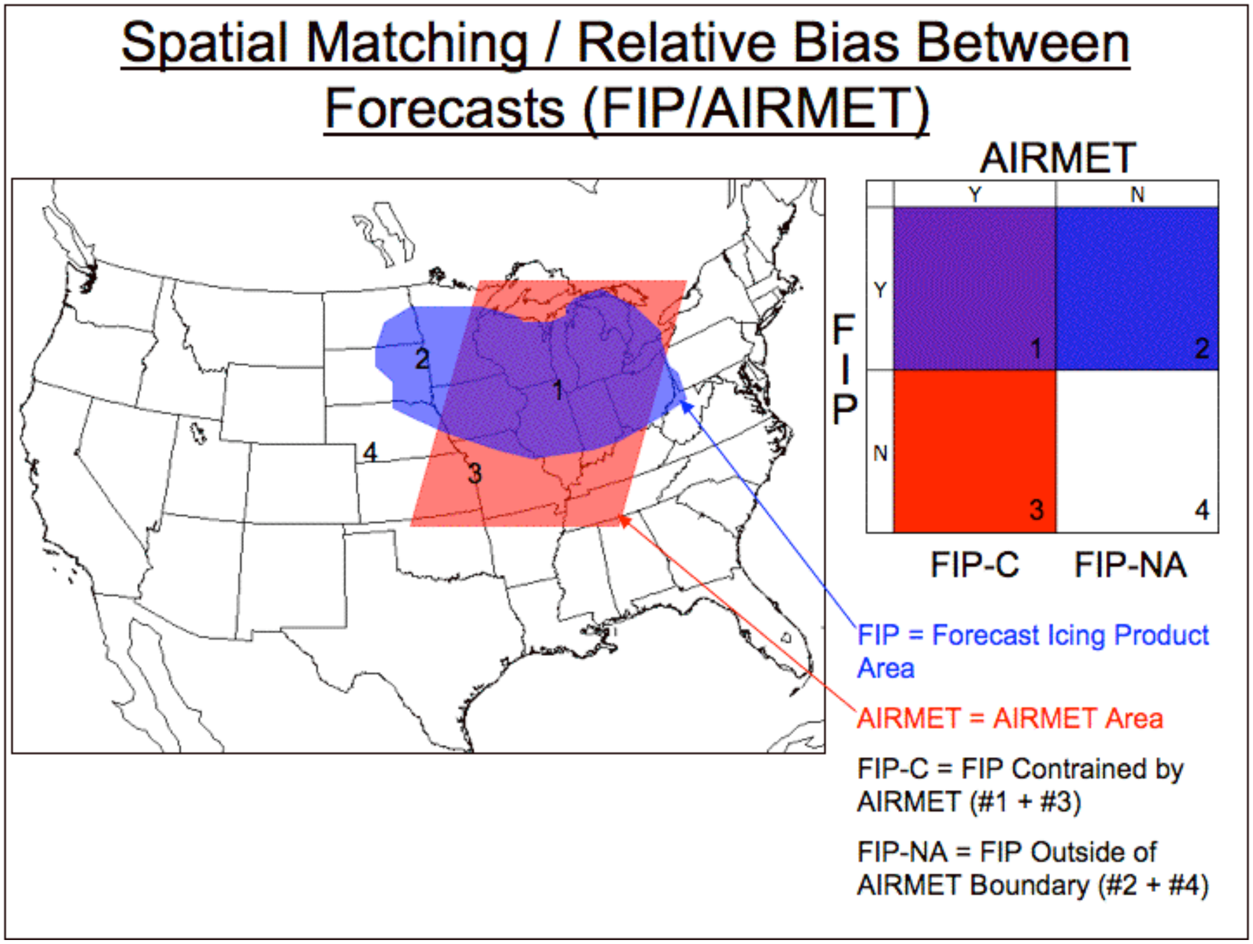


Figure 4.2. FIP/AIRMET Spatial Matching Depiction ("Relative Bias").

Temporal Matching – To facilitate a reasonable comparison between FIP and AIRMET fields, a FIP forecast aggregate was created by combining 1-h, 2-h, 3-h, and 6-h forecasts/verification into a time domain resembling that of an AIRMET (nominally, a "forecast" of conditions occurring or expected to occur within 6 hours (NWS 2007)). The FIP/PIREP matching was done as described above (Figure 4.1), though the results were found to be relatively insensitive to the choice of time window.

It is important to note that there are inherent inadequacies in comparing FIP forecasts and AIRMETs, and that an ideal comparison is not possible. Though an attempt is made for the benefit of the reader, it is not possible to remove entirely the "apples to oranges" nature of this comparison. For example, the AIRMET size criterion (3000 continuous square miles) is not something that can be replicated in a practical way by the FIP at this time. Also, the effect of the FIP aggregation/"smearing" needs to be accounted for, which is reflected in the volume efficiency statistics.

4.3 Verification and Statistical Methods

4.3.1 Joint Probability Distribution Composition and Structure

The verification methods used are consistent with the approach detailed by Brown et al. (1997) and similar to those used in previous studies. The standard 2x2 contingency table is used as a basis in this report. All of the forecasts and observations are treated as dichotomous events (yes or no). For FIP forecasts, this required the use of thresholds. A 2x2 contingency table is presented in Table 4.1. The columns in the table are a count of the distribution of the observations, while the rows represent the distribution of the forecasts. Combined, the contingency table characterizes the joint distribution of forecasts and observations.

Table 4.1. Standard 2x2 Contingency Table

2x2 Contingency Table

		Observations		
		Yes	No	Total
Forecasts	Yes	YY	YN	YY+YN
	No	NY	NN	NY+NN
	Total	YY+NY	YN+NN	YY+YN+NY+NN

4.3.2 Verification Limitations due to Assumption of PIREPs as "Truth"

Due to the non-systematic nature of the verification data set (PIREPs), the "yes" observations and "no" observations must be treated separately (Carriere et al. 1997). As a result, it becomes inappropriate to compute several common statistics that would otherwise be computed and analyzed (e.g. Critical Success Index, Bias, False Alarm Ratio). The rationale for this is well documented by Brown and Young (2000) and Carriere et al. (1997).

4.3.3 METAR-based SLD Interpretation Scheme

As introduced earlier, a scheme based on METAR data (originally developed for IFR/VFR determination (Loughe et al. 2007)) was applied to SLD event identification in this study. It is noteworthy that the scheme isolates SLD events and not individual SLD reports. The scheme starts by identifying episodes of freezing rain or freezing drizzle of at least 30 minutes and the transitions in and out of such events. Using this approach, an episode is ignored if less than 30 minutes in duration. A requirement is imposed that there be at least one METAR report 2 hours prior to the observed transition to frozen precipitation, at least one report two hours following the transition, and an average of one report every 90 minutes throughout the time period of the event. The same requirements are imposed on event-trailing edges, marking the transition back from the defined event. Additionally, a restriction is imposed such that longest data gap allowed between subsequent observations during an identified event is 2.25 hours.

4.3.4 Definition and Description of Primary Statistics Used

A description of the primary statistics used in this evaluation is given in Table 4.2.

Table 4.2. Summary Table of Verification Statistics.

Statistic	Formula	Description	Interpretation
PODy	YY/(YY+NY)	Probability of Detection of "Yes" Observations	Proportion of "Yes" Observations Correctly Forecast
PODn	NN/(YN+NN)	Probability of Detection of "No" Observations	Proportion of "No" Observations Correctly Forecast
Percent Volume	(VOLfcst / VOLtotal) * 100	Fraction / Percent of Total Volume Forecasted	Fraction / Percent of Total Volume Impacted by Forecast
Volume Efficiency	((PODy) * 100) / Percent Volume	PODy Per Unit Volume	Relative Skill / Efficiency of the Forecast

4.3.5 *Comparison Techniques*

Four basic plots were used throughout this study for the various results – strip charts, histograms, "skill plots", and height series plots. The strip chart is used to illustrate PODy, PODn, and volume efficiency along the abscissa for a specific lead time for 15 small regions. The histogram was used to depict the count and the distribution of FIP and PIREP attributes. A number of plots of skill versus lead-time or issue-time for the evaluation of forecast products were also used in this analysis. Finally, height series plots were used for the depiction of skill in the vertical across the different flight levels in 5kft intervals for the CONUS as a whole and for the sub-regions defined in Section 4.4.

Many skill plots presented utilize a 95% confidence interval (C.I.). The confidence interval (1) is based upon the central limit theory and Gaussian assumptions; however, for a large population (*N*) such as those being dealt with in this study, this estimate is an adequate coarse measure of the interval. In equation 1, the PODy skill may be interchanged with other skill scores used in this study, such as PODn and volume efficiency.

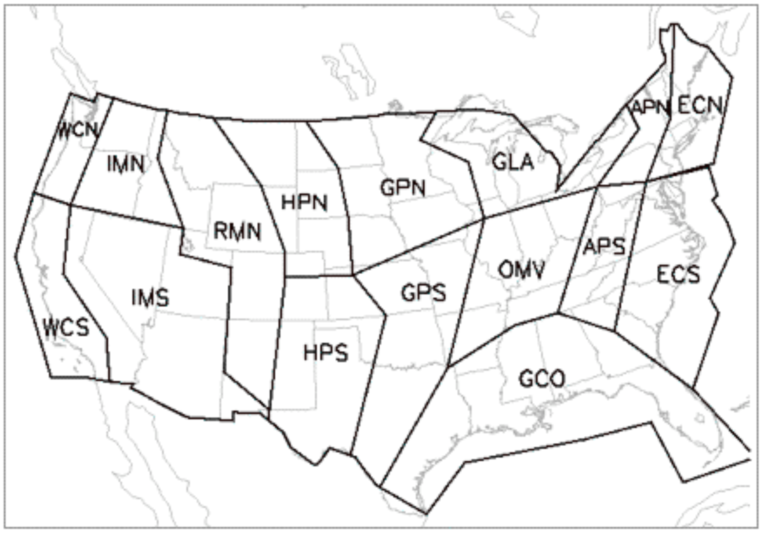
$$95\%\ \text{C.I.} = PODy \pm \sqrt{\frac{(PODy)(1-PODy)}{N}} \qquad (1)$$

4.4 Stratifications

The verification results presented are stratified in multiple ways. Combinations of stratifications are also used. The stratification categories are detailed below.

