NOAA Technical Memorandum OAR GSD-59
https://doi.org/10.7289/V5/TM-OAR-GSD-59

Icing Product Alaska Diagnosis (IPA-D) Assessment

August 2017

Kenneth R. Fenton Jr.
Matthew S. Wandishin
Joan E. Hart
Melissa A. Petty

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
August 2017

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR GSD-59

Icing Product Alaska Diagnosis (IPA-D) Assessment

Kenneth R. Fenton Jr.[1]
Matthew S. Wandishin[2]
Joan E. Hart[2]
Melissa A. Petty[1]

[1] Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

[2] Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD

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.

UNITED STATES DEPARTMENT OF COMMERCE

Wilbur Ross
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

RDML Tim Gallaudet, Ph.D., USN Ret.
Under Secretary for Oceans
And Atmosphere/NOAA Administrator

Office of Oceanic and Atmospheric Research

Craig N. McLean Assistant Administrator

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

The Quality Assessment (QA) Product Development Team (PDT) was tasked to complete an assessment of the Icing Product Alaska Diagnosis (IPA-D) algorithm. IPA-D was developed by the In-Flight Icing (IFI) PDT at the National Center for Atmospheric Research (NCAR) and provides a real-time analysis of icing conditions in the Alaska region by assimilating observations to a model background field. IPA-D was evaluated against the Icing Product Alaska Forecast (IPA-F), as well as the Forecast Icing Potential (FIP) and Forecast Icing Severity (FIS) algorithms over a period from September 15, 2016 through January 15, 2017.

The purpose of this verification study was to assess the IPA-D algorithm for potential use as a situational awareness tool and for potential transition into operations by the Alaska Aviation Weather Unit (AAWU). The results of this study are intended to inform the Technical Review Panel (TRP) as well as establish a baseline of performance for future versions of the product.

The assessment findings include:

- IPA-D exhibited similar skill to the IPA-F 2-Hour and FIP/FIS 0-Hour algorithms
- IPA-D produced smaller volumes of icing potential at the cost of a lower detection rate
- IPA-D had lower probabilities of detection (more missed events) than both IPA-F and FIP/FIS
- IPA-D had lower false alarm rates, thus indicating more available airspace than IPA-F and FIP/FIS
- IPA-D had a much larger percentage of fill values than IPA-F

IPA-D identified less icing overall than IPA-F and FIP/FIS, which was reflected in both the probability and severity fields (section 6.1.2). Specifically, IPA-D contained fewer counts of icing at all positive severity levels (light, moderate, and severe) than IPA-F. With the overall characteristic of less positive icing than FIS, IPA-D distributed the positive icing through fewer counts of light icing, but more counts of moderate and severe icing than FIS.

It was found in the probability distributions that IPA-D contained a significantly higher number of fill values compared to IPA-F (section 6.1.1). While the fill values in IPA-F were used to mask areas where the algorithm grid points were below ground level, the fill values in IPA-D extended to all vertical levels and regions, including over the oceans. The IFI PDT has been made aware of this issue and is working to modify the IPA-D code to address the fill values. For the purposes of the assessment, 17% of observations from the Tropospheric Airborne Meteorological Data Reporting (TAMDAR) data set were excluded from the statistics because the nearest IPA-D grid point contained a fill value.

IPA-D's fewer counts of icing led to smaller volumes of icing identified by the algorithm, which resulted in a greater number of missed events, but also more available airspace than the other algorithms. Using pilot reports (PIREPs), IPA-D had a lower probability of detection (POD) than IPA-F at all severity thresholds; however, IPA-D did so with lower volumes of icing identified (section 6.2.1). Compared to FIP/FIS, at the light or greater (LOG) icing threshold, IPA-D had a lower POD. At the moderate or greater (MOG) threshold, however, IPA-D had both a lower POD and a higher volume of icing than FIP/FIS, meaning that it identified less icing using a larger volume. In general, IPA-D had lower PODs than IPA-F and FIP/FIS, but

also lower probabilities of false detection (POFDs), illustrated in Figure 1. The resulting Pierce Skill Scores (PSSs) were similar among the three algorithms and within a 95% confidence level of one another. Additionally, IPA-D performed relatively better with light icing and relatively worse with MOG icing events.

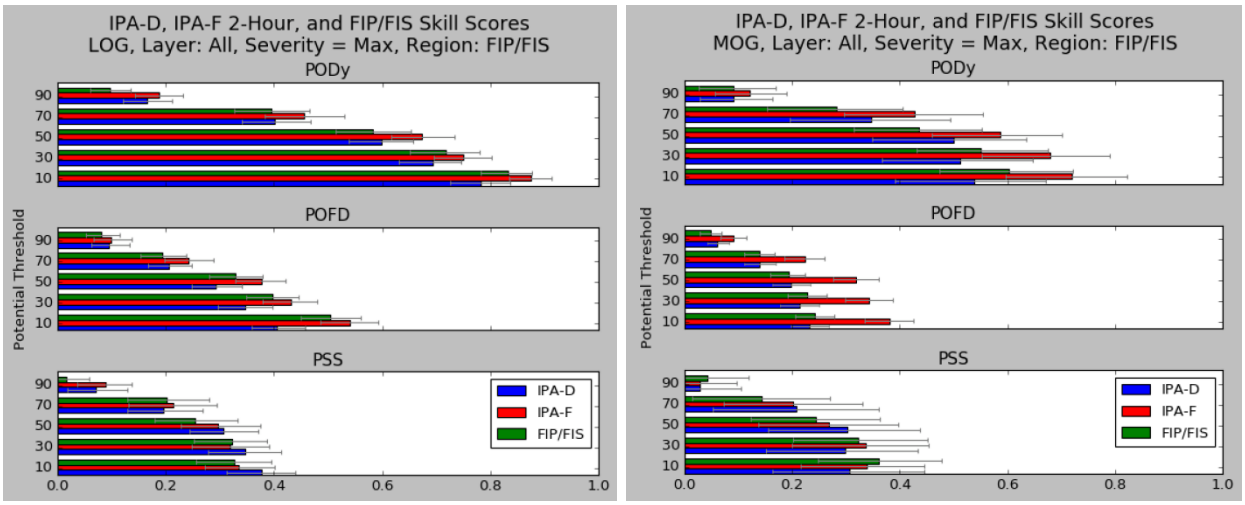

FIGURE 1. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE), IPA-F (RED), AND FIP/FIS (GREEN) IN THE FIP/FIS DOMAIN. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF LOG ICING (LEFT) AND MOG ICING (RIGHT). ERROR BARS INDICATE THE 95% CONFIDENCE LEVEL.

The results from the PIREP analysis were corroborated by the other observation platforms. Only positive reports of icing were used from TAMDAR, which meant that only POD could be calculated from this observation source. Similar to the PIREP results, IPA-D had lower PODs than IPA-F and FIP/FIS (section 6.2.2). As before, the difference in POD between IPA-D and IPA-F was greater at the MOG threshold than the LOG threshold. Additionally, the radiosonde data was used to calculate POD and POFD. Here IPA-D had lower PODs than IPA-F, but higher PODs than FIP/FIS (section 6.2.3). The radiosonde-derived POFDs for IPA-D were again lower than IPA-F, but similar to FIP/FIS. The satellite data revealed similar results: satellite-derived PODs and POFDs for IPA-D were lower than both IPA-F and FIP/FIS (section 6.2.4).

The results are summarized in Table 1. For all observation platforms, IPA-D yields a lower POD and POFD than IPA-F; IPA-D yields a similar or lower POD than FIP/FIS (except against radiosondes), with mixed results relative to FIP/FIS on POFD. As suggested by the POFD scores, the lower detection rate for IPA-D is associated with smaller forecast volumes (i.e., more free airspace). Despite these differences, the resulting Peirce Skill Scores (PSS; scored against PIREPs only) were similar among the three algorithms and within a 95% confidence level of one another. Note that, while the skill of IPA-D is similar to that of IPA-F and FIP/FIS, the use of IPA-D involves a different assessment of the relative risks of missed events and false alarms.

TABLE 1. SUMMARY OF POD AND POFD RESULTS BY EACH OBSERVATION PLATFORM.

	POD (higher is desirable)		POFD (lower is desirable)	
PIREP	IPA-D < IPA-F	IPA-D <= FIP/FIS*	IPA-D < IPA-F	IPA-D <= FIP/FIS**
TAMDAR	IPA-D < IPA-F	IPA-D < FIP/FIS	N/A	N/A
Radiosonde	IPA-D < IPA-F	IPA-D >= FIP/FIS	IPA-D < IPA-F	IPA-D > FIP/FIS
Satellite	IPA-D < IPA-F	IPA-D < FIP/FIS	IPA-D < IPA-F	IPA-D < FIP/FIS

* A reversal occurred when masking with the highest probability thresholds.
** For MOG icing, IPA-D and FIP/FIS yield similar POFDs

Some differences were noted when evaluating the performance of the algorithms regionally. First, IPA-D identified more icing than IPA-F in mountainous regions, in particular in the Alaska Range (section 6.3.1). This is counter to the other regions where IPA-F identified more icing than IPA-D, consistent with the relative distributions discussed above. The differences between IPA-D and FIP/FIS were most apparent in the Alaska Range and in the Coast and Columbia Mountains in Northern British Columbia. The PIREP-computed skill scores for all three algorithms were similar through the six regions of Alaska, with IPA-D finishing with a slightly higher PSS in the North Slope, Kotzebue and Norton Sound, South Central, and Gulf and Southeast regions (section 6.3.2). Again, the general trend was for IPA-D to have both lower PODs and POFDs than IPA-F and FIP/FIS in each region.

In addition, the results were analyzed by vertical layer. Once more, the skill scores were similar and IPA-D had the lowest PODs and POFDs among the three algorithms (section 6.4.2). IPA-D performed relatively better at altitudes at and below 12,000 ft.

Next, the SLD output of IPA-D was compared to that of IPA-F. The percentage of grid points that were positive for SLD was similarly small for both algorithms, each being less than 0.01% (section 6.5.1). The skill scores based on METARs treated IPA-D and IPA-F grid points that had an unknown value for SLD as positive identifications of SLD. The resulting POD, POFD, and PSS were nearly identical between the two algorithms.

Finally, three case studies were selected by joint coordination between the QA PDT and Alaska Aviation Weather Unit (AAWU). The first case featured 11 PIREPs of light and moderate icing around the Anchorage Bowl on December 1, 2016 (section 6.6.1). IPA-D performed the best in this scenario, having the highest probabilities for icing of the three algorithms, although the fill values near some of the PIREPs were a concern. The second case occurred on December 22, 2016 over western Alaska, with 13 reports of light through severe icing (section 6.6.2). This event highlighted the tailored areas of icing that IPA-D identified, as opposed to the larger and broader areas of IPA-F and FIP/FIS. All three algorithms struggled to accurately identify the icing in this case, but IPA-F and FIP/FIS were able to miss less of the icing than IPA-D due to their larger spatial extents of icing coverage. Lastly, a case with five PIREPs of icing near Ketchikan on January 13, 2017 was analyzed (section 6.6.3). All three algorithms did well in this scenario and captured higher probabilities and MOG icing near the PIREP locations. IPA-D performed slightly better than the other two algorithms in this event, owing to its more tailored areas of icing that correctly identified the severities.

2 INTRODUCTION

Icing Product Alaska Diagnosis (IPA-D) is the complimentary algorithm to the Icing Product Alaska Forecast (IPA-F), previously evaluated by the QA PDT in August 2016. The QA PDT used similar data sources and metrics in this verification analysis of IPA-D. A notable difference though, was the inclusion of an additional verification source, the Tropospheric Airborne Meteorological Data Reporting (TAMDAR) aircraft reports.

There are several anticipated use cases for this product. First, IPA-D may be used during shift change to inform the on-coming shift of the current state of icing over the Alaska region. Second, IPA-D could be used as a daily calibration of the Icing Product Alaska Forecast (IPA-F) during the manual model initialization performed by forecasters. Finally, IPA-D may also be used as a situational awareness tool for forecasters from which they can make comparisons to the forecasts of IPA-F and amend those forecasts as necessary.

This assessment compares IPA-D against algorithms currently in operational or experimental use by the AAWU. Specifically, IPA-D output was quantitatively compared to the two and three-hour forecasts of IPA-F and the zero-hour forecasts the Forecast Icing Potential (FIP) and Forecast Icing Severity (FIS) models. The observational data against which these models was evaluated included pilot reports (PIREPs), TAMDAR, aviation routine weather reports (METARs), radiosonde upper air observations, and satellite data from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) and the CloudSat satellites.

3 DATA

This section describes the situational awareness products and observation data that were used during the assessment. Additionally, spatial and geographic stratifications are discussed. The time period for the data collection was 15 September 2016 through 15 January 2017.

3.1 SITUATIONAL AWARENESS PRODUCTS

3.1.1 IPA-D

The IPA-D algorithm assimilates satellite and observation data against a model background field to produce a diagnostic of the icing environment. Specifically, IPA-D is able to use METARs, PIREPs, Geostationary Operational Environmental Satellite (GOES), and Polar Operational Environmental Satellite (POES) inputs to modify the Alaska cutout of the Rapid Refresh (RAP) grid. The receipt of a new GOES image triggers the algorithm to start processing all observation data available at that time. The product issuance frequency is configurable and can be set to produce up to four diagnostics per hour. This assessment used diagnostics produced at the top of the hour and did not consider the other products valid at 15, 30, or 45 minutes past the hour. One exception occurred when IPA-D was not available at the top of the hour, in which case the XX:45 diagnostic from the previous hour was used in its place.

The IPA-D output was made available at 11.25-km horizontal grid spacing and 500-foot vertical increments from 500 through 30,000 feet. Only vertical levels at thousand foot intervals were evaluated to match those available from IPA-F and FIP/FIS. The grid-based output includes icing probability, icing severity, and super-cooled large droplet (SLD) potential fields.

3.1.2 IPA-F SHORT-TERM FORECASTS

The IPA-F algorithm is the forecast counterpart to IPA-D with similar outputs. The output from IPA-F includes calibrated icing probability, icing severity, and potential for super-cooled large droplets (SLD). The implementation of IPA-F is similar to the logic behind the FIP, which is documented in McDonough et al. (2003), Brown and Bernstein (2006), and Wolff et al. (2008). IPA-F is produced hourly from 13-km RAP output that is projected onto a 11.25-km horizontal grid. Since the IPA-F does not produce a zero-hour forecast, the two- and three-hour IPA-F forecasts were used for comparison to IPA-D. Accounting for the latency of model runs, comparing the three-hour IPA-F forecast to the IPA-D will mean the same underlying RAP grid was input into both algorithms.

3.1.3 FIP AND FIS

The FIP and FIS are currently run over the Alaska region by the AAWU. The FIP and FIS algorithms are applied to the 40-km North American Mesoscale (NAM) model, from which 6-km gridded forecasts are produced. Forecasts are made in three-hour intervals, from zero through 24 hours. While there is no explicit diagnostic output from the FIP or FIS, the zero-hour forecast is the product's initialization of the icing environment and could be used as a situational awareness tool. Therefore, the zero-hour forecast was used as a reference product for comparison in this verification study.