- FIP Mask → Masks of 0 (no mask), 0.25, and 0.5.
- Season → Winter, Spring, Summer, and Fall.
- Lead Time →The available forecast lead times are 1-h, 2-h, 3-h, 6-h, and 12h.
- Altitude → Altitude bins of 10,000 feet were used (0-10kft; 10-20kft; 20-30kft). All altitudes are Above Ground Level (AGL).
- Region → The CONUS domain was divided into air-traffic-based sub-regions. A map of these sub-regions is presented in Figure 4.3. Further, the sub-regions are ranked in order of air traffic density, as depicted in Figure 4.4.

CONUS w/Regions



ID	Region
WCN	West Coast North
WCS	West Coast South
IMN	Intermountain North
IMS	Intermountain South
RMN	Rocky Mountain
HPN	High Plains North
HPS	High Plains South
GPN	Great Plains North
GPS	Great Plains South
GLA	Great Lakes
OMV	Ohio and Mississippi Valley
GCO	Gulf Coast
APP	Appalacians
ECN	East Coast North
ECS	East Coast South

Figure 4.3. Definition of CONUS Sub-Regions.

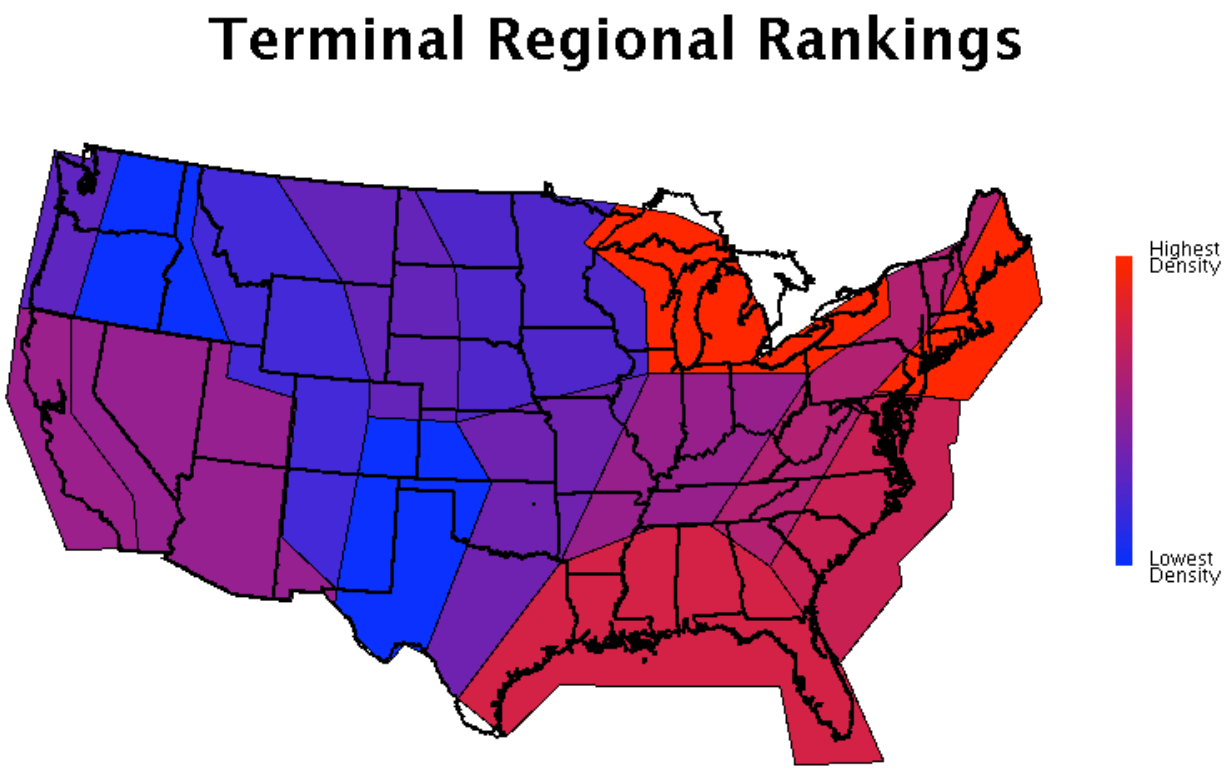


Figure 4.4. A map of the sub-regions colored by a measure of terminal air traffic density. The density, computed from empirical air traffic data, was based on the departure and arrival rates for all major airports resident in each region.

In order to identify the sub-regions most relevant to air traffic planners utilizing icing forecasts, a coarse measure of terminal air traffic density was implemented. The algorithm computed the departure and arrival rates for 35 major airports in the CONUS. Then, it aggregated the total rates for each of the sub-regions. The result, shown in Figure 4.4, shows where low-level air traffic is significant, and allows FIP skill to be examined with respect to operational use.

4.5 Time Period of Study

The time period for this study is broken down by season as follows:

Winter = January-March 2006 and January-March 2007

Spring = May 2006

Summer = July-August 2006

Fall = October 2006 and October-November 2007

5. Results

5.1 Assessment Overview

Framework

This study assesses the FIP product in two ways: primarily as a product used to supplement the icing AIRMETs, and secondarily as an independent forecast. FIP will therefore be examined when constrained to the interior of the AIRMET boundaries as well as when it is constrained to the outside of the AIRMET boundaries. In addition, FIP will be assessed as a supplementary tool in the summer season when the icing AIRMET volume is drastically reduced, likely due to additional coverage from Convective SIGMETs for icing events in the proximity of thunderstorms.

Special attention is paid to the winter season (January-March 2006 and 2007) due to the importance of icing forecasts during that time, and the number of observed events captured by PIREPs (Figure 3.4). The experimental Aviation Digital Data Service (ADDS) provides images of FIP severity forecasts with an option to apply one of three probability masks: none, 0.25, and 0.5. The discussion will focus on a comparison between using the 0.25 probability for masking the FIP severity field and using no probability masking. Application of the 0.5 probability mask excludes a dramatic amount of volume from the severity forecast; assessment of this mask was deemed not meaningful (Figure 5.13 right). Finally, the 3-h lead-time will serve as a proxy for the value of FIP to supplement the AIRMET as it is valid at the midpoint of the AIRMET valid time window. The skill over all issue times for the 3-h lead is relatively uniform.

Forecast Agreement

To examine the agreement between FIP severity and AIRMET icing forecasts, the joint probability distribution (JPD) between the two products was computed. The JPD reveals the relative bias of the forecasts, which is the fractional agreement both within AIRMET polygons and outside of AIRMET polygons. Table 5.1 shows the overall JPD for the number of FIP forecast grids compared to the AIRMET forecast for all altitudes during the winter season when using a probability mask of none and 0.25. FIP to AIRMET forecast agreement is considered when FIP issues a less-than-MOG (LTM) forecast grid outside an AIRMET and a MOG forecast grid inside an AIRMET. Strikingly, the FIP forecast grid inside of AIRMET polygons agrees with the AIRMET MOG icing forecast only about 22% of the time in the winter season when using no probability mask and only 15% when using the 0.25 probability mask. Outside of the

AIRMET polygons, where MOG icing should be considerably less, the distribution shows good agreement between the two forecasts.

Table 5.1. The overall joint probability distribution (JPD) for the number of FIP forecast grids compared to the AIRMET forecast using no probability mask and the 0.25 probability mask for all altitudes during the winter season. Agreement quadrants are highlighted.

No Probability Mask		
LTM	25325707	2443368
MOG	235970	683473
LTM	99.08%	78.14%
MOG	0.92%	21.86%
0.25 Probability Mask		
LTM	25412402	2667751
MOG	149275	459090
LTM	99.42%	85.32%
MOG	0.58%	14.68%
	Outside AIRMET	Inside AIRMET

A case study, depicted in Figure 5.1, qualitatively illustrates the findings from the JPD examination. The images contain FIP and AIRMET icing areas from 6 different flight levels. The significant areas of icing in the FIP grid primarily reside within the AIRMET polygons, but fill a much smaller volume. In effect, the grid describes the structure of the icing within the broad area of concern indicated by the AIRMET polygon. The FIP grid highlights significant regions of MOG icing within the AIRMET near the mid-levels, while correctly indicating negligible icing above 20kft above-ground-level (AGL). These qualitative observations demonstrate the nature of the disagreement of the forecasts within AIRMET polygons indicated by the JPD.

Disagreement between forecasts can indicate the potential for a supplementary relationship that adds significant value. When FIP is constrained within the AIRMET polygon boundaries, the grid would add value by identifying the structure of areas that have no icing (increasing the PODn statistic) while retaining the significant areas of icing broadly outlined by the polygon (maintaining a high PODy). Outside the AIRMET boundary, the FIP grid would add value by identifying MOG icing not captured by the AIRMETs (increasing PODy) while keeping PODn high. Sections 5.2 and 5.3 investigate these questions in detail.

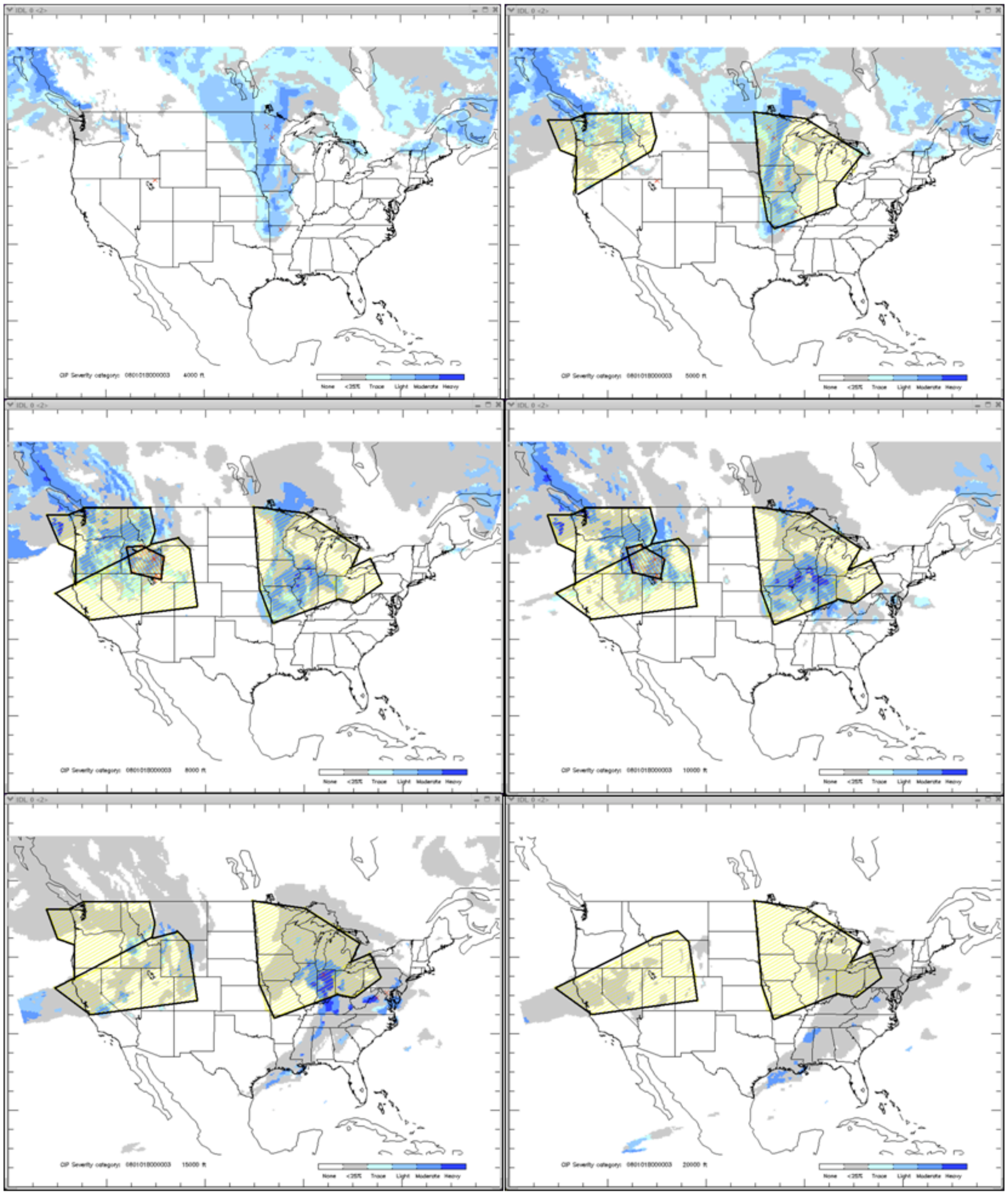


Figure 5.1. A case study of the FIP icing severity (blue shading) overlaid with AIRMET boundary (stippled yellow) and SIGMET boundary (stippled red) at different altitudes from 1 January 2008 (3-h lead-time valid at 2100UTC). Top left, 4kft AGL; top right, 5kft AGL; middle left, 8 kft AGL; middle right, 10kft AGL; bottom left, 15kft AGL; bottom right, 20kft AGL.

5.2 FIP Constrained to the AIRMET (FIP-CON)

This section examines FIP's supplemental value when being constrained within the boundary of an AIRMET (Regions 1 and 3 in Figure 4.2)

5.2.1 FIP-CON CONUS Evaluation

The partial JPD in Table 5.2 shows the level of agreement when FIP is constrained to the AIRMET for different altitude ranges (0-10kft, 11-20kft, 21-30kft, and 0-30kft AGL) using no probability mask and a 0.25 probability mask. In this case, the probability of detection of MOG icing events (PODy) for the AIRMET is 1, while the probability of detecting LTM (PODn) is 0. The volume of MOG icing in the FIP grids is considerably smaller than the bounding AIRMET volume for the winter season, with the majority of the icing volume of the FIP grids residing in the 11-20-kft AGL altitude segment. Agreement between the forecasts is highest (~25%) in that layer. From the table, the 0-10-kft layer has the most disagreement with the AIRMET forecast (>99%) regardless of the probability mask used. In addition, there is a large drop in agreement at the highest altitude range when applying the probability mask, as lower probabilities for icing occurrence seem to prevail at this altitude range.

Table 5.2. The count of FIP grids reporting LTM icing and MOG conditions while constrained inside an AIRMET for different altitude ranges and for no probability mask (above) and a 0.25 probability mask (below). Areas of forecast agreement are highlighted.

	0-10kft	11-20kft	21-30kft	0-30kft
No Probability Mask--- Forecast Grid Count				
LTM (disagree)	324130	1568704	550534	2443368
MOG (agree)	2314	521584	159575	683473
Percentage				
LTM (disagree)	99.29%	75.05%	77.53%	78.14%
MOG (agree)	0.71%	24.95%	22.47%	21.86%
0.25 Probability Mask--- Forecast Grid Count				
LTM (disagree)	324221	1684774	658756	2667751
MOG (agree)	2223	405514	51353	459090
Percentage				
LTM (disagree)	99.32%	80.60%	92.77%	85.32%
MOG (agree)	0.68%	19.40%	7.23%	14.68%

Skill by altitude over the CONUS, shown in Figure 5.2, reveals that FIP (when constrained to an AIRMET) adds value on a broad scale by effectively reducing the

MOG icing forecast airspace bounded by the AIRMET. Including all altitudes over the CONUS, the PODn values are approximately 0.42 and 0.49 while PODy values are approximately 0.66 and 0.59, with no probability mask and a 0.25 probability mask, respectively. In the constrained context, it is important to maximize PODn while keeping PODy relatively high. High PODn values for FIP are found in the lowest (0-10kft) and highest levels (21-30kft). At the mid-levels (11-20kft) FIP still identifies roughly 30% of less than MOG PIREPs while retaining greater than 65% of the MOG reports. It appears that not applying a probability mask to the severity field maximizes the value added to the AIRMET forecast by FIP.

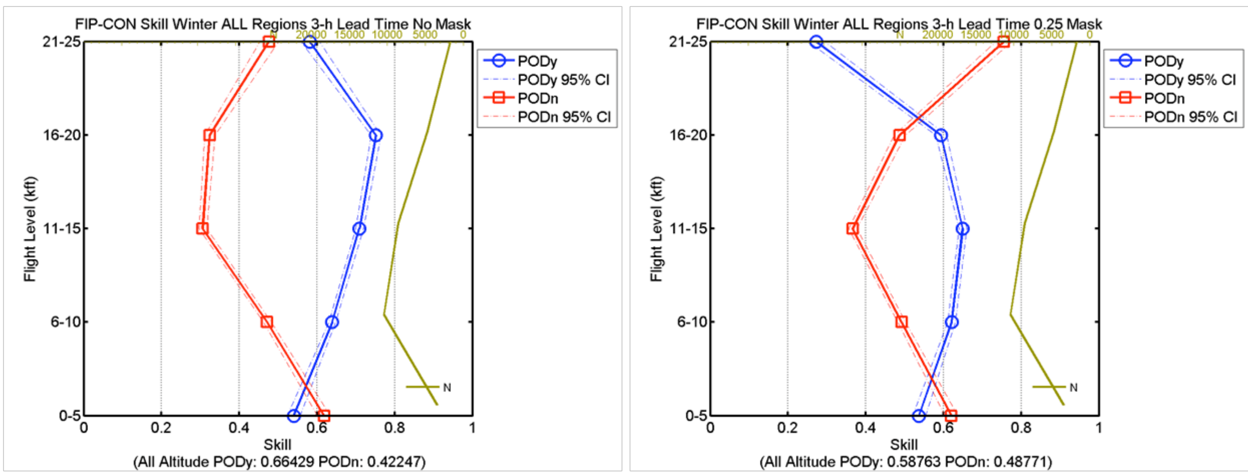


Figure 5.2. Height series of skill (PODy blue circles, PODn red squares) in 5-kft layers for all regions when FIP is constrained to the AIRMET with no probability mask (left) and a 0.25 probability mask (right). The number of PIREPS used in this verification appears as a solid brown line and corresponds to the top x-axis. Confidence intervals are shown as dashed lines.

5.2.2 FIP-CON Regional Evaluation

Although few differences are exhibited in the JPDs over the 15 regions below 20kft AGL, the skill of FIP can vary significantly. This section will explore the geographic differences of FIP when constrained to the AIRMET in the winter season for 3 different height stratifications. The geographic plots in strip plot format are ranked from top to bottom based on a measure of air traffic density using terminal activity, which is described in Section 4.4.

0-10kft AGL

The skill of FIP forecasts at the lowest flight layer varies significantly from region to region; however, within most of the high traffic regions it exhibits skill in detecting less than MOG icing PIREPs while maintaining relatively high detection of MOG icing

PIREPs. The 0-10-kft layer shows almost total disagreement between FIP and the AIRMET using the JPDs from Table 5.2, which tends to indicate that FIP forecasts of MOG icing are more sensitive to atmospheric conditions at the lower levels. Figure 5.3 shows the PODn and PODy statistics by region for this layer without a probability mask. Regions such as WCN and RMN tend to capture more icing events at the lowest levels as evident from the high PODy values, while the corresponding low PODn values may be due to elevation decreasing the total volume of airspace. The APP region and ECS region have similar high PODy values while still getting more than half of the negative icing PIREPs, adding considerable value to an AIRMET forecast at the lowest levels for these regions.

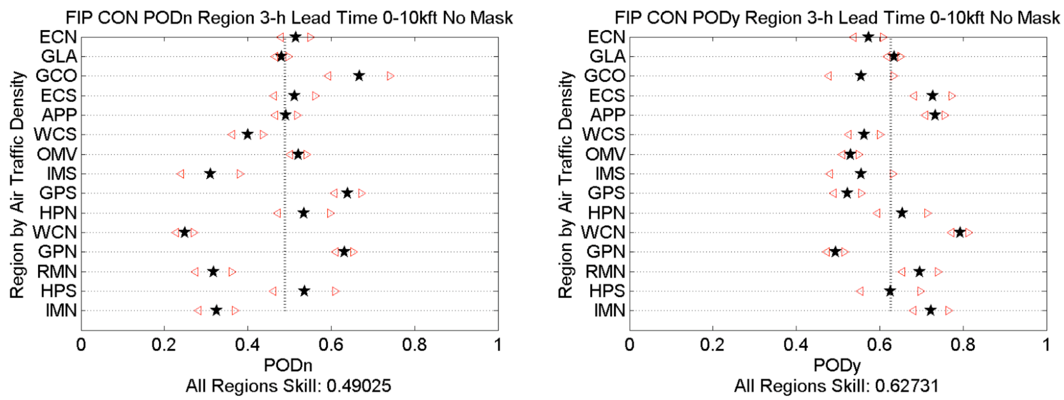


Figure 5.3. PODn (left) and PODy (right) by region with no mask for FIP constrained to the AIRMET (FIP CON) in the 0-10-kft AGL layer. The vertical dotted line represents the average skill of all regions.