The FIP and FIS output an uncalibrated value between zero and one for icing potential and severity, respectively. The FIS values are made comparable to IPA output using thresholds from the IFI PDT. FIP potentials are converted using the following equation from the IFI PDT (Kucera et al. 2007):

$$\text{probability} = \text{potential} * (-0.033*\text{fcst\_lead\_hours}+0.84) \quad (1)$$

A comparison of the three situational awareness products follows in Table 2 below.

TABLE 2. COMPARISON OF THE CHARACTERISTICS OF IPA_D, IPA_F, AND FIP/FIS.

	IPA-D	IPA-F	FIP/FIS
Product Frequency	Every hour	Every hour	0000, 0600, 1200, and 1800 UTC
Probability Fields	Probability of Icing	Probability of Icing	Icing Potential (FIP) {10, 30, 50, 70, 90}; converted to probability; adjusted for lead time per equation 1
Severity Fields	Icing Severity • None/Trace • Light • Moderate • Severe	Icing Severity • None/Trace • Light • Moderate • Severe	Icing Severity (FIS) PDT/AAWU • 0.000 = None/Trace • 0.175 = Light • 0.375 = Moderate • 0.700 = Severe
Super-cooled Large Drops	Present	Present	Not present
Forecast Time Used for Comparison	N/A	2 and 3-hour	0-hour
Horizontal Resolution	11.25 km	11.25 km	6 km (from 40 km)
Altitudes	500–30,000 ft	1,000–30,000 ft	1,000–30,000 ft
Vertical Resolution	500-ft increments	1,000-ft increments	1,000-ft increments at and below 6,000 ft

	used 1,000 ft increments to match IPA-F	1,000-ft increments available for assessment	3,000-ft increments above 6,000 ft

3.2 OBSERVATIONS

3.2.1 PIREPS

PIREPs are observations of meteorological variables by pilots that can include the existence or absence of icing and the corresponding type and intensity. The reports are made at irregular intervals at the discretion of the pilot. PIREPs of icing include the location, altitude or range of altitudes, type of aircraft, air temperature, and the type and intensity of icing (NWS 2007). All PIREPs of icing in the Alaska domain, both positive and negative observations, were used in this study, leveraging the usual thresholds listed below. PIREPs not containing reported icing conditions have in some cases been used as implicit observations of no icing. This approach was not employed for this assessment – PIREPs excluding the icing condition were not used in this assessment. In addition to light or greater (LOG) icing cases, special attention will be paid to reports of moderate or greater (MOG) icing.

Icing intensity

1. Trace: Ice becomes perceptible. The rate of accumulation is slightly greater than sublimation. Deicing/anti-icing equipment is not utilized unless encountered for an extended period of time (more than 1 hour).
2. Light: The rate of accumulation may create a problem if flight is prolonged in this environment (more than 1 hour). Occasional use of deicing/anti-icing equipment removes/prevents accumulation. It does not present a problem if deicing/anti-icing is used.
3. Moderate: The rate of accumulation is such that even short encounters become potentially hazardous, and use of deicing/anti-icing equipment or diversion is necessary.
4. Severe/Heavy: The rate of accumulation is such that deicing/anti-icing equipment fails to reduce or control the hazard. Immediate diversion is necessary.

All icing types, listed below, were included in the study as well.

Icing types

1. Rime: Rough, milky, opaque ice formed by the instantaneous freezing of small super-cooled water droplets.
2. Clear: A glossy, clear, or translucent ice formed by the relatively slow freezing of large super-cooled water droplets.
3. Mixed: This is a combination of rime and clear.

Figure 2 shows the distribution of PIREPs during the assessment period. There were large quantities of both positive and negative reports of icing throughout the period. Within the IPA domain, PIREPs indicating the absence of icing were relatively more likely over Alaska than in the Pacific Northwest. The geographic distribution of PIREPs over Alaska was fairly even across the six regions.

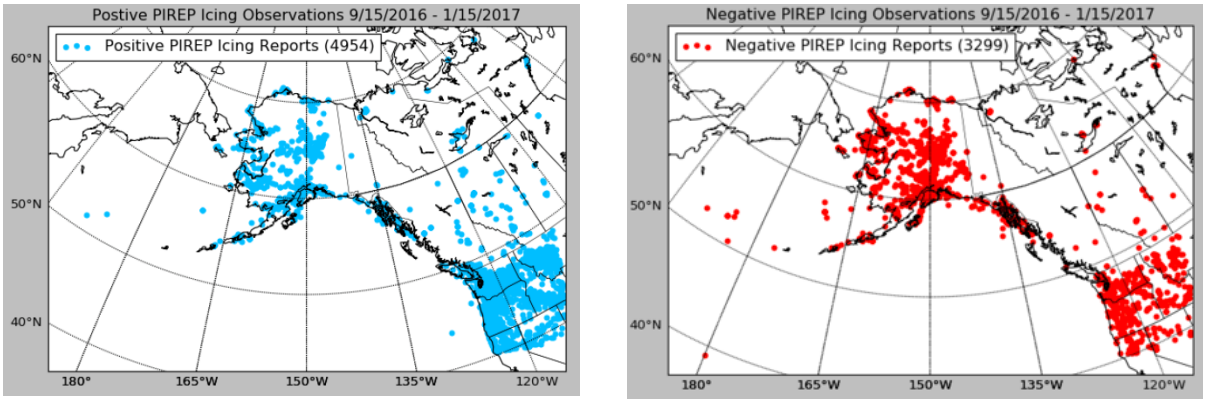

FIGURE 2. GEOGRAPHIC DISTRIBUTION OF POSITIVE (LEFT) AND NEGATIVE (RIGHT) PIREPS OF ICING IN THE IPA DOMAIN DURING THE ASSESSMENT PERIOD.

3.2.2 METARs

METARs are automated and human-produced observations of the meteorological conditions occurring at a surface location. Among the many possible variables that can be reported, an observation of freezing rain or freezing drizzle can be used to indicate the presence of SLD and icing conditions. When freezing rain or freezing drizzle is reported in a METAR, SLD and icing is inferred to exist between the surface and the cloud base, the lowest cloud layer of broken or greater coverage (Madine 2008). These METARs were used to verify SLD in IPA-D and IPA-F.

There are fewer METAR observations over the IPA-D region than over the CONUS. As shown in Figure 3, there are 810 METAR locations in the IPA-D domain compared to 4,147 sites in the CONUS.

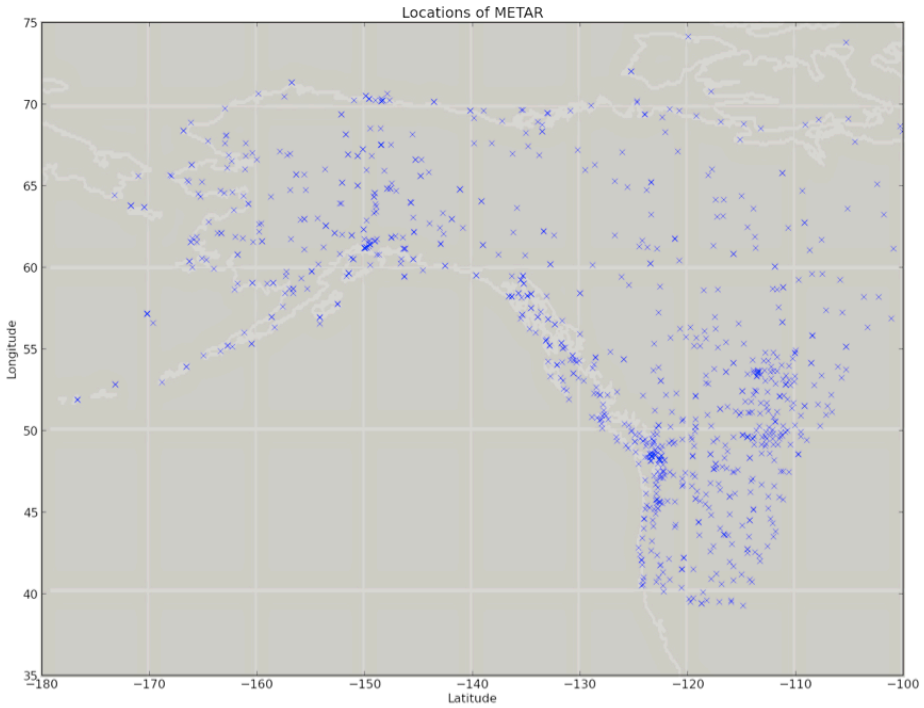

FIGURE 3. LOCATION OF METAR OBSERVATIONS IN THE ALASKA DOMAIN.

3.2.3 RADIOSONDE UPPER AIR OBSERVATIONS

Radiosonde upper air observations were used to supplement the direct measurements of icing conditions. Radiosonde observations include a vertical atmospheric profile of temperature and humidity, among other meteorological parameters, that can be used to indicate conditions that are conducive to icing. The classification criteria used for this study is documented in section 4.2.4.

Within the IPA-D domain there were 37 sites that regularly sent up radiosonde balloons at 12-hour intervals (0000 or 1200 UTC; Figure 4). Although the balloons drift with the wind and collect data over a non-zero time period, all observations used in this study will be attributed to the model grid location closest to the launch location. The choice to use launch location is supported by the results from Seidel et al. (2011), which showed the median balloon drift to be less than 10 km through 500 mb. Most icing observations occur below 500 mb (~18,000 ft) and the grid spacing of IPA-D is 11 km, which means that the majority of radiosonde data applicable to this data will be best attributed to its nearest model grid point.

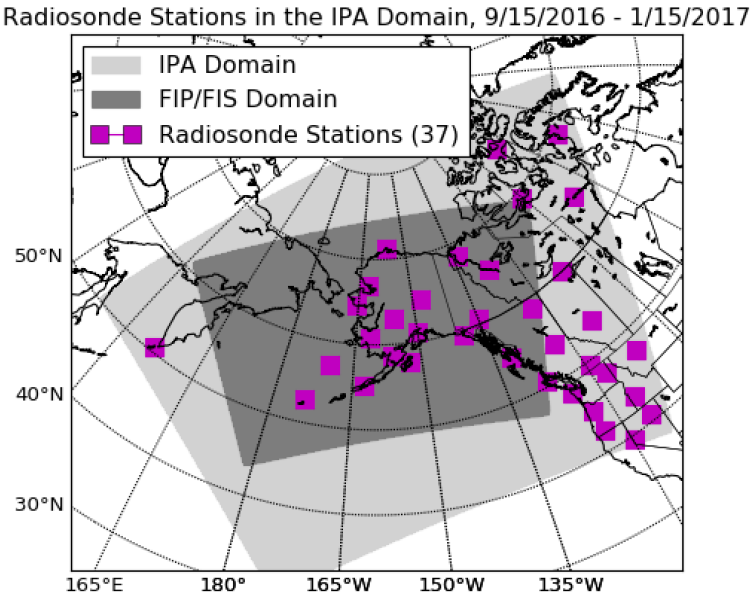

FIGURE 4. LOCATION OF THE RADIOSONDE UPPER AIR OBSERVATION SITES WITHIN THE ALASKA IPA-D DOMAIN.

3.2.4 TAMDAR

TAMDAR is a set of meteorological sensors that are mounted on commercial aircraft and report the in-situ weather conditions experienced by the aircraft. TAMDAR is a proprietary sensor of Panasonic Avionics Corporation and is typically mounted on regional aircraft. Each TAMDAR sensor is capable of recording temperature, pressure, wind speed, wind direction, relative humidity, icing, and turbulence. TAMDAR observations occur at a much higher frequency than PIREPs and there are thousands of observations made world-wide each day.

Of particular importance to this study is the icing sensor that is part of the TAMDAR system. Included in each TAMDAR report is the icing condition and a quality indicator, with the possible values shown in Figure 5 and Figure 6 below. For this assessment, only icing reports with a "good data" quality control flag were included. Additionally, icing reports were treated as a binary condition, with any positive icing value, trace through severe, considered as a yes value for the presence of icing. Only positive icing values were used and TAMDAR observations of no icing were not used in this assessment. The TAMDAR observations of no icing were not used because Panasonic's updated quality control flag, including an improved algorithm for heater activation, were not implemented until April 2017.

Icing Intensities

value0: No icing
value1: Light icing
value2: Light icing in Cloud *
value3: Light icing in Precipitation *
value4: Moderate icing
value5: Moderate icing in cloud *
value6: Moderate icing in precipitation *
value7: Severe icing *
value8: Severe icing in cloud *
value9: Severe icing in precipitation *
value10: Trace of icing *
value11: Trace of icing in cloud *
value12: Trace of icing in precipitation *
value13: Reserved
value14: Reserved-1
value15: Missing value

FIGURE 5. ICING INTENSITIES REPORTED BY TAMDAR. THE "NO ICING" REPORTS (VALUE0) WERE NOT USED IN THIS ASSESSMENT. *: VALUES THAT ARE NOT TYPICALLY OBSERVED IN THE TAMDAR DATA.

Quality Flags

value0: Good Data
value1: Slightly Suspect
value2: Highly Suspect
value3: Bad Data

FIGURE 6. ICING REPORT QUALITY INDICATORS FOR TAMDAR OBSERVATIONS.

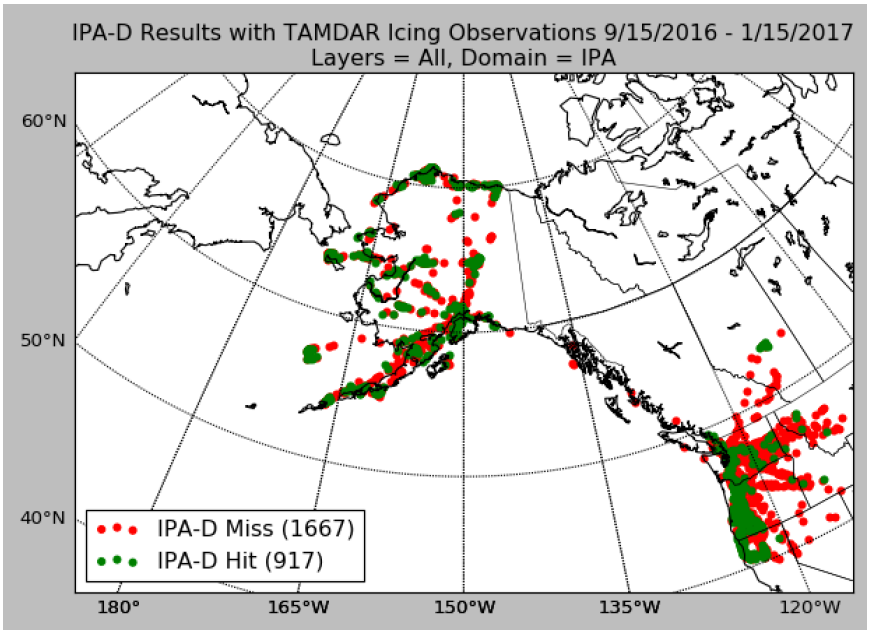

FIGURE 7. MAPS SHOWING THE LOCATION AND NUMBER OF TAMDAR REPORTS OVER THE ALASKA DOMAIN DURING THE ASSESSMENT PERIOD.

3.2.5 SATELLITE

Data products from the satellites CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) were combined with a RAP temperature field to specify the boundaries for which icing conditions exist. CloudSat and CALIPSO are polar-orbiting satellites within the A-Train constellation. Each flies in a sun-synchronous orbit that is 705 km above Earth's surface. CloudSat and CALIPSO maintain a close formation, providing near-simultaneous and collocated observations with the instruments on

these two platforms. The ground track repeats every 233 orbital revolutions, or every 16 days (Stephens et al. 2002).

CloudSat carries a Cloud Profiling Radar (CPR), which sends out a series of short pulses at a 94-GHz frequency, providing a very detailed view of clouds from space. The CPR on CloudSat offers desirable sensitivity to cloud particles and provides data with an along-track resolution of ~1.7 km, a cross-track resolution of ~1.4 km, and a vertical resolution of 480 m with oversampling every 240 m from the surface up to 30 km. While CloudSat has a history of use in verification studies (Kay et al. 2009), the satellite has had technical issues, and at present, only operates during daylight hours.

The quantitative evaluation of clouds and cloud processes in the global atmosphere, and the evaluation of the relationship between vertical profiles of liquid water and ice are among the primary objectives of the CALIPSO mission. Instruments on this satellite utilize three receiver channels to measure the intensity of a returned lidar signal as a function of distance from the device. Information on the size and type of particles is computed from a ratio of measurements taken at two different wavelengths. The lidar signal is more prone to attenuation (signal loss) than the microwave signal sent from CloudSat (Platt 2011). The vertical resolution of the lidar ranges from 30 m near the surface to 300 m at higher altitudes, with a horizontal resolution of 333 m at the surface (NASA 2010).

4 METHODS

Verification was performed for several stratifications to evaluate IPA-D effectiveness at different thresholds and over different geographic regions. The application of data to evaluate model output is also described below.

4.1 STRATIFICATIONS

Results were stratified according to icing intensity thresholds and spatial considerations.

4.1.1 ICING PROBABILITY AND SEVERITY STRATIFICATIONS

The icing output from IPA-D, IPA-F, and FIP was binned as a binary yes or no icing condition using relevant field thresholds. For the probabilistic and potential fields, the thresholds were based on potential values of 10, 30, 50, 70, and 90, which are commonly used for the FIP thresholds by the AAWU. The thresholds and FIP output were converted to probabilities using equation 1.

The FIS severity values were converted to the categories of trace, light, moderate, and severe according to the IFI PDT thresholds.

SLD output was also converted to a binary condition. Values of 0.05 and greater, as well as unknown values, were considered indicative of the presence of SLD in the product output while values of less than 0.05 were considered to specify the absence of SLD.

4.1.2 ALTITUDE BINS

Results were aggregated into the following altitude ranges:

Stratification	**Altitudes**
Near Surface	500 – 6,000 ft
Low	6,500 – 12,000 ft
Middle	12,500 – 18,000 ft
High	18,500 – 30,000 ft

4.1.3 SEVERITY STRATIFICATION

Two icing severity conditions were evaluated in this study, the presence of light or greater (LOG) icing and the presence of MOG icing. The presence of any icing was defined as LOG icing while the MOG condition was restricted to moderate and severe icing. Table 3 below shows the mapping between different products and observation sets to icing severities.

TABLE 3. COMPARISON OF THE ICING CATEGORIES ACROSS PRODUCTS AND DATA PLATFORMS. *: VALUES NOT TYPICALLY OBSERVED IN TAMDAR DATA.

IPA-D Category	IPA-F Category	FIS	PIREP	TAMDAR
None/Trace	None/Trace	0.000–0.174	Neg Neg-Clr Trace	value 0: No icing value 10: Trace of icing * value 11: Trace of icing in cloud * value 12: Trace of icing in precipitation *
Light	Light	0.175–0.374	Trace-Light Light	value 1: Light icing value 2: Light icing in cloud * value 3: Light icing in precipitation *
Moderate	Moderate	0.375–0.699	Light-Mod Moderate	value 4: Moderate icing value 5: Moderate icing in cloud * value 6: Moderate icing in precipitation *
Severe	Severe	0.700–1.000	Mod-Severe Severe	value 7: Severe icing * value 8: Severe icing in cloud * value 9: Severe icing in precipitation *

4.1.4 GEOGRAPHIC STRATIFICATION

The results of the verification study were aggregated across the IPA domain, the FIP/FIS domain, and segregated into regions. The regional definition from McDonough et al. (2010) was used, shown in Figure 8. The six regions are North Slope, Kotzebue and Norton Sound, Interior, Southwest and Bering, South Central, and Southeast and Gulf Coast.

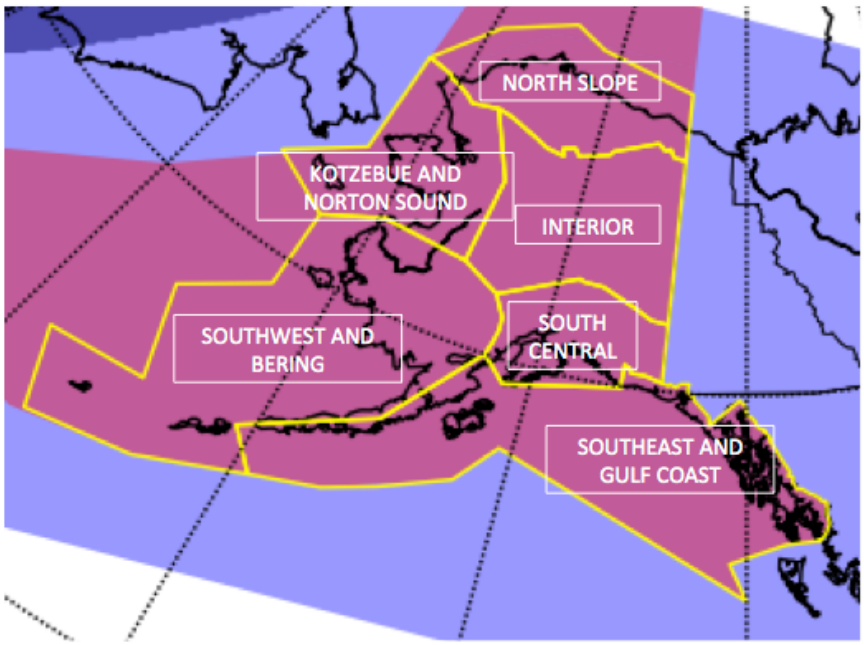

FIGURE 8. THE REGIONAL DIVISION OF THE ALASKA DOMAIN.

4.2 Techniques to Pair Observations and Product Output

The grid-based technique was used to match observations to product output for verification. The grid-based technique matched observation points to either the closest neighboring model grid point or a neighborhood of model grid points. Object-based techniques were used in the case studies and are described in section 5.4.

4.2.1 PIREP-based Techniques

PIREPs were paired to product output using a neighborhood method. Pearson and Sharman (2013) demonstrated that turbulence PIREPs have a median horizontal location error of 35 km and an average vertical error of 20 ft. For the purposes of this study, it was assumed that the errors for icing PIREPs will be similarly distributed. To account for the uncertainty in the horizontal location of PIREPs, a neighborhood approach was used. The neighborhood was composed of all model grid points within a circle with a radius of 35 km, centered at the nearest grid point to the PIREP location. The number of grid points included in neighborhoods constructed in this manner varied because the horizontal grid spacing of the IPA-D and IPA-F grids, based on the NCEP 242 Grid, vary with latitude, ranging from 9.5 km to 12.5 km.

Vertically, PIREPs reporting a single altitude were first matched to a neighborhood containing the closest model vertical level, as well as the two nearest adjacent vertical levels, +/- 1,000 ft. For PIREPs that reported a layer of icing rather than a single altitude, the vertical neighborhood ranged from the closest model level below the PIREP base, to the closest model level above the PIREP top. For a PIREP reporting no icing of unlimited height, the vertical neighborhood ranged from the closest model level below the PIREP base, to the maximum model level.

The horizontal and vertical neighborhood approaches were combined to produce volume neighborhoods matched to each PIREP. The cylinder had a horizontal radius of 35 km and a vertical height of at least 2,000 ft (for PIREPs reporting only a single altitude). Figure 9 below shows the example neighborhoods for the three PIREP cases.

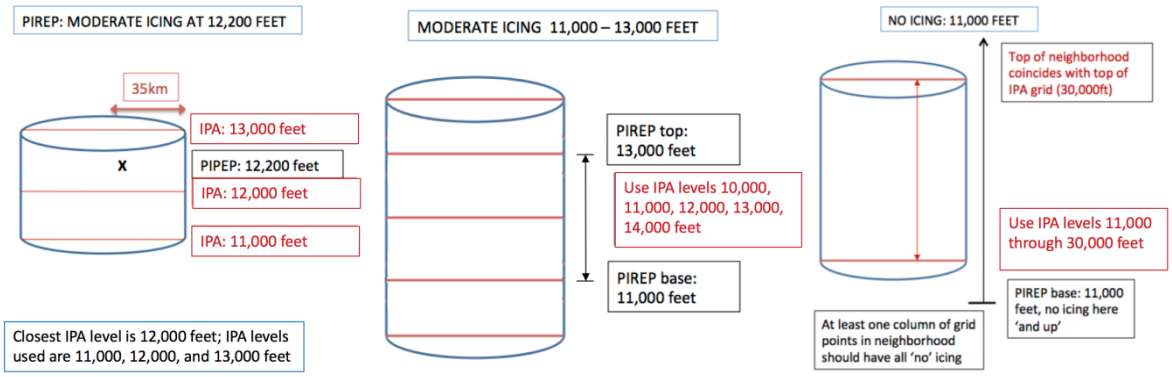

FIGURE 9. THREE POSSIBLE NEIGHBORHOODS SURROUNDING PIREPS. ON THE LEFT IS A NEIGHBORHOOD SURROUNDING A PIREP AT A SINGLE LEVEL. THE MIDDLE DEPICTS THE NEIGHBORHOOD AROUND A PIREP REPORTING A LAYER OF ICING. THE DIAGRAM ON THE RIGHT SHOWS THE CASE OF A PIREP FOR NO ICING ABOVE A GIVEN LEVEL.

From each neighborhood, the minimum, median, and maximum severity was calculated and compared against the icing severity of the PIREP. Hits were recorded for each of the minimum, median, and maximum severities when the PIREP severity was a match. Conversely, if the minimum, median, or maximum severities did not match the PIREP severity, then that category recorded a miss. This evaluation was done for both the LOG and MOG severity cases.

For temporal matching, all PIREPs within a time window of zero through 59 minutes after the diagnostic time were used for verification.

4.2.2 TAMDAR-BASED TECHNIQUES

TAMDAR observations rely upon precise GPS measurements to locate each observation in space and time. The value reported by the icing sensor refers to the value that was present at the sample time, as opposed to an average or peak value. This means that the TAMDAR observations do not suffer from the same location errors that afflict PIREPs. As a result, the TAMDAR observations were matched to the nearest IPA grid point.

Only positive reports of icing from TAMDAR were used in this assessment. Each positive icing report was matched to the nearest algorithm grid point and evaluated against the LOG and MOG threshold. To ensure the quality of the reports, several filtering measures were enacted. First, only observations with a "good quality" flag were accepted. Next, reports were filtered to ensure they were independent and not repetitive. This was done be removing additional reports that contained the same tail number as a previous report and were located within 1 km of horizontal distance, 25 meters of altitude, and within one hour of the first occurrence of that report. Finally, cases were removed from the analysis in which the nearest algorithm grid point was a fill value.

4.2.3 METAR-BASED TECHNIQUES

METARs were used to verify SLD model output. METAR reports of freezing rain and freezing drizzle are indicative of the presence of SLD from the surface upwards to the ceiling location. Figure 10 below diagrams the location where SLD was considered present for a positive METAR report of freezing rain or freezing drizzle. METAR reports of clear skies or snow were used as an indicator of the lack of SLD. METARs

were also matched with the nearest model grid point spatially and zero through 59 minutes after the model diagnostic time temporally.

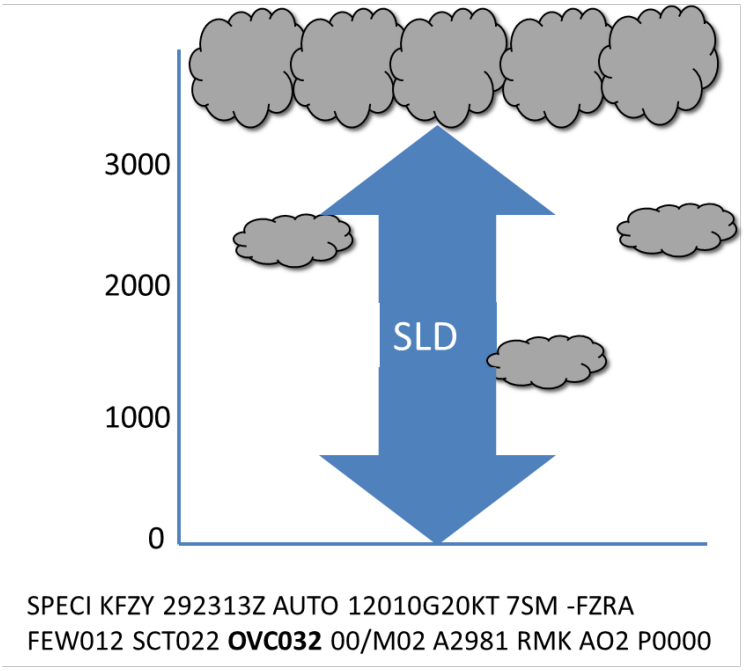

FIGURE 10. DIAGRAM SHOWING THE VERTICAL EXTENT ATTRIBUTED TO A METAR CONTAINING A REPORT OF FREEZING RAIN OR FREEZING DRIZZLE.

4.2.4 RADIOSONDE-BASED TECHNIQUES

Radiosonde observations were used to create a binary condition for whether or not icing is possible through a given air column. There are several accepted temperature ranges and humidity levels for which icing is considered possible. In the classification criteria of Schultz and Politovich (1992), icing is considered possible for layers of an atmospheric profile in which the temperature is between 0°C and -20°C and relative humidity is greater than 67%. This will be referred to as "Class 1" criteria and layers that are outside of those bounds are considered areas in which icing is highly unlikely. According to Bernstein et al. (2007), more stringent criteria are relative humidity values greater than 87% and temperatures between -2°C and -15°C, referred to as "Class 2" criteria.

This study utilized both Class 1 and Class 2 criteria. Model output from the horizontal grid point closest to the radiosonde location was matched to the radiosonde temperature and humidity data. Each grid point along the vertical profile was regarded as a hit (miss) if the product diagnosed (did not diagnose) icing in an area meeting Class 2 criteria. Likewise, a false alarm (correct negative) was recorded if the product diagnosed (did not diagnose) icing in an area within (outside of) Class 1 criteria. A sample of a radiosonde with a potential icing area between 5,000 and 17,000-ft is shown below in Figure 11.