11-20kft AGL

When constrained to the AIRMET, the FIP grid successfully retains regions of MOG icing in the mid-levels, areas of concern for aviation, while still identifying some areas of less than MOG icing conditions, albeit, to a lesser extent than the 0-10kft AGL layer. Statistics of PODy and PODn are shown in Figure 5.4 for the 15 regions in this layer where the JPD agreement is maximized between FIP and the AIRMET. The probability of detecting MOG icing for all regions using no mask is approximately 0.73, while detecting less than MOG is approximately 0.31. The low traffic regions of HPN and HPS seem to benefit most, with respect to available airspace, as indicated by PODn values of around 0.5 with corresponding high values of PODy.

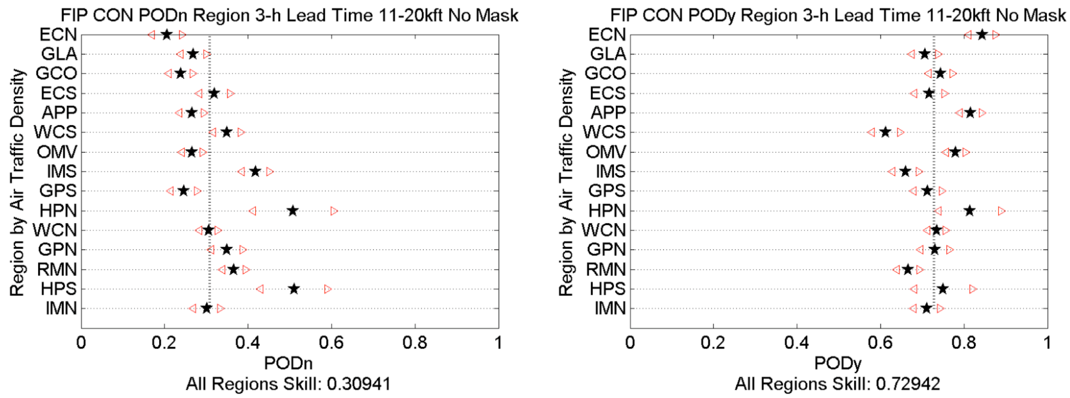


Figure 5.4. PODn (left) and PODy (right) by region with no mask for FIP constrained to the AIRMET in the 11-20-kft AGL layer. The vertical dotted line represents the average skill of all regions.

21-30kft AGL

Occurrence of MOG icing at the high altitudes is a rare event. An examination of the JPDs for the 15 regions reveals that there are a few notable differences in the level of agreement between FIP and AIRMETs at the 21-30-kft layer. For example, there is considerable agreement between AIRMET forecasts and FIP forecasts of MOG icing in the OMV region (~45% of forecast grids), but very low forecast agreement (~1% of forecast grids) in the RMN region (regional JPDs while calculated are not shown in this report). This may be due to the fact that icing may occur at higher altitudes in the more meteorologically dynamic OMV region and is being captured by the AIRMET, while in regions such as RMN icing is a much rarer occurrence at higher altitudes and the AIRMET tends to over-forecast the vertical extent of the icing threat. This may indicate that some AIRMETs bound large domains containing regions of differing atmospheric dynamics.

Although the number of icing PIREPs decreases considerably in the 21-30-kft AGL layer, it is useful to examine the regional statistics (without a mask) since the JPDs show differences in agreement across the regions. Figure 5.5 shows skill scores varying considerably over the different geographic regions at the 3-h lead-time. In this figure, the OMV region shows considerable skill in identifying icing-free regions above 21kft while still maintaining a high degree of positive MOG detections. Similar statistics are found in the RMN region; however, FIP and the AIRMET are in disagreement almost 99% of the time where in the OMV region there is fairly high agreement (~45%).

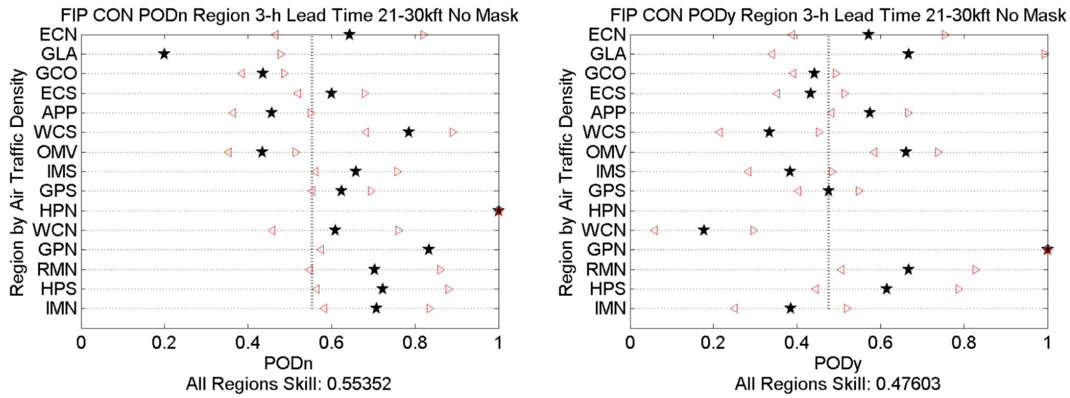


Figure 5.5. PODn (left) and PODy (right) by region with no mask for FIP constrained to the AIRMET (FIP-CON) in the 21-30-kft AGL layer. The vertical dotted line represents the average skill of all regions.

5.2.3 Summary of FIP-CON

It is apparent that value can be added to the AIRMET forecast when using FIP as a supplement in the constrained case by picking up regions of LTM icing severity within the AIRMET boundary. The greatest differences in skill between the 15 geographic regions occur in the 0-10-kft and 21-30-kft AGL layers. The JPDs show that FIP forecast grids and the AIRMET forecast disagree 99% of the time in the lowest layer; however, regions such as APP, ECS, and even the GLA have been shown to positively identify regions of negative icing around 50% of the time while still obtaining more than 60% of the MOG icing PIREPs. In the GLA region this amounts to 1448 out of 2955 LTM icing PIREPs being correctly identified (while still capturing 1069 out of 1646 MOG icing PIREPs) by the 3-h lead-time FIP forecast where there is a MOG icing forecast from the AIRMET. This adds value to the AIRMET forecast when the AIRMETs have volumes that may unnecessarily begin at the surface. The JPDs disagree over some geographic regions when examining the highest layer (21-30kft AGL), which is possibly due to the broad AIRMET covering meteorologically diverse events. This is evident in the drastic JPD differences between the OMV and RMN regions.

5.3 FIP outside the AIRMET Boundary (FIP-NA)

This section examines FIP's supplemental value when outside of the boundary of an AIRMET (Regions 2 and 4 in Figure 4.2)

5.3.1 FIP-NA CONUS Evaluation

Although examination of Table 5.2 will reveal that FIP constrained outside of the AIRMET is in agreement greater than 98% of the time with the AIRMET in terms of less than MOG icing events; it is important to note that icing is a rare event and that the area of disagreement, albeit ~1%, is of great importance. In this case, the probability of detection of MOG icing events (PODy) for the AIRMET is 0, while the probability of detecting LTM (PODn) is 1. As a side note, there are no significant geographic differences in the level of agreement between the forecasts when FIP is constrained outside the AIRMET over the various altitude stratifications.

Table 5.3. The count of FIP grids reporting LTM and MOG conditions while outside an AIRMET for different altitude ranges and for no probability mask (above) and a 0.25 probability mask (below). Agreement is highlighted.

	0-10kft	11-20kft	21-30kft	0-30kft
No Probability Mask---Forecast Grid Count				
LTM (agree)	**9224550**	**7333899**	**8767258**	**25325707**
MOG (disagree)	**12463**	**138828**	**84679**	**235970**
Percentage				
LTM (agree)	**99.87%**	**98.14%**	**99.04%**	**99.08%**
MOG (disagree)	**0.13%**	**1.86%**	**0.96%**	**0.92%**
0.25 Probability Mask--- Forecast Grid Count				
LTM (agree)	**9225035**	**7360939**	**8826428**	**25412402**
MOG (disagree)	**11978**	**111788**	**25509**	**149275**
Percentage				
LTM (agree)	**99.87%**	**98.50%**	**99.71%**	**99.42%**
MOG (disagree)	**0.13%**	**1.50%**	**0.29%**	**0.58%**

Figure 5.6, a height series plot of PODy and PODn, indicates that FIP adds value by correctly identifying regions of significant MOG icing when outside the AIRMET forecast volume. The PODy values are approximately 0.41 and 0.35 while PODy values are approximately 0.77 and 0.80, for no probability mask and a 0.25 probability mask respectively. While outside the AIRMET, it is important for FIP to maximize PODy while keeping PODn relatively high. The highest PODy values (greater than 0.4) for FIP-NA are found in the mid-levels at the expense of a reduction in PODn. FIP may tend to over-forecast at these flight levels. At the highest and lowest flight levels, FIP still picks up greater than 25% of the MOG icing observations while retaining 80% of the less than MOG icing reports. In the case of FIP outside the AIRMET, using the 0.25 probability mask decreases the number of MOG forecast grids which increases the PODn value while still picking up the significant regions of MOG icing.