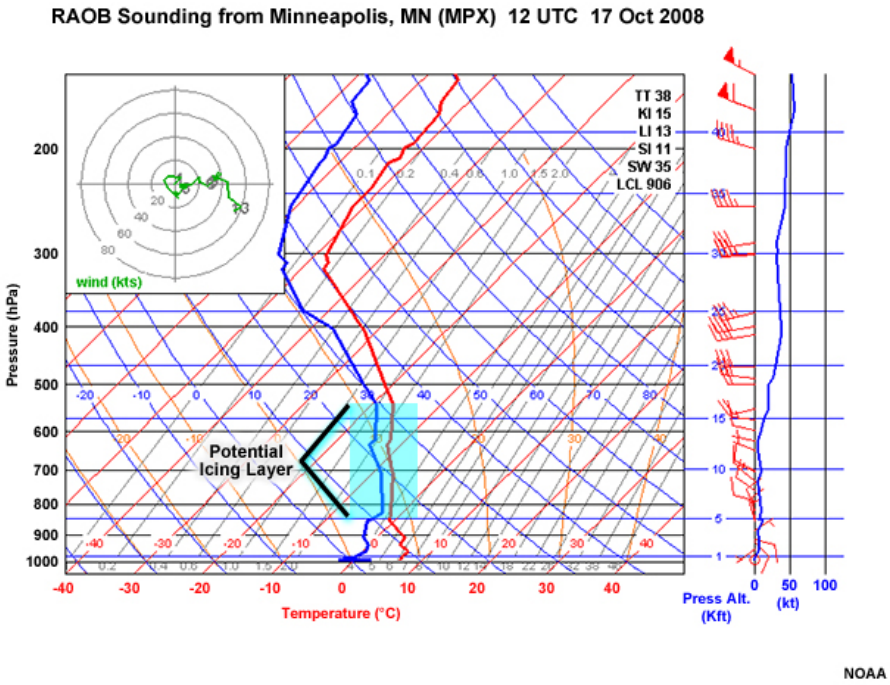

FIGURE 11. EXAMPLE OF A RADIOSONDE WITH A POTENTIAL ICING LAYER. THE LEVELS WITHIN THE BLUE BOX MEET THE CLASS 1 CRITERIA FOR POSSIBLE ICING.

4.2.5 SATELLITE-BASED TECHNIQUES

CALIPSO and CloudSat observations were used to bound regions where clouds were present and identify the cloud type. CALIPSO and CloudSat provided diagnostics of cloud horizontal extents as well as cloud bases and tops. This cloud classification was then combined with the temperature field from the RAP analysis to create a region within which the conditions for icing are possible.

The temperature analysis from the RAP was mapped to the IPA grids using a nearest-neighbor interpolation, resulting in a RAP temperature analysis with the same resolution and profile as the CloudSat and CALIPSO data. Using the cloud classification and temperature thresholds conducive to icing, the temperature and satellite analyses were combined to identify areas of potential icing conditions. The combinations of cloud type and RAP temperature that were used to identify areas of icing are outlined in Table 4. Figure 12 shows an example of the identification scheme.

TABLE 4. COMBINATIONS OF CLOUD TYPE FROM CALIPSO AND CLOUDSAT WITH TEMPERATURES FROM THE RAP THAT PRODUCE POSSIBLE ICING CONDITIONS. THE CLOUD TYPES ARE AS FOLLOWS: SC = STRATOCUMULUS, AC = ALTOCUMULUS, AS = ALTOSTRATUS, CU = CUMULUS, NS = NIMBOSTRATUS, CI = CIRRUS, CS = CIRROSTRATUS, CC = CIRROCUMULUS.

Cloud Types	Temperature range for icing
Sc, St	0°C to -10°C
Ac, As	0°C to -20°C
Cu, Ns, deep convective	0°C to -25°C
Ci, Cs, Cc	no icing

Two different approaches were used for combining the cloud and temperature information:

- Strict: Cloud type and temperature range by type (Note: Attenuated regions of CALIPSO not included)
- Relaxed: Any cloud type and full temperature range (0° to -25°C) (Note: Attenuated regions of CALIPSO are included)

Hits (misses) were recorded when the product identified (did not identify) icing within the strict criteria. False alarms (correct negatives) were recorded when the product identified (did not identify) icing outside of the relaxed criteria.

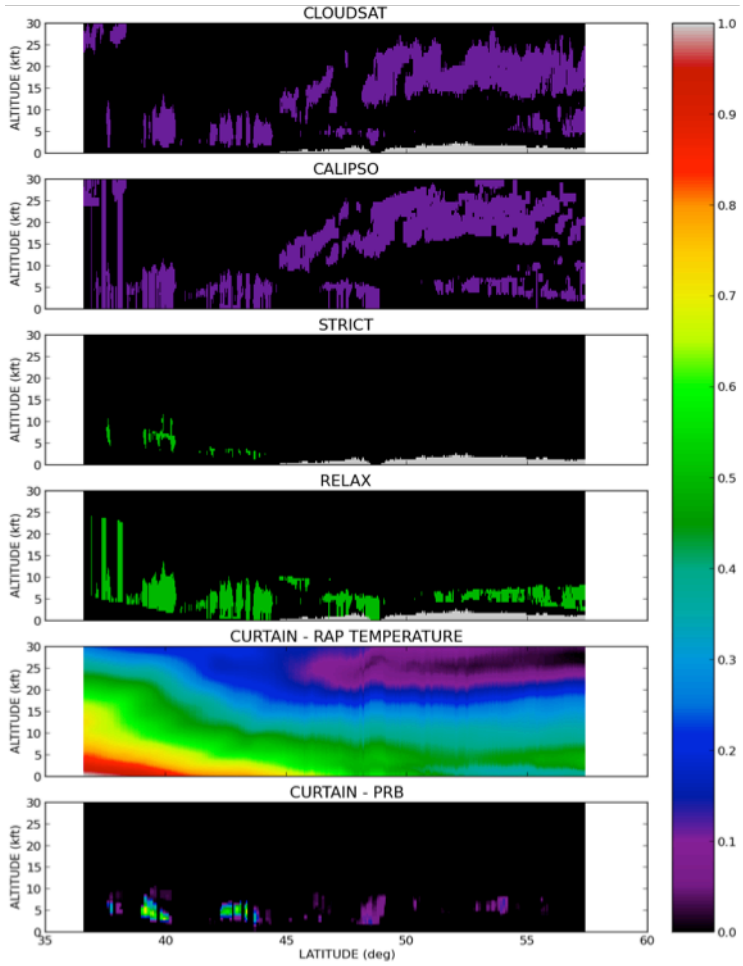

FIGURE 12. EXAMPLE OF POSSIBLE ICING REGIONS OBTAINED FROM COMBINING CLOUDSAT, CALIPSO, AND THE RAP TEMPERATURE FIELD. FROM TOP TO BOTTOM ARE THE CLOUDSAT CLOUD FIELD, CALIPSO CLOUD FIELD, STRICT COMBINATION, RELAXED COMBINATION, RAP TEMPERATURE FIELD, AND ICING PROBABILITY FIELD.

The CALIPSO and CloudSat observations and RAP temperature output were interpolated onto the IPA grid in order to make comparisons. Temporal matching was achieved by matching the IPA-D output to the satellite data that occurs zero through 59 minutes after the valid time of the algorithm.

5 EVALUATIONS

Terminology and score definitions are first provided for reference in the subsequent sections:

MOG:	Moderate-or-Greater Icing
LOG:	Light-or-Greater Icing
POD (= PODy):	Proportion of all observed events that are correctly forecast to occur, in this case, of detecting icing at or above a specific threshold
POFD (= 1 – PODn):	Proportion of all observed non-events that are mistakenly forecast to be events, in this case, detecting icing less than the specified threshold
PSS:	Pierce Skill Score, aka True Skill Score (POD – POFD)

The following table summarizes the performance evaluations that were completed in the performance evaluation section of this assessment and the products and observations that were used. IPA-D was the product evaluated in this assessment and the IPA-F 2-hr and FIP/FIS 0-hr products were the baseline products. The three products were evaluated for field distributions, climatological maps, POD, POFD, and PSS, where the specific metrics used varied according to observation set.

TABLE 5. LIST OF THE PERFORMANCE EVALUATIONS COMPLETED IN THIS ASSESSMENT, INCLUDING THE PRODUCTS AND OBSERVATIONS THAT WERE USED.

Performance Evaluations
Products Evaluated
IPA-D
IPA-F 2-hr
FIP/FIS 0-hr
Field Characteristics
Field Distributions
Climatological Maps
POD/POFD
PIREPs
TAMDARs (POD only)
METARs
Radiosondes
Satellite
PSS
PIREPs
METARs

5.1 Field Characteristics

5.1.1 Field Distributions

The makeup of the IPA-D fields was evaluated using value-based distributions. Distributions were generated for each field: bins for IPA-D severity were generated per severity category, and the probability and SLD values were binned from 0 to 1 using a bin size of 0.01. IPA-F and FIP/FIS distributions were also generated, all with bin sizes of 0.01 and severity categories depending on the field being considered.

5.1.2 Climatological Maps

Further characteristics of IPA-D were evaluated using climatology maps, comparing IPA-D to IPA-F and FIP/FIS. Spatial distributions were derived by aggregating counts of IPA-D field values exceeding a threshold (e.g., 50 for potential, MOG for severity) for each horizontal grid point, aggregated over the study time period for all vertical levels. Difference maps were also created by subtracting the IPA-F or FIP/FIS counts from the IPA-D counts.

5.2 Statistics Calculated Using Truth Data Sources

5.2.1 PIREPs

PIREPs were used to determine product skill in detecting IPA-D, IPA-F, and FIP/FIS severity. Due to the non-systematic nature of this dataset, 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, and False Alarm Ratio). The rationale for this is well documented by Brown and Young (2000) and Carriere et al. (1997).

The association of the algorithms to PIREPs as described previously yields the following contingency table:

Hit: forecast = yes; obs = yes

False alarm: forecast = yes; obs = no

Miss: forecast = no; obs = yes

Correct no: forecast = no; obs = no

'Yes' signifies that the forecast or observation equaled or exceeded a given threshold, and 'no' signifies that the forecast or observed value was less than the threshold. POD, POFD, and PSS were then computed from the contingency table.

In addition, error bars were added to the PIREP skill score plots. The error for each statistic (POD, POFD, PSS) was generated using a bootstrap method. The bootstrap method employed for this assessment sampled the contingency table from each day in the assessment period 1,000 times with replacement. The resulting distributions were normally shaped. The 95%-confidence level was calculated by taking the values at the 2.5% and 97.5% thresholds within the distribution.

5.2.2 METARs

POD, POFD, and PSS were calculated for IPA-D and IPA-F SLD using METAR observations of freezing rain and freezing drizzle. SLD potential in the model field >= 0.05 and unknown values were defined as a positive specification of SLD while values < 0.05 were treated as negative specification of SLD.

5.2.3 Radiosonde

The parameters of Class 1 and Class 2 criteria provide an upper and lower bound for regions where icing is possible. POD and POFD were calculated for each criterion against the IPA-D, IPA-F, and FIP outputs. Because the binary categorization of icing is not mutually exclusive and collectively exhaustive, they were not combined to compute PSS.

5.2.4 Satellite Based

The satellite observations of cloud coverage represent the maximum extent over which icing should be possible. The combination of the satellite observations and the RAP temperate field were compared against the IPA-D, IPA-F, and FIP/FIS identifications of icing. POD and POFD were computed for each time and location for which the data is available. For reasons similar to radiosonde, PSS was not computed.

5.3 Consistency

The consistency of the IPA-D output was also assessed and calculated in two ways. First, IPA-D was compared to IPA-D output at adjacent time steps. The top row of boxes in Figure 13 provides a visualization of the comparison. Here, the IPA-D valid at time *t+1* was compared to the IPA-D valid at time *t*.

Second, the IPA-D valid at time *t* was compared to the IPA-F 2-hr valid at time *t+1* and to the IPA-F 3-hr valid time, shown in the bottom row of boxes in Figure 13. Comparing the consistency of the IPA-F 2 and 3-hr products to the consistency of the adjacent IPA-D products will allow a quality judgement to be made regarding the consistency of IPA-D.

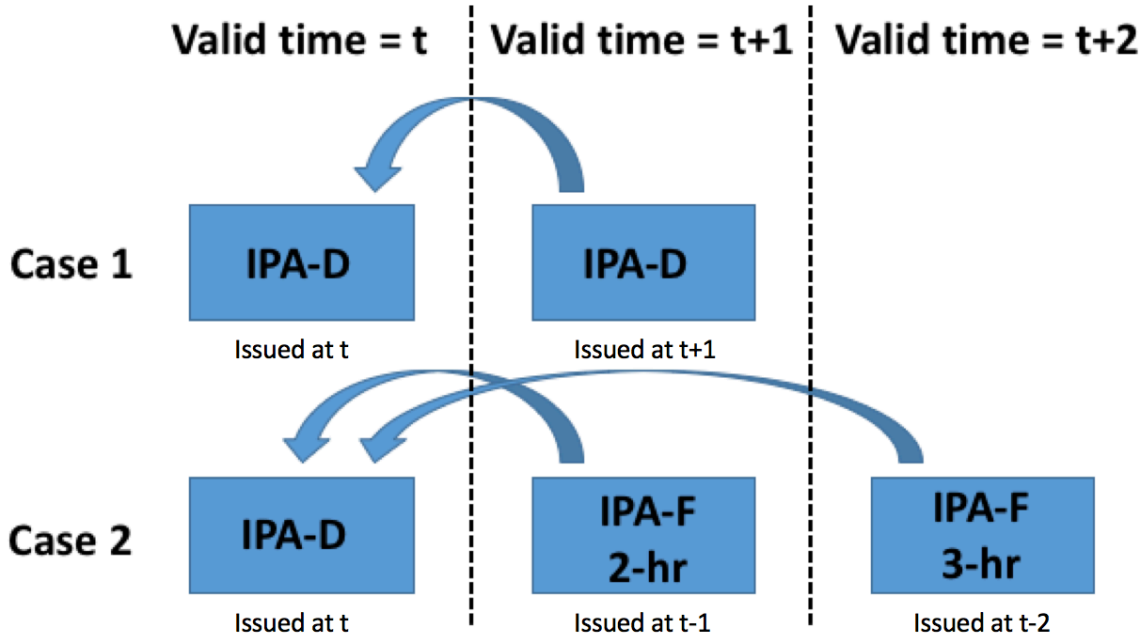

FIGURE 13. VISUALIZATION OF THE TWO CONSISTENCY CALCULATIONS. CASE 1 (TOP ROW): THE IPA-D AT TIME T+1 IS COMPARED TO THE IPA-D AT TIME T. CASE 2 (BOTTOM ROW): THE IPA-F 2-HR AND THE IPA-F 3-HR, VALID AT TIME T+1 AND T+2, RESPECTIVELY, ARE COMPARED TO THE IPA-D AT TIME T.

5.4 Case Studies

Case studies for specific events were also part of the analysis. The dates of these events were coordinated with the AAWU. From the set of interesting icing dates provided by the AAWU, three cases were selected based upon the number of MOG icing PIREPs and the temporal and spatial proximity of the PIREPs.

6 5 RESULTS

The results from the four-month comparison of IPA-D, IPA-F, and FIP/FIS are detailed in the subsequent sections. Performance metrics using the PIREP, TAMDAR, radiosonde, satellite, and METAR data sets are shared, as well as examinations of the algorithm output fields.

6.1 FIELD CHARACTERISTICS

6.1.1 FILL VALUES

Both IPA-D and IPA-F probability fields contained fill values, however, the distribution of the fill values was significantly different between the two algorithms. In IPA-F, the fill values are used to mark regions where the algorithm domain is below the terrain altitude in mean sea level. Figure 14 shows the location of the fill values at four selected vertical levels of IPA-F. It is clear from these images that the fill values in IPA-F act mostly as a terrain mask in the algorithm. The number of fill values in IPA-F shows a slight variation between files at each vertical level, meaning there are some fill values that are not associated with the terrain mask. The fill values in IPA-F composed between 1.8–2.1% of the daily summed IPA-F grid points and up to 35.4% of the grid points in a single vertical layer.

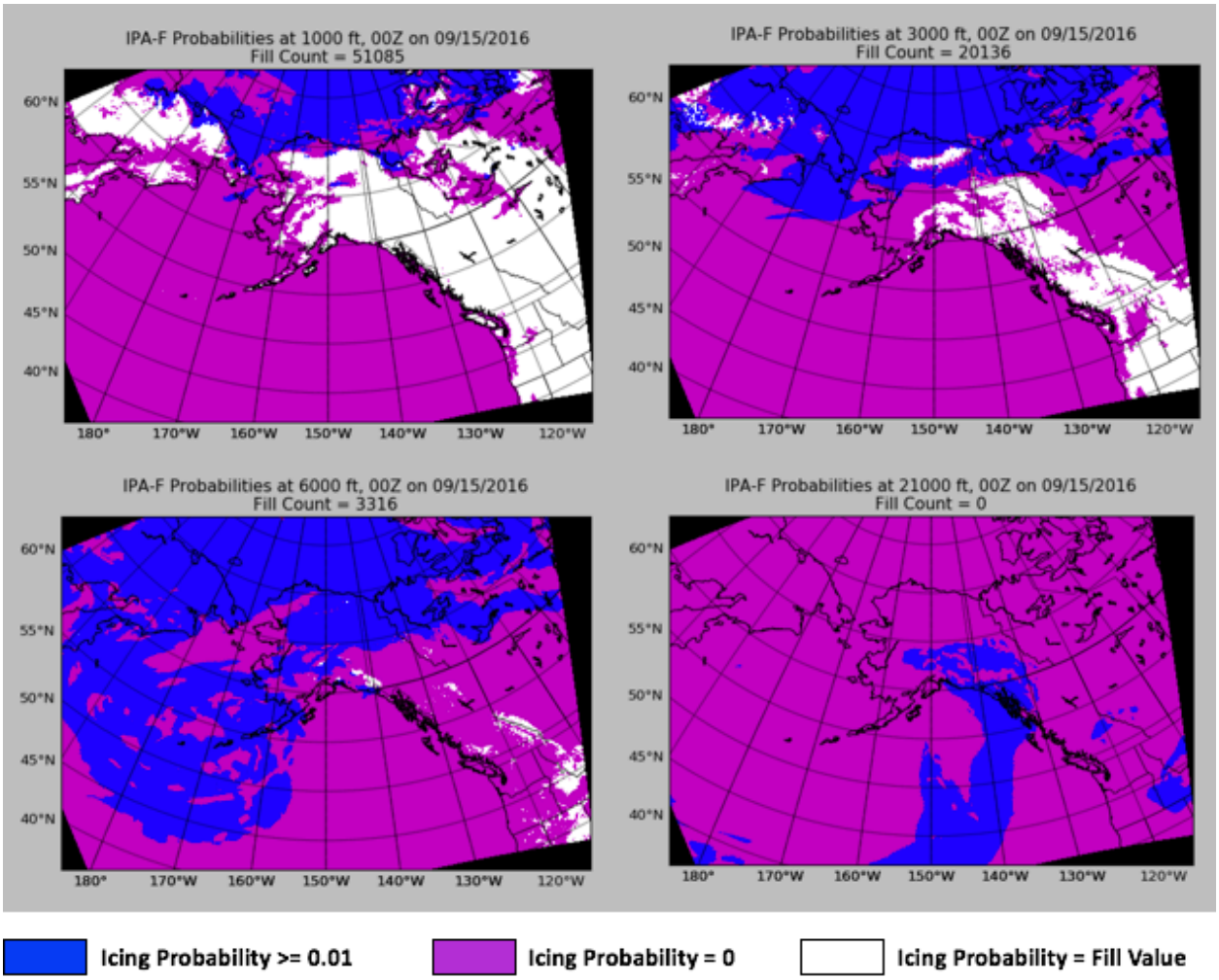

FIGURE 14. FOUR-PANEL IMAGE OF IPA-F 2-HOUR PROBABILITY FIELDS SHOWING THE LOCATION OF FILL VALUES (WHITE) AT 1 K, 3 K, 6K, AND 21 K FEET MSL.

IPA-D contained far more fill values than IPA-F. In addition to the occurrence of fill values in regions where the algorithm grid points are underground, IPA-D also had fill values throughout the grid in regions where the algorithm could not determine an icing value. This resulted in 15–37% of the daily IPA-D grid points being fill values. Figure 15 shows the location of the fill values at four selected vertical levels of IPA-D. In contrast to the IPA-F fill values in Figure 14, the IPA-D fill values occurred over the ocean and land regions at algorithm levels that were above ground level. Figure 15 also highlights the areal extent of the fill values, which varied between a minimum of 2.2% and a maximum of 91.3% of the grid points in a single vertical level. Additionally, the fill values displayed a seasonal variation, with a daily average of 27% of the IPA-D data in September before tailing off to 20% of the data through the colder months of the assessment period. The fill values were most prevalent at lower altitudes and decreased with height until about 15,000 feet, at which point the fill values remained at a constant number for each vertical level up to the algorithm top of 30,000 ft. The severity field of IPA-D did not contain fill values, but interestingly, there were some small areas of overlap where the LOG icing severity was co-located with a fill value in the probability field.

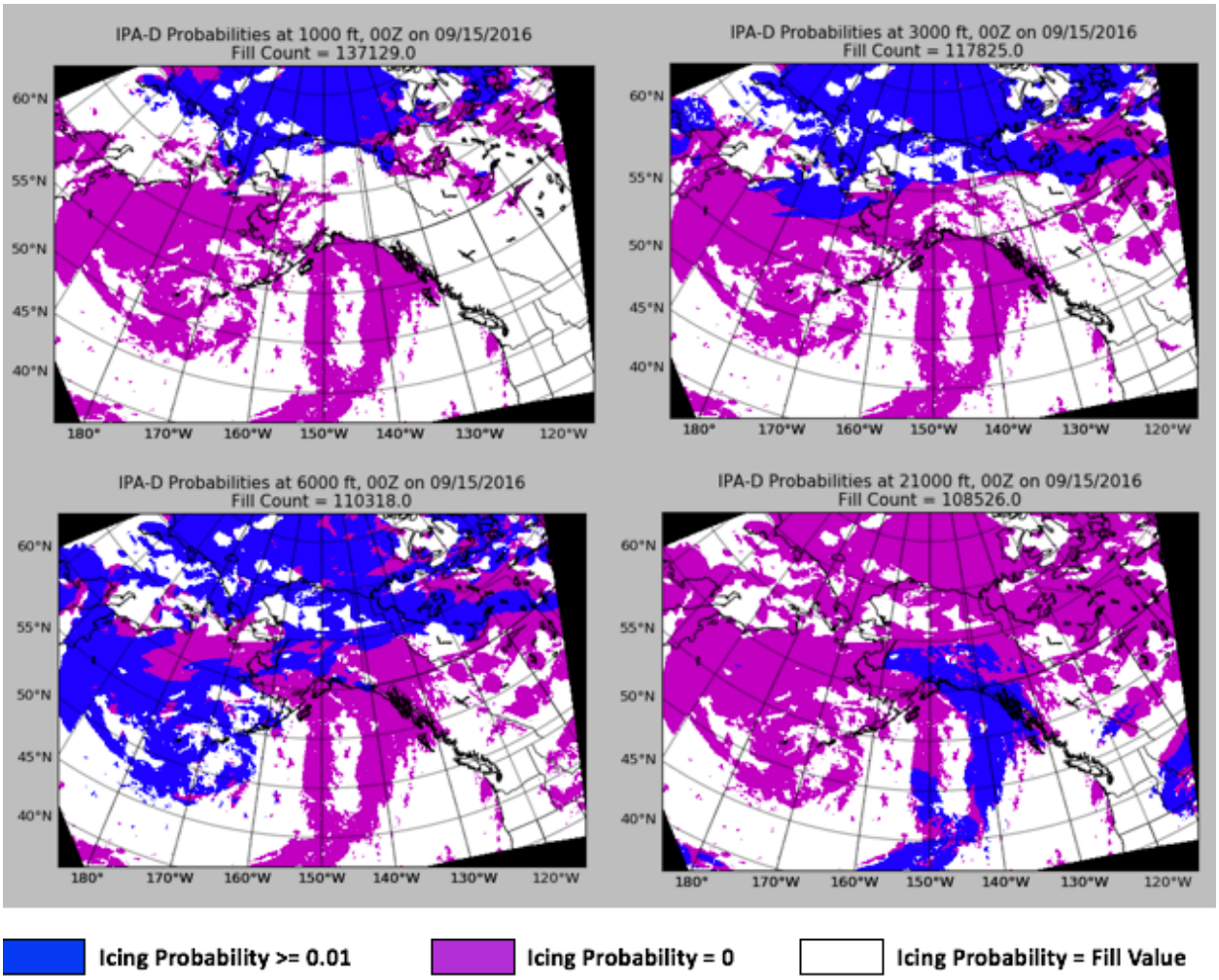

FIGURE 15. FOUR-PANEL IMAGE OF IPA-D PROBABILITY FIELDS SHOWING THE LOCATION OF FILL VALUES (WHITE) AT 1 K, 3 K, 6 K, AND 21 K FEET MSL. THE TIME OF THIS OUTPUT IS COINCIDENT TO THAT OF THE IMAGES FOUND IN FIGURE 14.

6.1.2 Field Distributions

As could be inferred from the fill value section, the prevalence of fill values in IPA-D resulted in lower total counts of valid IPA-D probability values. Therefore, the counts were normalized against the total number of non-fill value counts in order to compare the probability field of IPA-D to IPA-F and FIP. Figure 16 contains the distributions of the probability fields of IPA-D, IPA-F 2-Hour, and FIP. Over the assessment period, IPA-D probability fields only had usable, non-fill values in 80% of the data volume of IPA-F, meaning that 20% of the IPA-D volume could not be used to derive an icing probability for comparison. IPA-D had higher normalized counts near 0.2 and 0.4 icing probability, while IPA-F had higher counts near 0.6 and 0.75 probability. This trend was reversed when comparing IPA-D to FIP, where IPA-D had fewer normalized counts at the lower probabilities and more counts at the higher probability values. IPA-D and IPA-F had similar rates of occurrence of zero values, with 79% of the IPA-D data and 80% of the IPA-F data being zero values. The percentage of IPA-D zero values in the FIP/FIS domain decreased to 78% and was greater than the 72% of FIP data that were zero values.

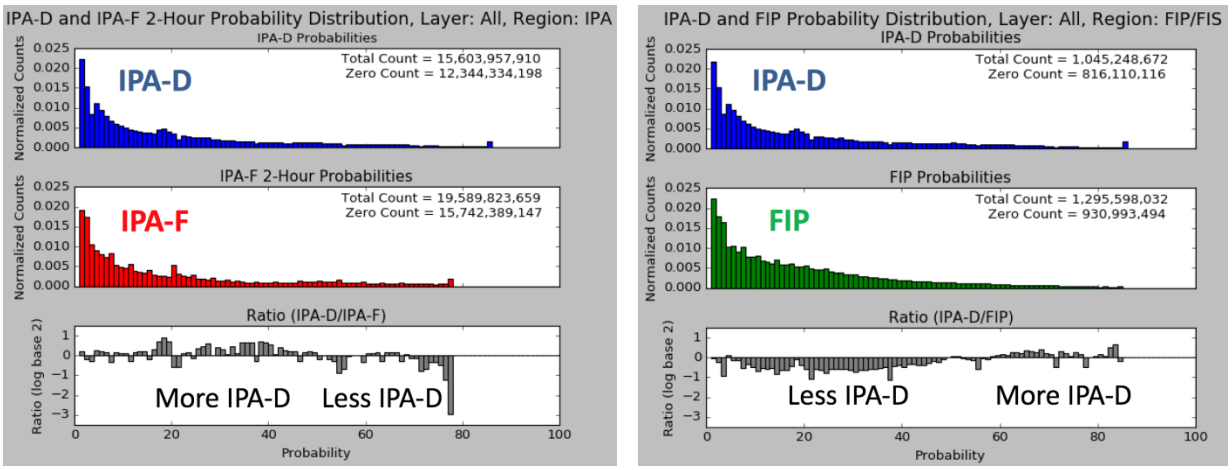

FIGURE 16. DISTRIBUTION OF PROBABILITY VALUES FOR IPA-D (TOP LEFT) AND IPA-F (MIDDLE LEFT) IN THE IPA DOMAIN AND IPA-D (TOP RIGHT) AND FIP (MIDDLE RIGHT) IN THE FIP/FIS DOMAIN. THE BOTTOM PANEL OF EACH GRAPHIC SHOWS THE LOG BASE 2 RATIO OF THE NORMALIZED COUNTS, WITH POSITIVE VALUES INDICATING MORE RELATIVE COUNTS OF IPA-D.

Next, the severity distributions of IPA-D were compared to IPA-F in the IPA domain and FIS in the FIP/FIS domain. While the IPA-D data did not contain fill values, the counts for each algorithm were again normalized to make a fair comparison, the results of which are presented in Figure 17. As expected, the most common condition for each algorithm was the no-icing condition, defined as none or trace icing. The no-icing condition was more frequently found in IPA-D than in both IPA-F and FIS. As icing severity increased from light to moderate to severe, each algorithm saw a decrease in grid points with higher icing severities. Light, moderate, and severe icing conditions occurred less frequently in IPA-D than in IPA-F for the IPA domain. The results were mixed when comparing IPA-D to FIS in the FIP/FIS domain. Here, the light icing condition was less frequent in IPA-D than in FIS, but the moderate and severe icing conditions were more frequent in IPA-D.