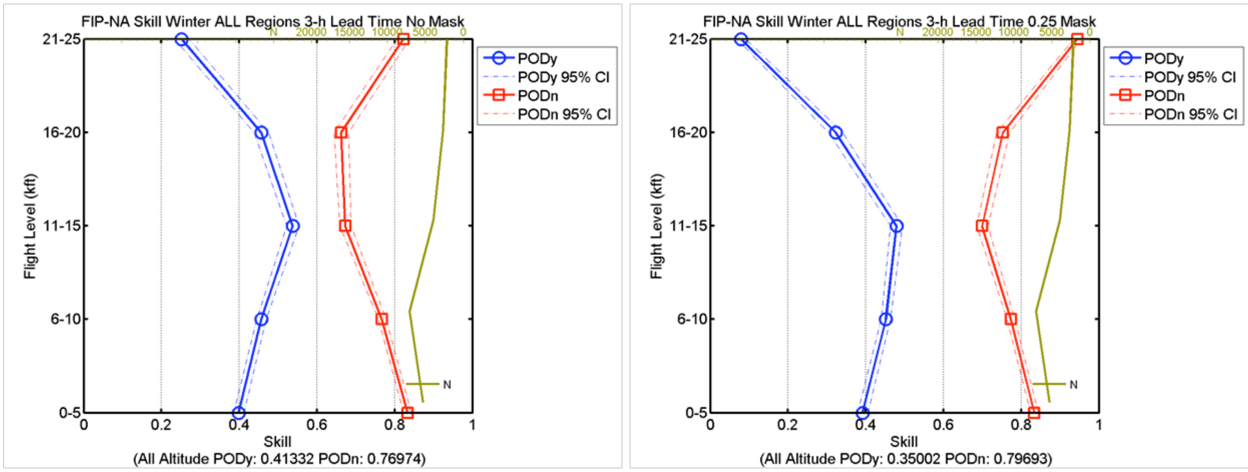


Figure 5.6. Height series of skill (PODy blue circles, PODn red squares) in 5-kft layers for all regions when FIP is outside of the AIRMET with no probability mask (left) and a 0.25 probability mask (right). The number of PIREPS used in this verification appears as a solid brown line and corresponds to the top x-axis. Confidence intervals are shown as dashed lines.

5.3.2 FIP-NA Regional Evaluation

Unlike the FIP-CON case, the JPDs show very little change in the level of agreement between the forecasts for the different geographic regions for all altitudes. The skill scores over the different regions vary less than the FIP-CON case, but some differences are still of significance. This section will explore the geographic differences of FIP when outside the AIRMET in the winter season for three different height stratifications. The geographic plots in strip plot format are ranked from top to bottom based on a measure of air traffic density using terminal activity. Additionally, statistics are produced using the 0.25 probability mask below 20kft.

0-10kft AGL

In the lowest altitude stratification when FIP is forecasting MOG outside the AIRMET, many of the high traffic regions observe MOG icing PIREPs (Figure 5.7). Regions such as OMV, ECN, and GLA appear to be forecasting MOG icing with a higher degree of skill than other regions; correctly identifying about 50% of the MOG PIREPs where the AIRMET indicates no icing while still picking up at least 75% of the less than MOG PIREPs. There are some hints of over-forecasting trends for FIP outside the AIRMET in the RMN and WCN regions, which have higher PODy values combined with relatively lower PODn values than the other regions.

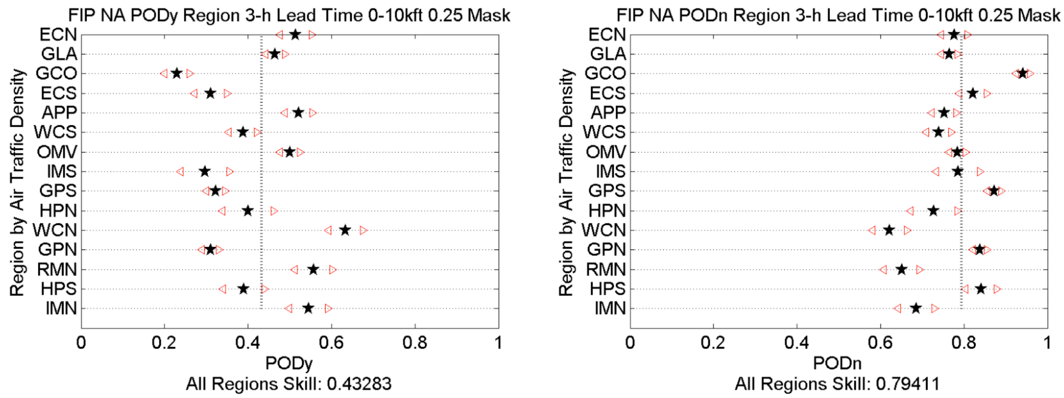


Figure 5.7. PODy (left) and PODn (right) by region with the 0.25 probability mask applied for FIP constrained outside the AIRMET for the 0-10-kft AGL layer. The vertical dotted line represents the average skill of all regions.

11-20kft AGL

In the 11-20-kft AGL layer, there appears to be an indication of FIP over-forecasting relative to the other altitude layers. PODy remains relatively high, but PODn decreases. However, there are a few high terminal traffic regions (ECN, GLA) that exhibit notable skill (Figure 5.8). When there is no AIRMET issued for the ECN region ($PODy_{AIRMET}$=0), FIP is capturing more than 50% of the MOG icing PIREPs while still correctly verifying more than 80% of the less than MOG PIREPs.

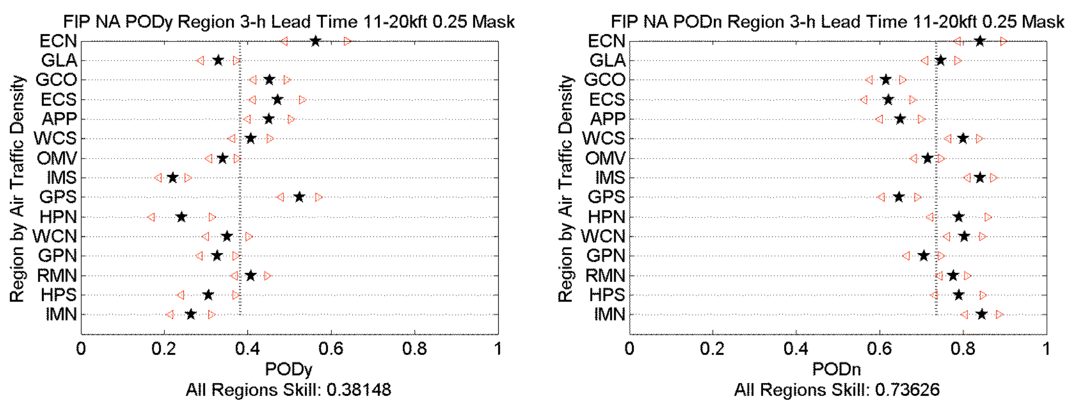


Figure 5.8. PODy (left) and PODn (right) by region with the 0.25 probability mask for FIP outside the AIRMET in the 11-20-kft AGL layer. The vertical dotted line represents the average skill of all regions.

21-30kft AGL

The PODn approaches 1 in the 21-30-kft layer within most regions, with the same high traffic regions (ECN, WCS, and OMV) exhibiting high PODy for MOG icing (Figure 5.9). In this case, no probability mask was used because FIP probability appears to decrease with increasing elevation. Applying the mask at the highest layer reduces the already small number of PIREPs, which eliminates the chance for meaningful statistics. The WCS, OMV, and ECN regions retains a PODn well above 80% while still obtaining 30% of the MOG or PIREPs at this altitude layer.

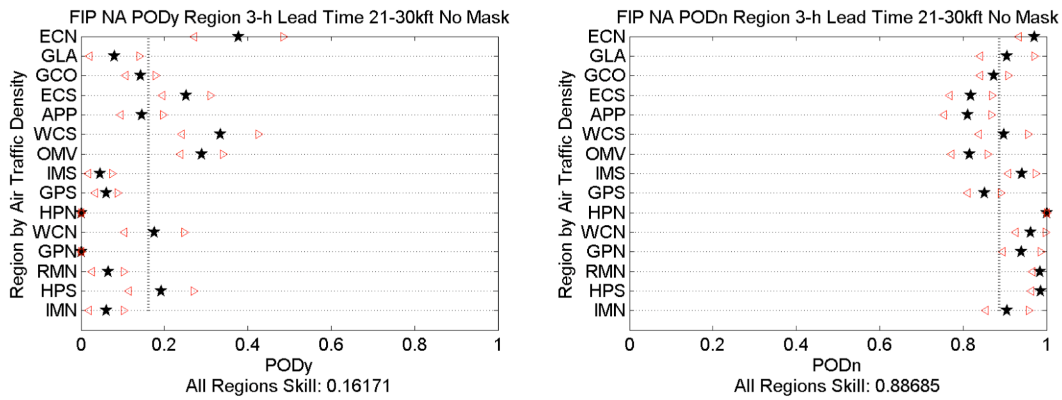


Figure 5.9. PODy (left) and PODn (right) by region without a probability mask applied for FIP constrained outside the AIRMET for the 21-30-kft AGL layer. The vertical dotted line represents the average skill of all regions.

5.3.3 Summary of FIP-NA

Overall, FIP adds value within most regions by identifying significant areas MOG icing where there is no AIRMET ($PODy_{AIRMET}$=0). It detects 30% of the MOG icing PIREPs when outside of the AIRMET while still identifying close to 80% of the less than MOG icing PIREPs for all flight levels within the CONUS. There is some evidence that FIP tends to over-forecast MOG icing in the 11-20-kft AGL layer. When FIP is constrained outside of the AIRMET polygons, it exhibits similar scores for PODn and PODy within most of the regions and throughout the different levels, with some notable exceptions. In the RMN and WCN regions, FIP tends to over-forecast MOG icing in all levels, while in the important air traffic regions ECN, GLA, and OMV it appears to add more value than average for the CONUS. For the 3-h lead-time FIP product in the OMV region, this amounts to capturing 252 out of 639 MOG icing PIREPs that the AIRMET misses while still obtaining 1728 out of 2196 LTM icing PIREPs.

5.4 FIP as a Summer Supplement

During the warmer seasons, when AIRMET icing issuance decreases considerably, FIP may have supplemental value (Table 5.3). The decrease in AIRMET issuances seems to be due to the fact that icing becomes considerably scarcer as the freezing level rises across the CONUS and icing events begin to be found in warm season convective weather (i.e. in the proximity of thunderstorms). Convective SIGMETs are often used to account for some of the missing icing AIRMET volume by including the possibility of icing for this latter point. FIP might provide an independent forecast for explicitly capturing icing events in the warm season and to supplement icing discussions in convective SIGMETs where necessary. The FIP algorithm attempts to identify both broad scale icing and those events on the convective scale.

Table 5.4. Forecast volumes for MOG icing for FIP with and without probability masks compared to the AIRMET for the winter 2006 and summer 2006 seasons.