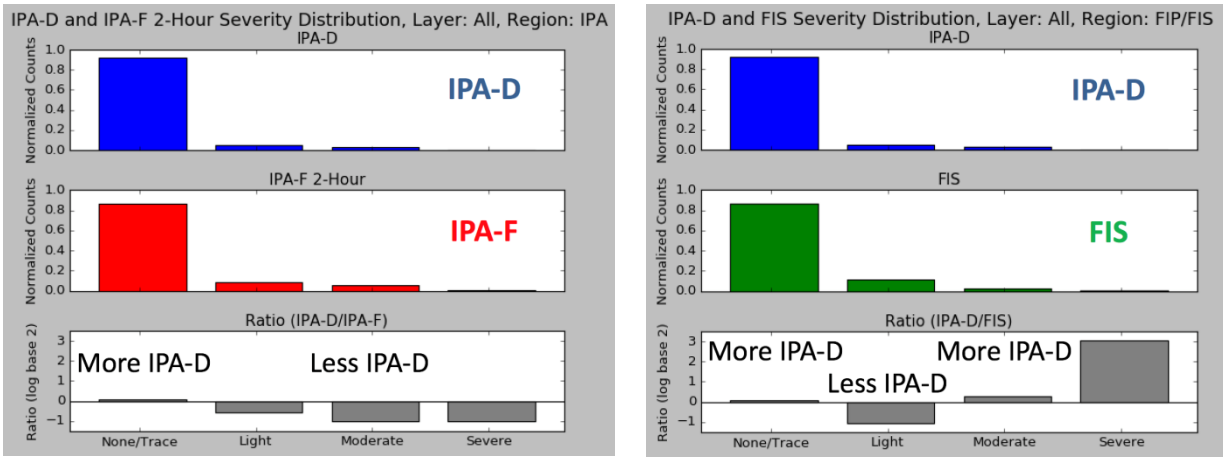

FIGURE 17. DISTRIBUTION OF SEVERITY VALUES FOR IPA-D (TOP LEFT) AND IPA-F (MIDDLE LEFT) IN THE IPA DOMAIN AND IPA-D (TOP RIGHT) AND FIS (MIDDLE RIGHT) IN THE FIP/FIS DOMAIN. THE BOTTOM PANEL OF EACH GRAPHIC SHOWS THE LOG BASE 2 RATIO OF THE NORMALIZED COUNTS, WITH POSITIVE VALUES INDICATION MORE RELATIVE COUNTS OF IPA-D.

Summarizing the field characteristic results, 15–37% of the grid points in the IPA-D probability field were fill values in the data covering the assessment period. The high percentage of fill values in the IPA-D probability field resulted in a lower total count of zero and positive probability values in IPA-D compared to IPA-F and FIP. Overall, IPA-D identified fewer counts of positive probabilities of icing than IPA-F and FIP, resulting in smaller icing volumes in IPA-D compared to the other two algorithms. In addition, IPA-D had a greater number of grid points that were identified in the no-icing condition than both IPA-F and FIS. This resulted in IPA-D diagnosing less light, moderate, and severe icing than IPA-F and less light icing than FIS.

6.2 Performance Metrics

The following sections present the results from comparisons of IPA-D, IPA-F, and FIP/FIS output to the truth data sets from PIREP, TAMDAR, radiosonde, and satellite identifications of icing.

6.2.1 Performance Metrics Using PIREPS

Skill scores using PIREPs as the truth data set were calculated for IPA-D and IPA-F in the IPA domain and IPA-D, IPA-F, and FIP/FIS in the FIP/FIS domain. The IPA-F 2-Hour forecast and the IPA-F 3-Hour forecast exhibited very similar scores, and therefore only the IPA-F 2-Hour forecasts will be presented in order to make the comparisons clearer (hereafter referred to as simply IPA-F).

The comparison of probabilities and severities in section 6.1.2 showed that IPA-D generally had smaller volumes of icing diagnosed than IPA-F and FIP/FIS. Here, the effectiveness of that smaller volume will be evaluated by measuring each algorithm's POD against the identified volume at each severity threshold. Figure 18 illustrates the POD versus volume results for IPA-D, combining the results from the LOG, MOG, and severe PIREP severity levels with the maximum, median, and minimum values from the neighborhood of IPA-D grid points surrounding the PIREP. Using the maximum value from the neighborhood resulted in the highest PODs, significantly greater than the PODs for the median and minimum values at each severity level. Additionally, there is the expected drop-off in POD score as the severity increases from LOG to MOG to severe. Given the substantially greater PODs generated by using the maximum neighborhood severity, the remainder of this assessment will focus on results using the maximum neighborhood value.

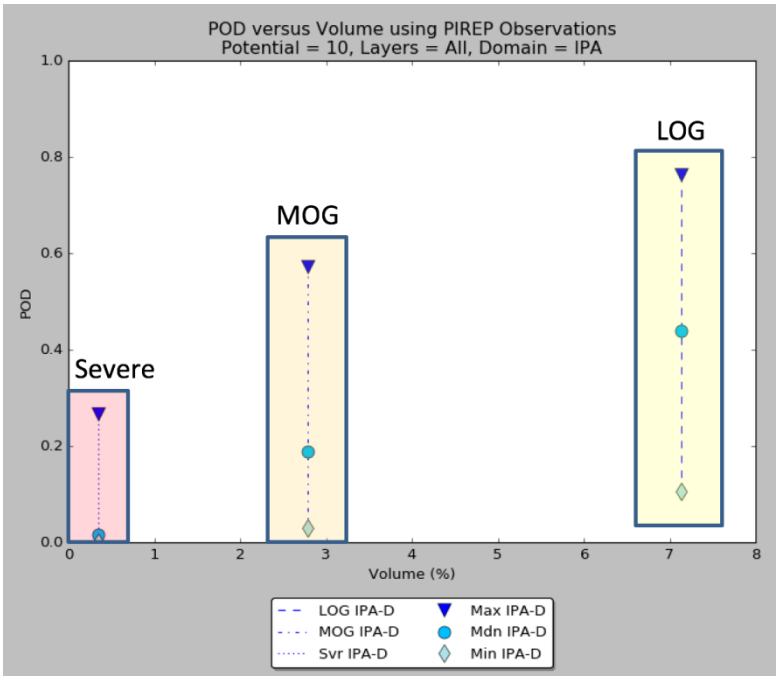

FIGURE 18. POD VERSUS VOLUME PLOT FOR IPA-D. LOG (RIGHT BOX), MOG (MIDDLE BOX), AND SEVERE (LEFT BOX) PIREPS MATCHED TO THE MAXIMUM (TRIANGLE), MEDIAN (CIRCLE), AND MINIMUM (DIAMOND) VALUES FROM THE ALGORITHM NEIGHBORHOOD SURROUNDING THE PIREP ARE DISPLAYED.

Figure 19 compares IPA-D to IPA-F for POD versus volume in the IPA domain. At each severity threshold, IPA-F has a higher POD than IPA-D while also having a larger volume. This created a trade-off between volume and detection, where the additional volume diagnosed at a given severity threshold by IPA-F allowed the detection rate to rise. The differences in POD between IPA-D and IPA-F (IPA-F minus IPA-D) were greater at the MOG (0.18) and severe (0.13) thresholds than at the LOG threshold (0.10), which was consistent with the increasing difference in the relative frequency between IPA-D and IPA-F as severity increases (shown in Figure 17). It is also significant to note that IPA-F was able to achieve a comparable POD at the MOG level (0.75) to IPA-D at the LOG level (0.76), while identifying a smaller volume (5.3% versus 7.1%).

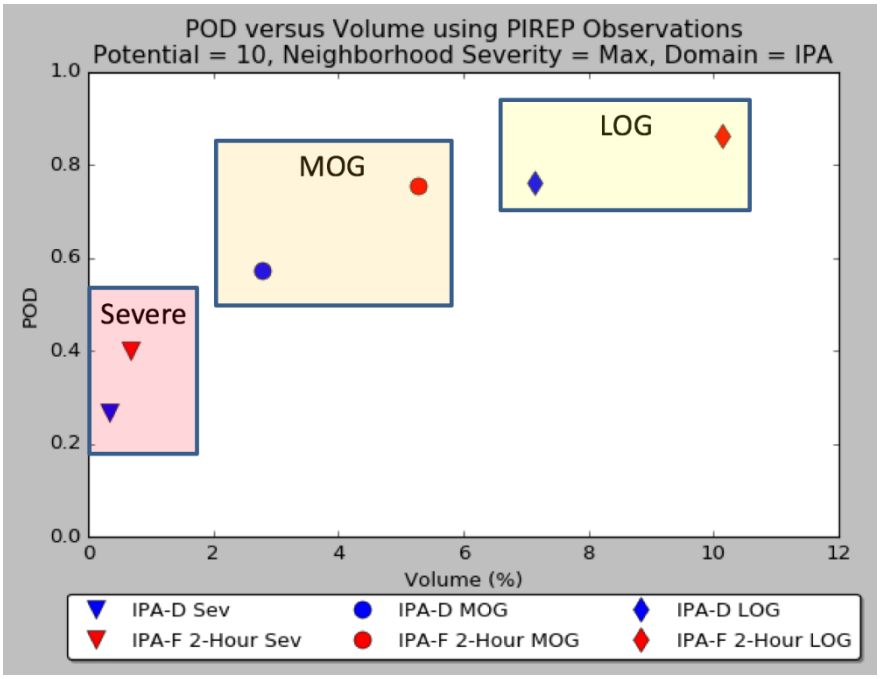

FIGURE 19. POD VERSUS VOLUME PLOT FOR IPA-D (BLUE) AND IPA-F (RED) IN THE IPA DOMAIN. LOG (RIGHT BOX), MOG (MIDDLE BOX), AND SEVERE (LEFT BOX) PIREPS WERE MATCHED AGAINST THE MAXIMUM VALUE FROM THE ALGORITHM NEIGHBORHOOD TO CREATE THE POD.

Similar to the above, Figure 20 displays POD versus volume for IPA-D, IPA-F, and FIP/FIS in the FIP/FIS domain. Only LOG and MOG severities are shown because there were no PIREPs of severe icing in the FIP/FIS region at the four FIP/FIS valid times during the assessment period. Again, the differences in PODs between the three algorithms at the LOG threshold were smaller than differences at the MOG threshold. IPA-D had the smallest volume of the three algorithms at the LOG level, but FIP/FIS had the smallest volume at the MOG level. FIP/FIS's higher POD (0.60 versus 0.54) over a smaller volume (2.1% versus 3.1%) than that of IPA-D makes FIP/FIS the unambiguously better choice between the two algorithms at the MOG threshold.

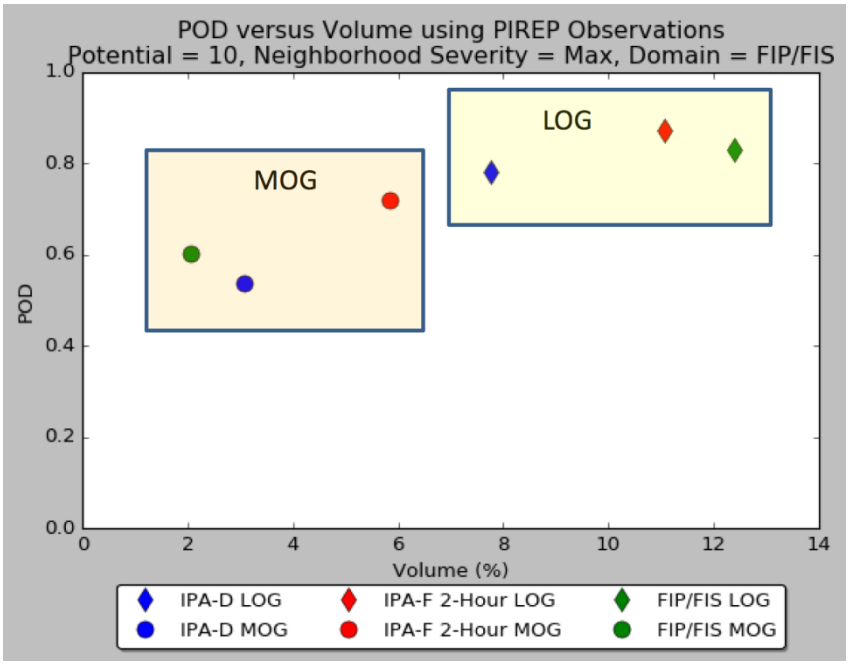

FIGURE 20. POD VERSUS VOLUME PLOT FOR IPA-D (BLUE), IPA-F (RED), AND FIP/FIS (GREEN) IN THE FIP/FIS DOMAIN. LOG (RIGHT BOX) AND MOG (LEFT BOX) PIREPS WERE MATCHED AGAINST THE MAXIMUM VALUE FROM THE ALGORITHM NEIGHBORHOOD TO CREATE THE POD.

POFDs and PSSs were also calculated to add more context to the POD results. The POD, POFD, and PSS for IPA-D and IPA-F in the IPA domain are displayed in Figure 21 for LOG PIREPs and in Figure 22 for MOG PIREPs. In both cases, the maximum severity from the neighborhood of algorithm grid points surrounding the PIREP was selected for comparison to the PIREP severity value. The maximum severity from the neighborhood yielded significantly higher PODs than the median and minimum values and was selected for display in these graphics.

In the LOG case (Figure 21), IPA-F PODs were higher than IPA-D at each potential level (10, 30, 50, 70, 90) with statistically significant differences at the 95% confidence level. IPA-F also had higher POFDs than IPA-D at each potential level. The difference between each algorithm's POD and POFD at each potential level was similar, which resulted in similar PSSs for IPA-D and IPA-F. In fact, IPA-D had slightly higher PSSs than IPA-F at potential thresholds of 10 and 30, although the differences were not statistically significant at any potential threshold for the LOG case. In the MOG case (Figure 22) the results were similar. Again, IPA-F had higher PODs and POFDs at every potential threshold that were statistically significant. The resulting PSSs were also close between IPA-D and IPA-F, although IPA-F had a marginally higher PSS at each potential threshold in the MOG case.

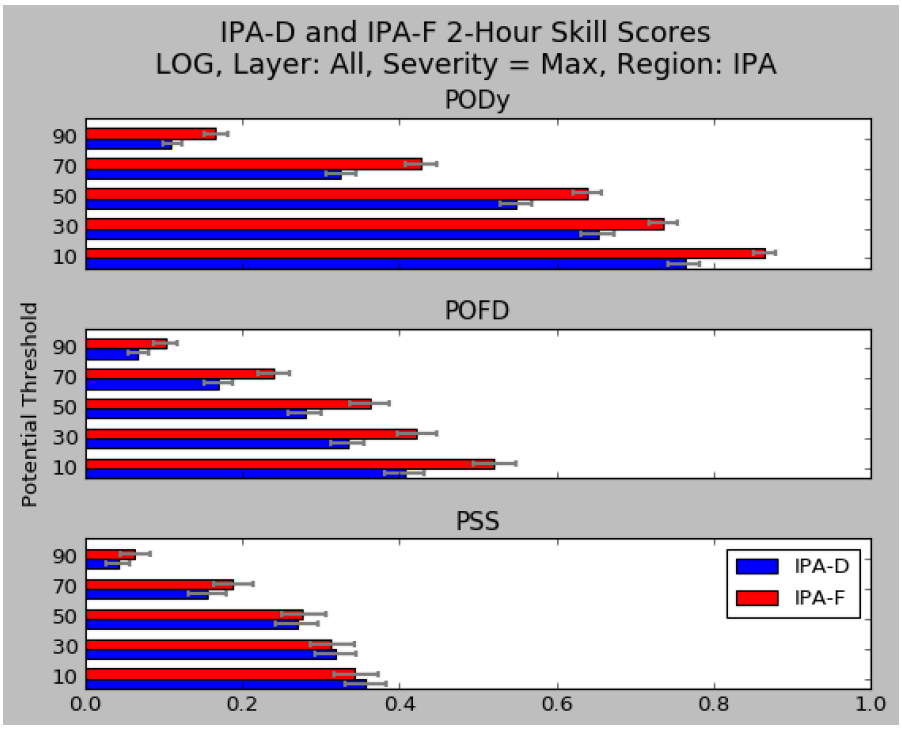

FIGURE 21. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE) AND IPA-F RED) IN THE IPA DOMAIN. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF LIGHT OR GREATER ICING.