Volumes (10^6 km^3)		
	Winter 2006	Summer 2006
AIRMET	7.834	0.347
FIP (No Mask)	2.592	1.716
FIP (0.25 Mask)	1.735	1.313

Skill for the summer season is shown for the FIP forecast in height series format for the 3-h lead-time for all regions in Figure 5.10. Due to the lack of volume of the icing AIRMET the probability of detecting MOG icing PIREPs is low compared to the independent FIP product above 10kft AGL. For all altitudes in summer season FIP captures greater than 60% of the MOG icing reports as well as 70% of the less than MOG PIREPs.

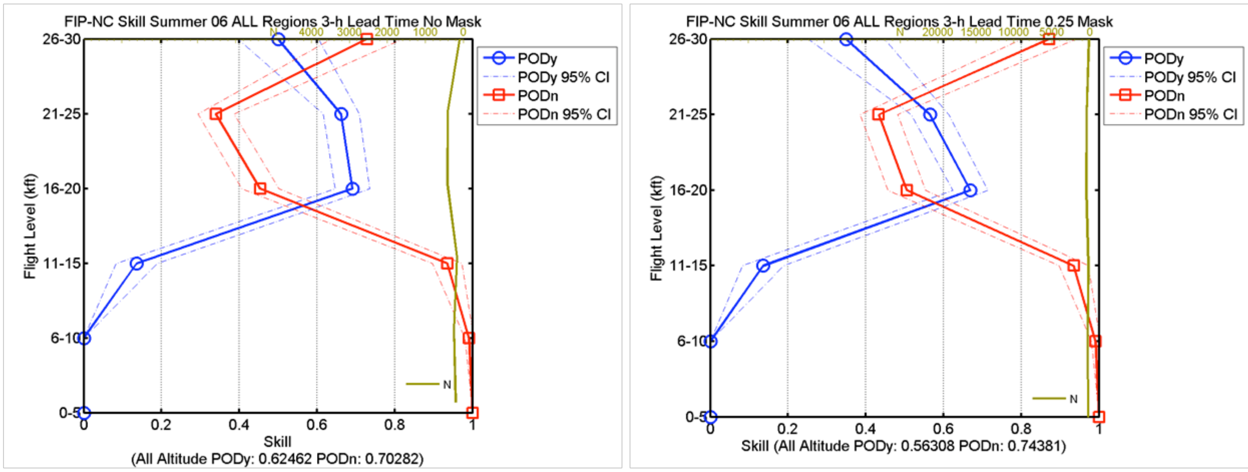


Figure 5.10. Height series of skill (PODy blue circles, PODn red squares) in 5-kft layers for all regions for the independent FIP forecast with no probability mask (left) and a 0.25 probability mask (right). The number of PIREPS used in this verification appears as a solid brown line and corresponds to the top x-axis. Confidence intervals are shown as dashed lines.

5.5 FIP Performance over Lead Time

This section evaluates the discrimination of the 5 FIP severity categories with PIREP intensity values. It also examines the reliability of the FIP probability field, which is used to mask the severity field in this study. In addition, an evaluation of the independent FIP product to FIP constrained inside the AIRMET over lead-time is performed.

5.5.1 FIP Severity Discrimination and Reliability of Probability

Overall, the distributions of both FIP severity and PIREP intensity do not vary for altitudes below 20kft or FIP lead-times. Trends in forecasting icing severity from FIP are shown by distributions of the count of the grids forecasting specific icing severity levels over all lead-times when conditioned on a PIREP observation. Figure 5.11 shows that for all altitudes (left) and for the 21-30-kft layer (right) the forecast distribution of severity does not change with lead time. There is a notable difference, however, in the distribution in FIP severity for the 21-30-kft layer. In the layers less than 20kft AGL, FIP tends to forecast just as many MOG icing grids (FIP severity category 3 and 4) than less than MOG icing grids (FIP severity category 0,1, and 2). In the layers greater than 20kft AGL, FIP begins to decrease the number of MOG icing grids in favor of less than MOG. PIREPs tend to observe trace and light amounts of icing (PIREP intensity category 1) for all FIP lead-times and all altitudes.

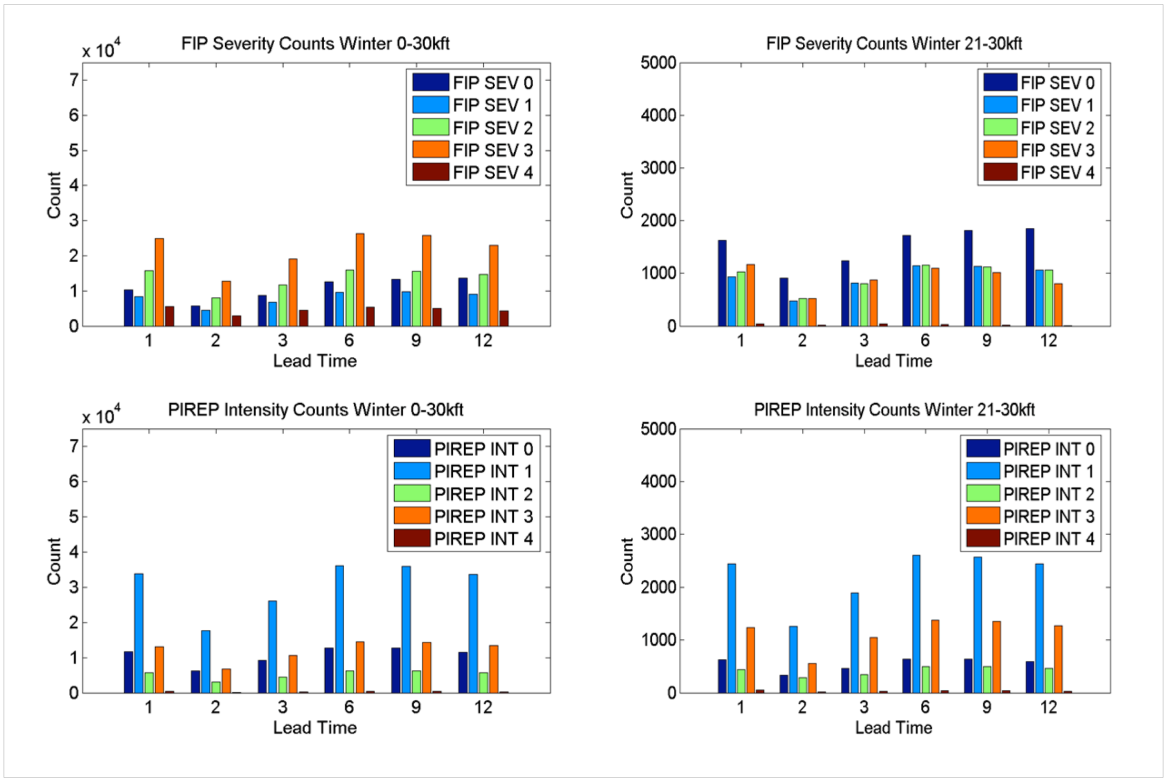


Figure 5.11. FIP severity counts by category (top) and PIREP intensity counts by category (bottom) for all lead-times and all altitudes (left) and at the 21-30-kft AGL layer (right).

This distribution result is reflected in the discrimination plot of FIP severity forecasts to PIREP intensity observations in Figure 5.12. FIP tends to forecast MOG icing for all PIREP intensity categories greater than 1 for all altitudes (Figure 5.12 left). For the altitudes 21kft and greater there is more of a 1:1 relationship between FIP severity and PIREP intensity which disappears when PIREP intensity reaches category 3 or higher. FIP no longer supports as many MOG forecasts in the upper altitudes where icing is a much rarer event.

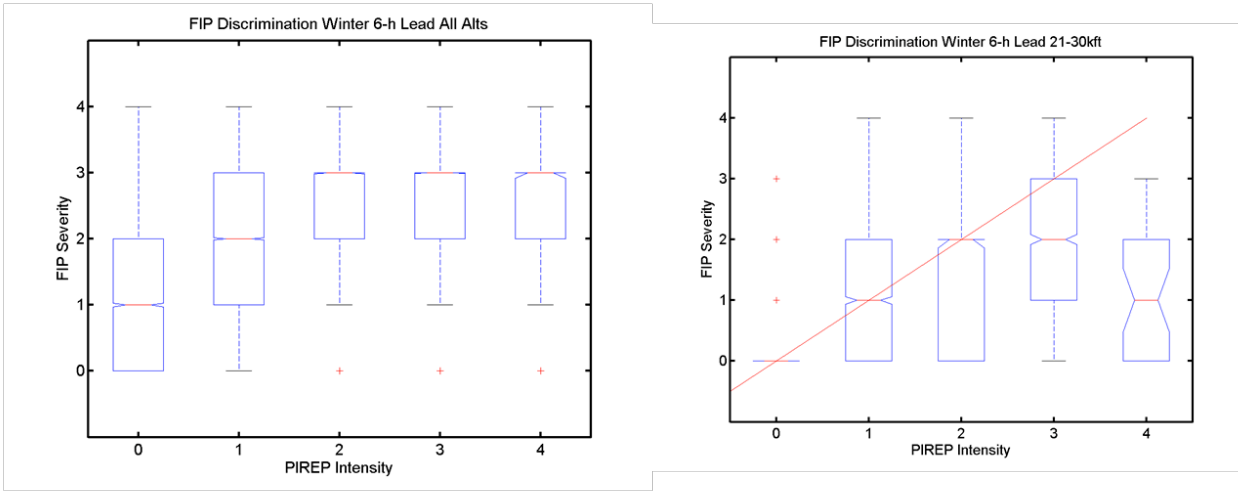


Figure 5.12. FIP severity corresponding to PIREP intensity for the 6-h lead-time for all altitudes (left) and for the 21-30-kft AGL layer (right) with the 1:1 trend line (red) added.

FIP appears to handle forecast uncertainty by only suppressing the probability field rather than suppressing both severity and probability as lead time increases. It is therefore necessary to explore the FIP probability field and the reliability of the probability field. Figure 5.13 shows the decrease in mean probability and volume for MOG icing forecasts over lead-time for all altitudes. The result is a decrease in volume over lead-time for MOG icing for different probability masks (Figure 5.13 right). Figure 5.14 shows a box plot diagram of reliability illustrating FIP has no resolution beyond the 0.25 probability threshold during the winter 2006 season. Similar reliability trends are found in all seasons.

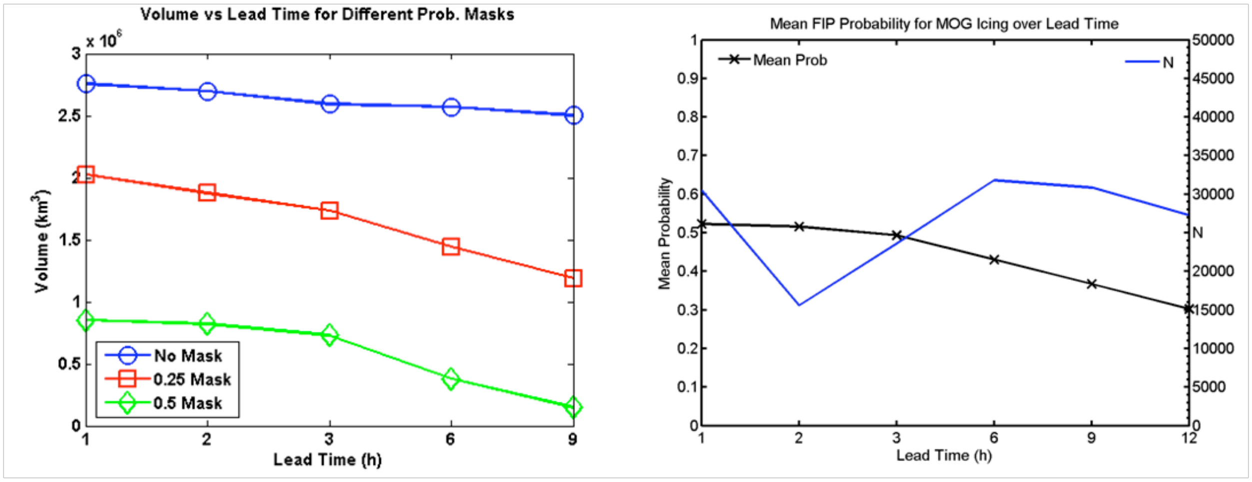


Figure 5.13. FIP volume (km^3) using no mask (blue), the 0.25 mask (green), and the 0.5 mask (red) over lead time (left), and mean FIP probability (black) with N (blue) for MOG icing over lead time (right).

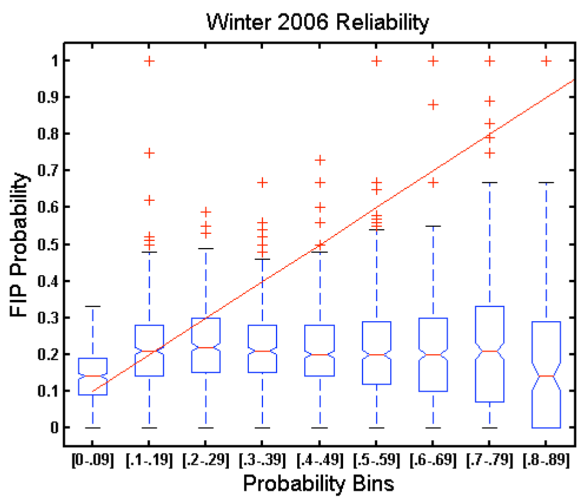


Figure 5.14. FIP reliability with a 1:1 trend line (red) added for the winter 2006 season.

5.5.2 FIP-CON vs. FIP-NC over Lead-Time

The skill of FIP constrained to the AIRMET with no probability mask applied is consistently higher than the independent FIP product over lead time as evident by Figure 5.15. The probability of detecting MOG icing when FIP is constrained inside the AIRMET is higher than the unconstrained FIP at the expense of PODn. However, an examination of volume efficiency amplifies the skill of FIP in the constrained cases. In this case, the AIRMET acts as a filter for FIP in the constrained case and will therefore have a smaller volume percentage than the independent FIP counterpart. Volume efficiency is more than a factor of 1.5 higher for the constrained FIP product.

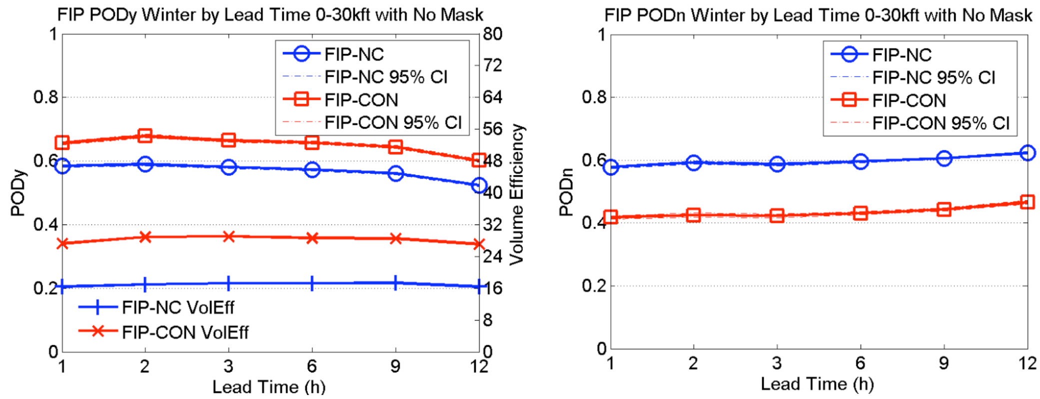


Figure 5.15. FIP constrained inside the AIRMET (FIP-CON) skill compared with independent FIP (FIP-NC) over lead time in terms of PODy and volume efficiency (left) and PODn (right) for all altitudes over winter 2006 and 2007 using no probability mask.

In terms of volume efficiency, using the AIRMET as a filter for FIP may add additional value in the forecasting process when the 0.25 mask is applied (Figure 5.16). As found in the previous section, FIP probability decreases over lead time, thus decreasing FIP's forecast volume over time and causing volume efficiency to increase with increasing lead time. Volume efficiency nearly doubles when using the 0.25 probability mask instead of no mask over this period for both FIP-CON and FIP-NC. However, skill in terms of PODy has a steeper decline over lead time than using no probability mask.

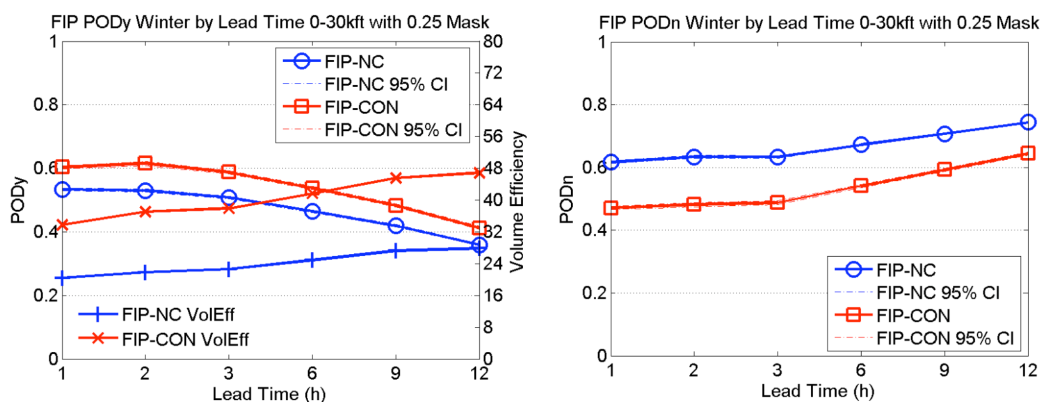


Figure 5.16. FIP constrained inside the AIRMET (FIP CON) skill compared with independent FIP (FIP NC) over lead time in terms of PODy and volume efficiency (left) and PODn (right) for all altitudes over winter 2006 and 2007 using the 0.25 probability mask.

5.6 FIP as an Independent Product

Although FIP is designated as a supplementary product for operational use, in this section FIP is assessed as an independent product using AIRMET skill for a baseline. The FIP algorithm does not provide a “smeared” output, so a direct comparison of the products is impossible. An aggregate FIP forecast is used as a proxy by averaging skill for the 1-,2-,3-, and 6-h lead-times aligned with each AIRMET issue time and valid period for the winter 2006 and 2007 season. No probability mask is applied to the FIP forecast in order to preserve the maximum amount of forecast volume. This methodology is explained in more detail in Section 4.

Relative to FIP, the probability of detecting MOG icing is greater for the AIRMET, at the expense of forecast volume. The AIRMET volume is much larger when compared to FIP volume for MOG icing at any given time. However, it is important to note that if FIP was smeared the volume would increase by a factor of roughly 1.5 to account for the movement of the icing feature over time. Figure 5.17 shows FIP aggregate skill scores compared to the AIRMET skill scores for all altitudes over the issue times of the AIRMET. When accounting for the volume of FIP compared to the volume of the AIRMET, FIP becomes the more effective forecast in terms of the volume efficiency statistic for all issue times. However, after considering increasing the FIP volume by a factor of 1.5, the volume efficiency for FIP decreases to approximately 11, a value closer to the AIRMET volume efficiency. In addition, the over-forecasting trend of the AIRMET is reflected in the lower PODn statistic. PODy and PODn are similar for both the independent FIP product and the AIRMET above 10kft AGL; however, FIP achieves similar skill with lower volume.

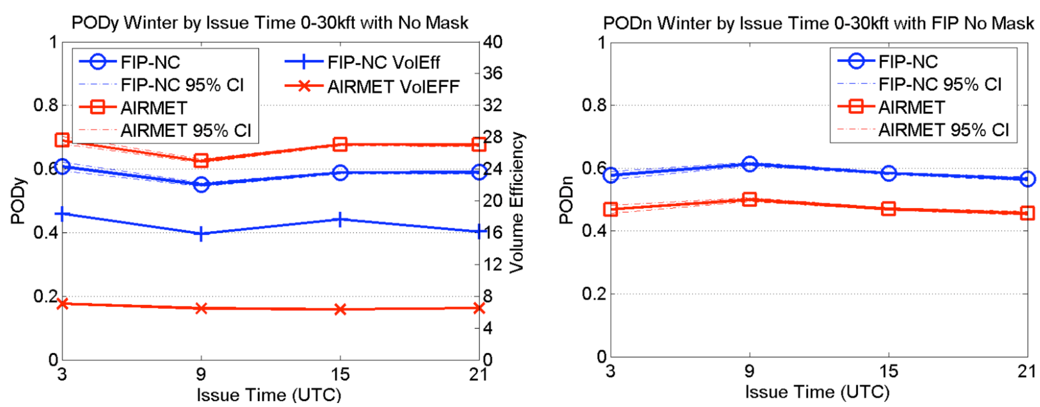


Figure 5.17. PODy and volume efficiency (left) and PODn (right) for detecting MOG icing PIREPs for the FIP aggregate (blue) compared to the AIRMET (red) over the issue time of the AIRMET for all altitudes.