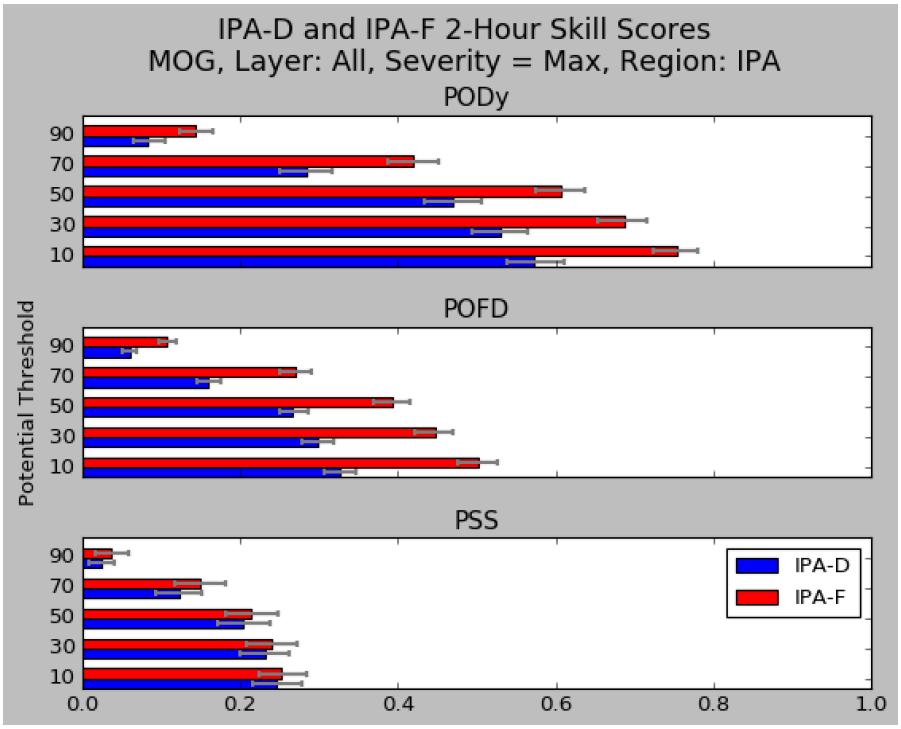

FIGURE 22. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE) AND IPA-F (RED) IN THE IPA DOMAIN. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF MODERATE OR GREATER ICING.

The POD, POFD, and PSS were evaluated for the FIP/FIS overlap region and are presented in Figure 23 for the LOG case and Figure 24 for the MOG case. Similar to the results in the IPA domain for the LOG case, IPA-D has lower PODs at every potential threshold than IPA-F, while also having lower POFDs than IPA-F. IPA-D has lower PODs than FIP/FIS at potential thresholds of 10 and 30, and also has lower POFDs at potential thresholds of 10, 30, and 50. The resulting PSSs for IPA-D are again higher than IPA-F and FIP/FIS at potential thresholds of 50 and below, but the scores are not different at a statistically significant level. In the MOG case at potential thresholds of 10 and 30, IPA-D trails IPA-F and FIP/FIS in POD, but counters with lower POFDs. At the potential thresholds of 50 and 70 in the MOG case, IPA-D has lower PODs and POFDs than IPA-F, and higher PODs and POFDs than FIP/FIS. For MOG PIREPs, IPA-D has lower PSSs than IPA-F and FIP/FIS at potential thresholds of 10 and 30, but higher PSSs at potential thresholds of 50 and 70. Noteworthy is that IPA-D and IPA-F exhibited higher PSSs in the FIP/FIS domain than the IPA domain as a whole, indicating increased accuracy over the Alaska region. The increase in PSS was driven largely by lower POFDs in the FIP/FIS domain.

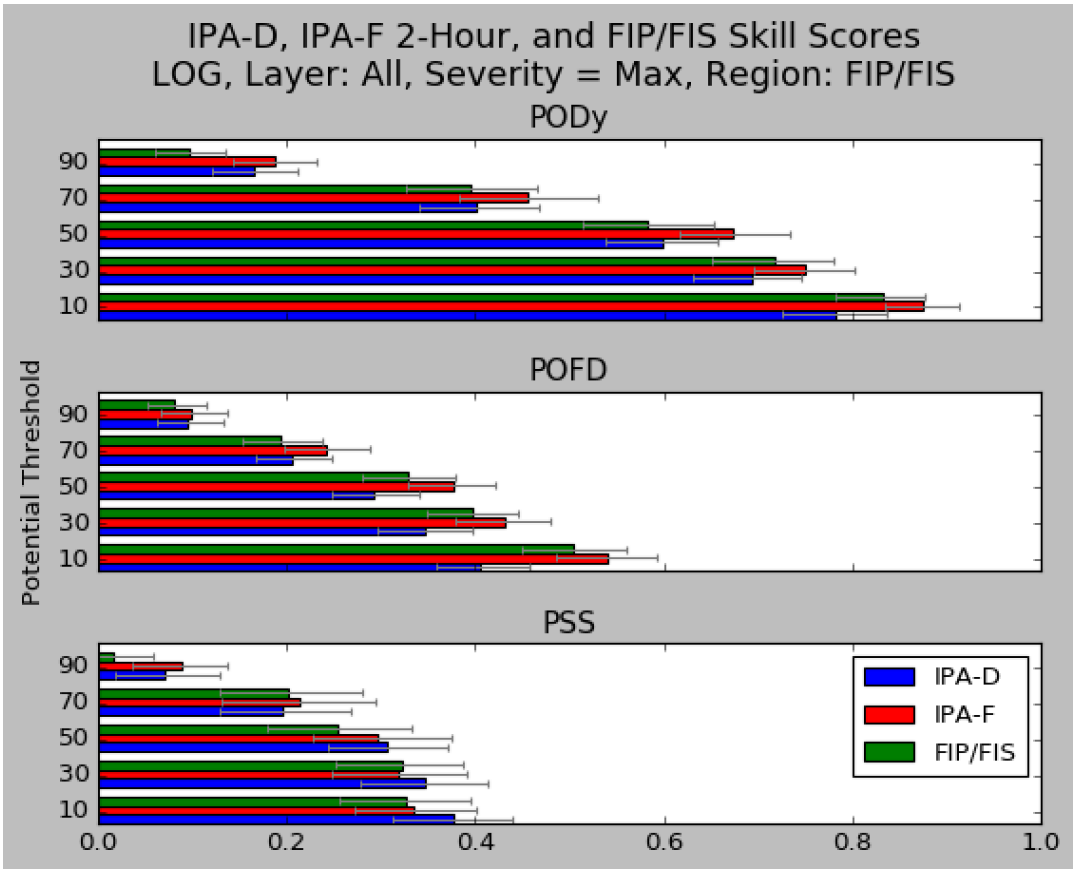

FIGURE 23. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE), IPA-F (RED), AND FIP/FIS (GREEN) IN THE FIP/FIS DOMAIN. ERROR BARS INDICATE THE 95% CONFIDENCE LEVEL. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF LIGHT OR GREATER ICING.

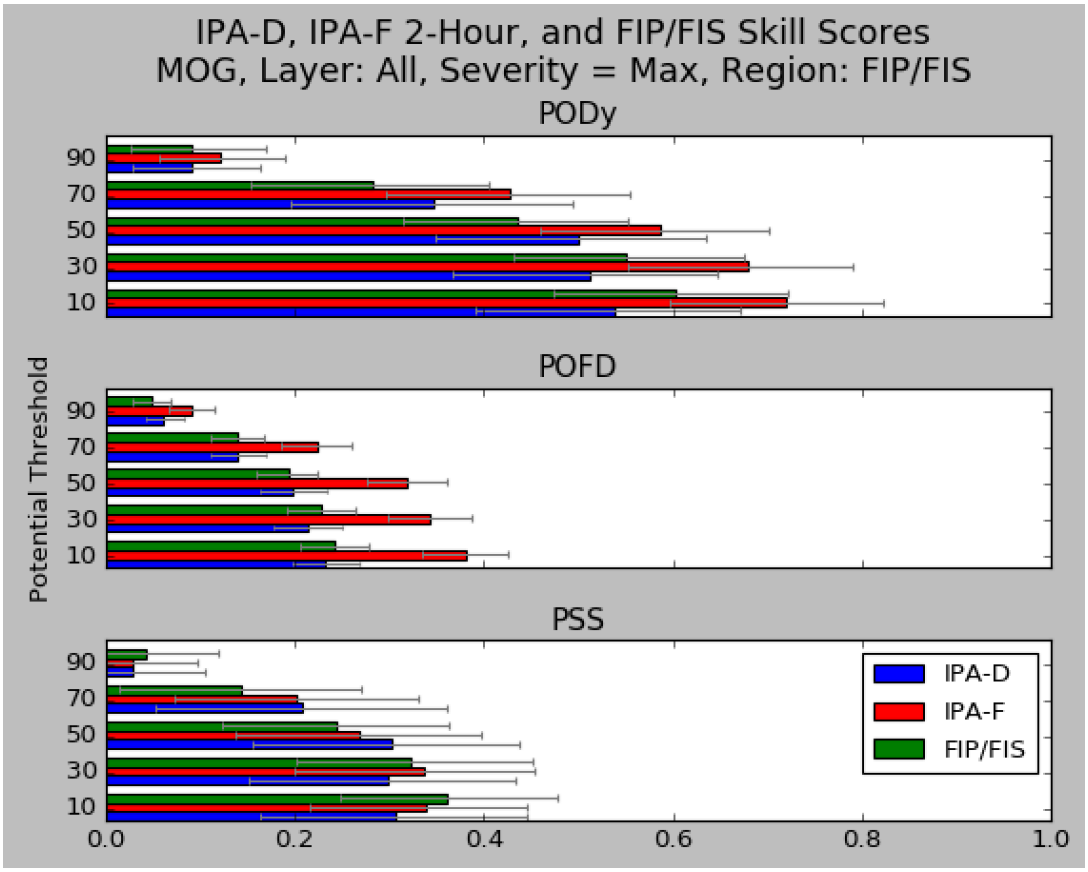

FIGURE 24. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE), IPA-F (RED), AND FIP/FIS (GREEN) IN THE FIP/FIS DOMAIN. ERROR BARS INDICATE THE 95% CONFIDENCE LEVEL. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF MODERATE OR GREATER ICING.

Skill scores were also divided by hour of day to look for diurnal trends in the algorithm performance. In the IPA region (Figure 25), there were few PIREPs between 06Z and 12Z which created the noise seen during those times. IPA-D performed best from 13–15Z and again at 00Z. IPA-F performed best from 18–22Z, but none of the differences attained statistical significance. The results for the FIP/FIS region (Figure 26) contains fewer valid hours because the FIP/FIS is only issued four times per day and the 06-Z and 12-Z valid times did not contain enough associated PIREPs to create usable information. Again, IPA-D performs best at 00Z while IPA-F and FIP/FIS perform better at 18Z. Interestingly, all the algorithms had lower POFDs at 00Z than at 18Z in the FIP/FIS domain.

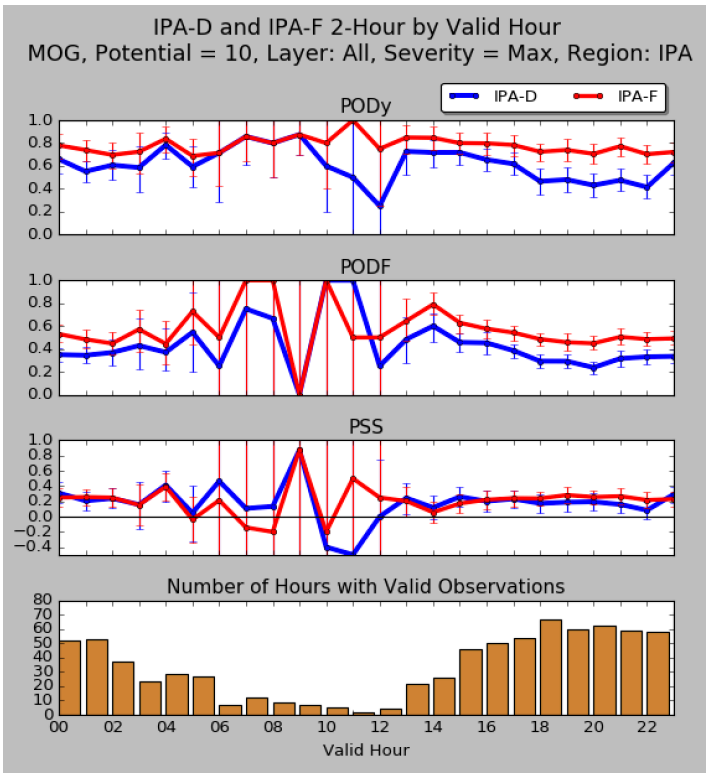

FIGURE 25. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE) AND IPA-F (RED) IN THE IPA DOMAIN WITH ERROR BARS INDICATING THE 95% CONFIDENCE LEVEL. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF MODERATE OR GREATER ICING.

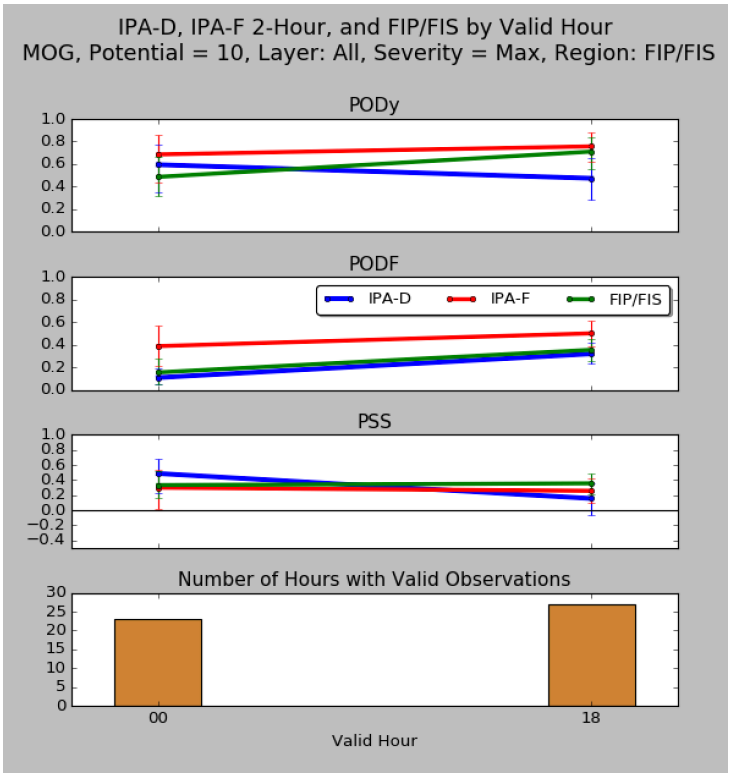

FIGURE 26. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D (BLUE), IPA-F (RED), AND FIP/FIS (GREEN) IN THE FIP/FIS DOMAIN WITH ERROR BARS INDICATING THE 95% CONFIDENCE LEVEL. THE MAXIMUM SEVERITY FROM THE NEIGHBORHOOD OF EACH ALGORITHM'S GRID POINTS WAS COMPARED TO PIREPS OF MODERATE OR GREATER ICING.