A different story appears when using the independent FIP aggregate for detecting icing of any intensity compared to the AIRMET's skill of capturing a PIREP of any icing (PIREP intensity greater than 0). The independent FIP product over-forecasts icing volume when considering icing of any intensity (FIP severity greater than 0 with no probability mask). This is illustrated by examining volume efficiency in Figure 5.18. The adjusted "smeared" volume of FIP would further decrease the volume efficiency.

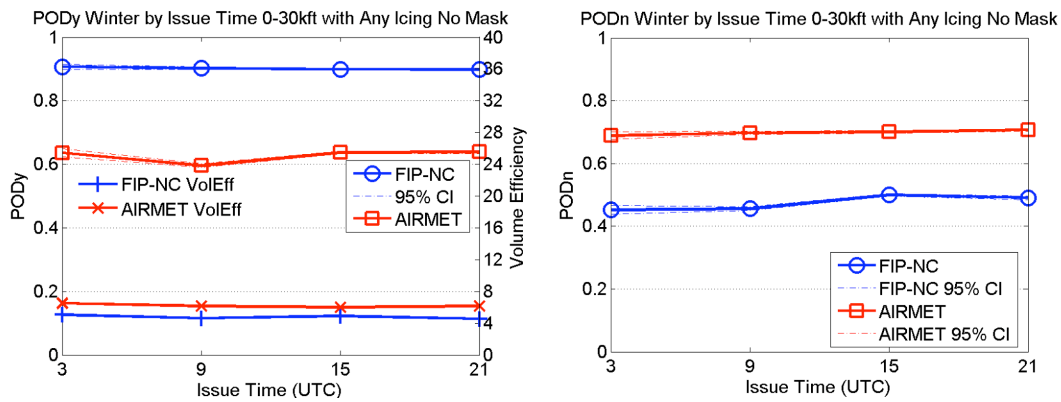


Figure 5.18. PODy and volume efficiency (left) and PODn (right) for detecting PIREPs of any icing for the FIP aggregate (blue) compared to the AIRMET (red) over the issue time of the AIRMET for all altitudes.

5.7 FIP Supercooled Large Droplet (SLD) Performance

Two methods are used to assess the skill of SLD performance in the FIP product. The first is a direct comparison of a PIREP report of SLD to a corresponding FIP forecast while the second involves a METAR-based approach.

An SLD PIREP contains an SLD, freezing rain, or freezing drizzle remark, or reports a severe intensity value (intensity 5) of clear icing type. The population of SLD PIREPs is relatively small compared to the population of PIREPs for all seasons. Most SLD PIREPs were found to occur below 20kft in the winter season, and were scattered throughout the CONUS. For the 3-h lead-time, there were 27 PIREPs of SLD and only 1 of those was matched with a positive SLD forecast. For the 6-h lead time, only 8 out of 58 PIREPS were captured by FIP SLD. Similar statistics can be found for other lead and issue time combinations. Overall, there is too small of an SLD PIREP population to yield useful statistics on the performance of FIP SLD.

METAR-based SLD approach is based upon the detection of freezing rain or freezing drizzle at a surface observing station. The midpoint of the duration of the freezing precipitation event is used as a valid time to compare to the FIP SLD valid time. In addition, the ceiling value reported by the METAR is used to estimate the depth of the observed SLD layer. Three different layer thicknesses, 1-, 3-, and 5-kft above the observed ceiling, are used in the spatial matching of the observed event to the FIP grid. Table 4 shows the FIP SLD identification success rate for the 3- and 6-h lead-times for the three layers used. Overall, FIP successfully identifies at least half of the METAR observations of SLD correctly. This adds meaningful context to SLD skill as PIREPs have been shown to be sparse and unreliable for SLD occurrence.

Table 5.5. FIP matches to METAR SLD observations for the 3- and 6-h lead-time for all FIP issue times.

3-h Lead-Time										
FIP Issue (UTC)	0	3	6	9	12	15	18	21	Total	POD
SLD METAR Count	118	188	190	252	208	174	114	126	1370	
FIP Matches-1000 ft	56	97	81	104	111	94	63	69	675	**49.27%**
FIP Matches-3000 ft	64	108	90	128	118	105	69	78	760	**55.47%**
FIP Matches-5000 ft	66	110	98	131	122	109	72	82	790	**57.66%**
6-h Lead-Time										
SLD METAR Count	258	263	374	347	258	182	176	190	2048	
FIP Matches-1000 ft	133	136	169	159	164	99	86	93	1039	**50.73%**
FIP Matches-3000 ft	148	152	195	177	176	102	93	106	1149	**56.10%**
FIP Matches-5000 ft	155	159	210	186	180	103	101	110	1204	**58.79%**

6. Conclusions

This report primarily evaluated FIP as a supplement to the icing AIRMET. Agreement between the two forecasts was measured. Then, the skill of the supplemental product was examined in two ways: constraining the grid to within the boundaries of AIRMET polygons and constraining the grid to outside of the boundaries of the polygons. The performance of FIP was also assessed during the summer season, a time when icing AIRMET issuances substantially decrease. Finally, FIP was considered as an independent product and a reasonable attempt was made to compare its skill with that of the icing AIRMET. Results from the study indicated:

- Qualitatively, FIP appears to effectively identify the structure of icing within the broad area outlined by the AIRMET polygons. Within an AIRMET polygon, FIP agrees with the operational forecast of icing in only about one-fifth of the forecast volume. However, FIP captures over 42% of the no-icing PIREPS while retaining a PODy of 0.66.
- FIP appears to identify some areas of icing that either weren't captured by the AIRMET or didn't meet minimum criteria for issuance. Outside of an AIRMET polygon, FIP agrees with the operational forecast of no icing in over 98% of the volume. FIP captures over 41% of the yes-icing PIREPS while retaining a PODn of 0.77.
- FIP provides significant support to the icing AIRMET forecast in the summer season by capturing 60% of the icing reports and 70% on the no-icing reports.
- In terminal areas, where rare events of SLD present a significant hazard, the FIP SLD forecast identifies over 50% of the observed freezing rain and freezing drizzle events.
- When FIP is considered as an independent product, its performance is similar to that of the icing AIRMET with respect to measures of PODy and PODn. However, FIP demonstrates an apparent improvement in the volume efficiency of capturing reports of icing.
- Overall, this study found FIP severity to perform best with no probability mask. The experimental Aviation Digital Display Service (ADDS) allows the icing severity field to be “masked” by the icing probability field. Values include no mask, 0.25 probability, or 0.5 probability.

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 or position of the FAA. The authors would like to thank the In-Flight Icing Research Team for providing the forecast data that was needed for the evaluation.

7. References

Benjamin, S.G., J.M. Brown, K.J. Brundage, B.E. Schwartz, T.G. Smirnova, and T.L. Smith, 1998: The operational RUC-2, *Preprints, 16th Conference on Weather Analysis and Forecasting,* Phoenix, AZ, American Meteorological Society (Boston), 249-252.

Benjamin, S.G., T.G. Smirnova, K.J. Brundage, S.S. Weygandt, T.L. Smith, B. Schwartz, D. Devenyi, J.M. Brown, and G.A. Grell, 2004: A 13-km RUC and beyond: Recent developments and future plans. *11th Conference on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, October 2004, American Meteorological Society (Boston).

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., 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, Sep. 11-15, American Meteorological Society (Boston), 393-398.

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).

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, 13-16 May, American Meteorological Society (Boston), 311-314.

Carriere, J.M., S. Alquier, C. LeBot, and E. Moulin, 1997: Statistical Verification of Forecast Icing Risk Indicies, *Meteorological Applications*, Vol 4, Issue 2, p.115-130.

Chapman, M., M. Pocernich, A. Holmes, P. Boylan, P. Kucera, B.G. Brown, J.L. Mahoney, and J.T. Braid, 2007: Quality Assessment Report: Forecast Icing Product (FIP) – Severity. Report to the FAA Aviation Weather Technology Transfer Board. Available from the Quality Assessment Research Team (Jennifer.Mahoney@noaa.gov).

Federal Aviation Administration, 2008: Aeronautical Information Manual (AIM), Official Guide to Basic Flight Information and ATC Procedures. Available at http://www.faa.gov/atpubs/, p. 7-1-4 – 7-1-5.

Kane, T.L., B.G. Brown, and R.T. Bruintjes, 1998: Characteristics of Pilot Reports of Icing, *Preprints, 14th Conference on Probability and Statistics*, Phoenix, AZ, Jan. 11-16, American Meteorological Society (Boston), p. 90-95.

Loughe, A.F., S. Madine, J. Mahoney, and M. Graf, 2008: A Lead-Time Metric for Assessing Skill in Forecasting the Onset of IFR Events. Preprints, 13th Conference on

Aviation, Range, and Aerospace Meteorology, New Orleans, LA, Jan. 21-24, American Meteorological Society (Boston).

McDonough, F., B.C. Bernstein, and M.K. Politovich, 2003: The Forecast Icing Potential (FIP) Technical Description. Report to the FAA Aviation Weather Technology Transfer Board. Available from M.K. Politovich (NCAR, P.O. Box 3000, Boulder, CO, 80307), 30 pp.

NWS, 2007: Aviation Weather Services, Advisory Circular AC 00-45F. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, and U.S. Department of Transportation, 393 pp.

Wolff, C.A., F. McDonough, M.K. Politovich, B.C. Bernstein, and G.M. Cunning, 2003: FIP Severity Technical Document. Report to the Aviation Weather Technology Transfer Technical Review Board. Available from the Inflight Icing Research Team (marcia@ucar.edu), 44 pp.