At the potential threshold of 10, which corresponds to the maximum PSSs of each algorithm, IPA-D performed slightly better than IPA-F and FIP/FIS with LOG PIREPs and slightly worse than IPA-F and FIP/FIS with MOG PIREPs. In general, IPA-D was relatively better at identifying light icing while IPA-F and FIP/FIS were relative better at identifying moderate icing.

6.2.2 Performance Metrics Using TAMDAR

In addition to icing from PIREPs, icing from TAMDAR-equipped aircraft was used to identify the presence of icing. Only the positive reports of icing were used from TAMDAR, which resulted in POD being the only score calculated for each algorithm. The resulting PODs for IPA-D, IPA-F, and FIP/FIS are displayed in Figure 27 for the LOG and MOG thresholds of each algorithm. As was the case with the PODs derived from PIREPs, IPA-D's POD was generally lower than the PODs of both IPA-F and FIP/FIS in the LOG case. At the MOG threshold, IPA-D had lower PODs than IPA-F at all but the 90-potential level. As expected, the PODs for all algorithms were lower at the MOG threshold than the LOG threshold, but similar to the PIREP PODs, IPA-D experienced a larger decrease in POD than IPA-F when moving from the LOG threshold up to the MOG threshold. Results in the IPA domain (not shown) were similar.

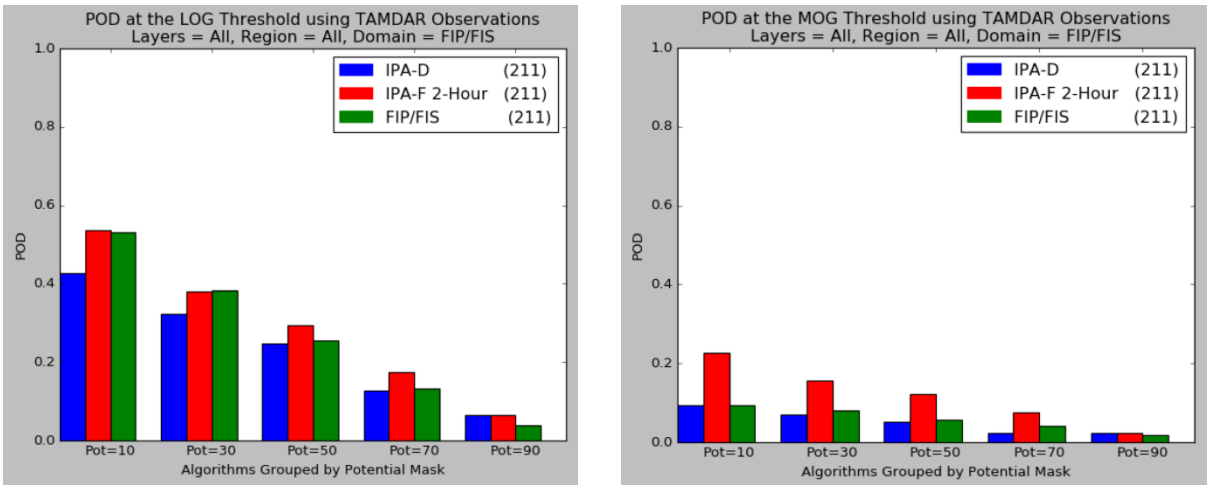

FIGURE 27. POD CALCULATED FROM TAMDAR FOR ICING ALGORITHMS AT THE LOG (LEFT) AND MOG (RIGHT) THRESHOLDS IN THE FIP/FIS DOMAIN.

6.2.3 PERFORMANCE METRICS USING RADIOSONDE

The performance metrics calculated from the radiosonde observations showed a similar story. Figure 28 shows the POD and POFD calculated against radiosonde observations at the MOG threshold in the FIP/FIS domain. Again, IPA-D had a lower POD and POFD than IPA-F at each potential level. One difference using this platform was that IPA-D was able to achieve higher PODs than FIP/FIS, while having similar POFDs.

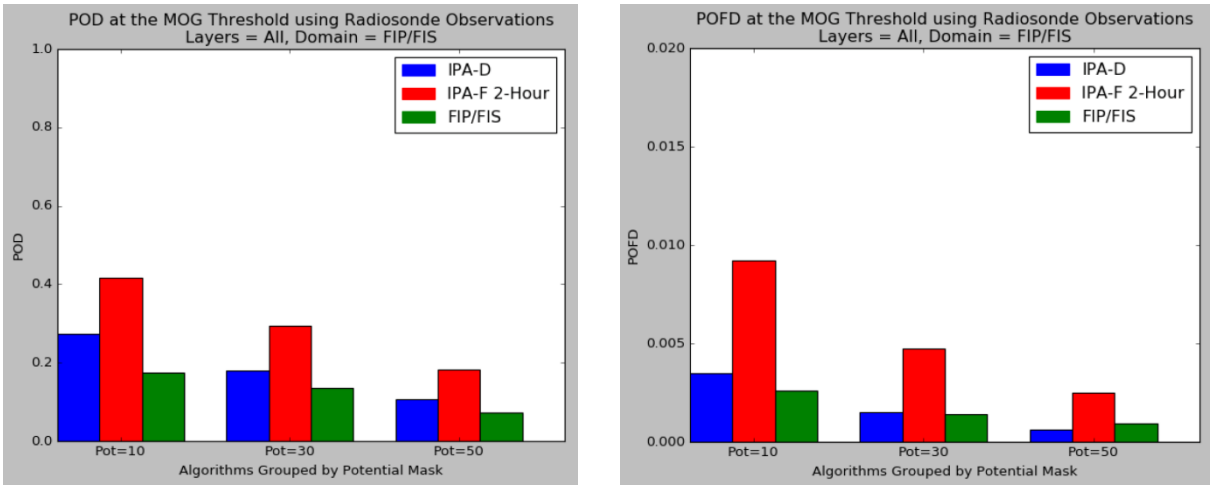

FIGURE 28. POD (LEFT) AND POFD (RIGHT) CALCULATED FROM RADIOSONDE OBSERVATIONS FOR ICING ALGORITHMS IN THE FIP/FIS DOMAIN. NOTE THAT THE Y-AXIS FOR POFD IS REDUCED TO HIGHLIGHT THE DIFFERENCES BETWEEN THE ALGORITHMS. THE POFDS ARE SMALLER USING RADIOSONDE OBSERVATIONS BECAUSE THERE WERE A LARGE NUMBER OF CASES IN WHICH CORRECT IDENTIFICATIONS OF NO ICING WERE MADE.

6.2.4 Performance Metrics Using Satellite

Satellite observations from CALIPSO and CloudSat were also used to derive performance metrics for the three algorithms and are shown in Figure 29. IPA-D had lower PODs and POFDs than IPA-F and FIP/FIS at all valid times, which reinforces the results using PIREPs, TAMDAR, and Radiosondes. It is important to keep in mind that each hour also corresponds to a specific geographic region of Alaska given the sun-synchronous orbits of CloudSat and CALIPSO. The ground tracks of the satellites move from east to west over Alaska throughout the day, beginning near the Yukon border at 21Z and proceeding towards the Aleutians by 00Z. The relative difference in POD and POFD between IPA-D and IPA-F increased slightly during the four hours that satellite data was present daily over Alaska. The maximum differences occurred at 00Z, which aligned geographically with the Aleutians and Bering Sea.

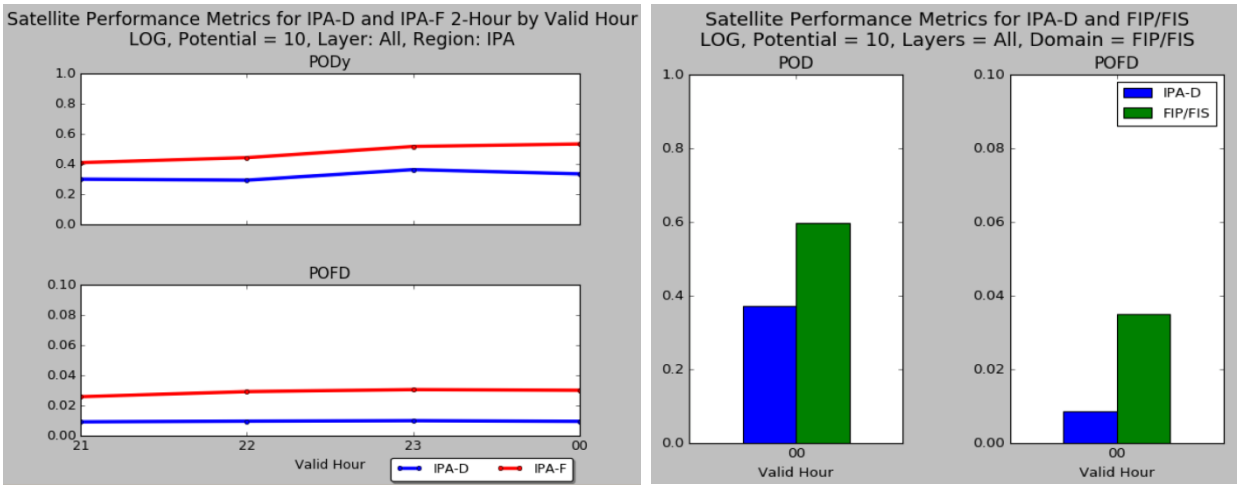

FIGURE 29. POD (LEFT GRAPHIC, TOP PANEL) AND POFD (LEFT GRAPHIC, BOTTOM PANEL) FOR IPA-D AND IPA-F IN THE IPA DOMAIN. POD (RIGHT GRAPHIC, LEFT PANEL) AND POFD (RIGHT GRAPHIC, RIGHT PANEL) FOR IPA-D AND FIP/FIS IN THE FIP/FIS DOMAIN. THE HOURS FROM 21 TO 00Z WERE SELECTED IN THE IPA DOMAIN BECAUSE THAT IS WHEN THE DAILY SATELLITE PASSES WERE OVER ALASKA. 00Z WAS THE ONLY TIME AVAILABLE IN THE FIP/FIS DOMAIN WHEN USING THE UNION OF SATELLITE OBSERVATIONS AND FIP/FIIS VALID TIMES.

6.3 Analysis by Region

6.3.1 Geographic Distributions

There were differences in the geographic distributions of icing in IPA-D as compared to IPA-F and FIP/FIS. As was mentioned in section 6.1.2, IPA-D had less icing identified than IPA-F. In Figure 30, it is clear from the widespread extent of red coloring that most geographic areas have more icing identified by IPA-F than by IPA-D, as was expected from the field distribution analysis. There are important differences, however, in mountainous regions such as the Alaska Range and in northern British Columbia. Here, IPA-D identified more icing, which is significant given the rarity of that occurrence. The differences between IPA-D and IPA-F decreased as the potential threshold increased from 10 through 90. The same was true for severity, with the differences decreasing as severity increased from LOG, to MOG, to severe. Figure 30 also highlights the influence of METAR observations in the IPA-D data. In the Northwest Territories and Alberta,

there are several darker red circles, which are regions centered around METARs where IPA-D removed icing.

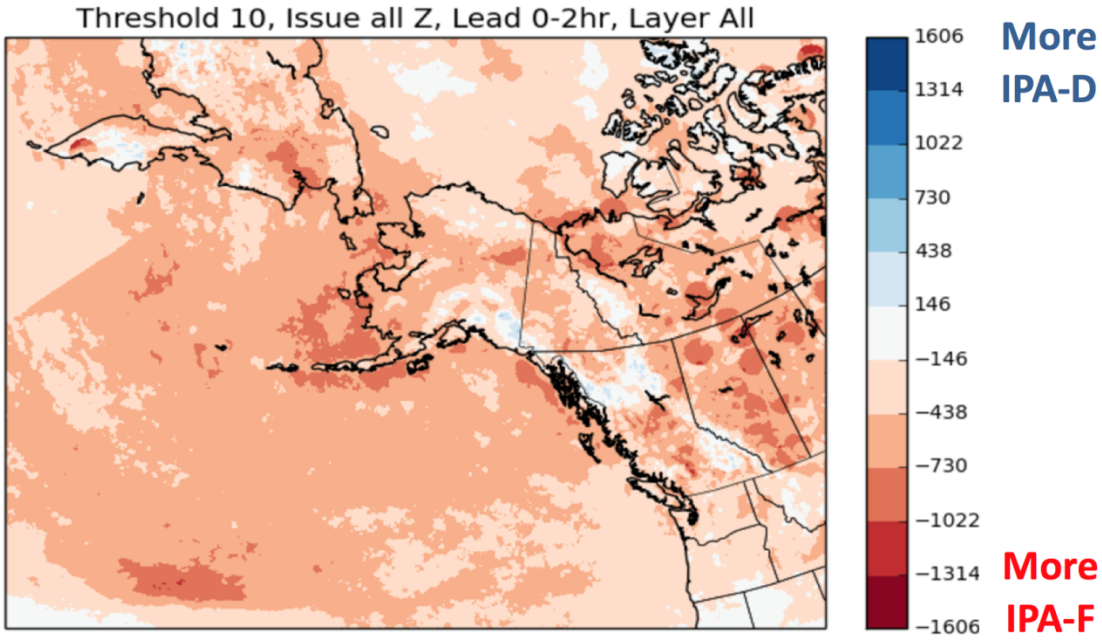

FIGURE 30. GEOGRAPHIC DISTRIBUTION OF IPA-D MINUS IPA-F ICING POTENTIALS AT A POTENTIAL THRESHOLD OF 10.

There were also differences in the probability and severity fields of IPA-D compared to FIP/FIS. Similar to IPA-F, FIP had more icing than IPA-D in most areas except for the mountainous areas. Again, the differences decreased with increasing potential thresholds. The geographic differences at the LOG severity between IPA-D and FIS (not shown) mirrored those between the probability fields of IPA-D and FIP, namely that FIS had more LOG icing in most areas except for the mountains. Increasing the severity to the MOG threshold, the story changes and IPA-D was more prevalent than FIS in all geographic areas, as was expected from the distribution analysis. In particular, the differences were the largest in the Bering Sea and in the coastal mountains of British Columbia.

6.3.2 Performance Metrics by Region

PIREPs and TAMDAR icing reports were widely distributed throughout Alaska and lent themselves well to being segregated by region. Figure 31 shows the PIREP-based skill scores at the LOG threshold for each region. In all regions, IPA-D had a lower POD, as well as a lower POFD, than IPA-F. The same held true for IPA-D when compared to FIP/FIS, except in the North Slope, where IPA-D had a higher POD than FIP/FIS. IPA-D achieved a higher PSS than IPA-F and FIP/FIS in four of the six regions, namely the Kotzebue and Norton Sound, North Slope, South Central, Gulf, and Southeast regions. Skill scores for the three algorithms generally varied between 0.3 and 0.6, except in the North Slope region where skills scores were much lower due to higher POFDs in that region. When moving up in severity to MOG icing (not shown), the PODs among the three algorithms showed a greater decrease in the northern three regions when compared against the southern three regions. POFD also decreased more in the northern regions

compared to the southern regions, which resulted in generally lower skill scores in the northern regions. At the MOG severity threshold, IPA-D achieved a higher PSS than IPA-F and FIP/FIS in two of the six regions, the Kotzebue and Norton Sound and Gulf, and Southeast regions.

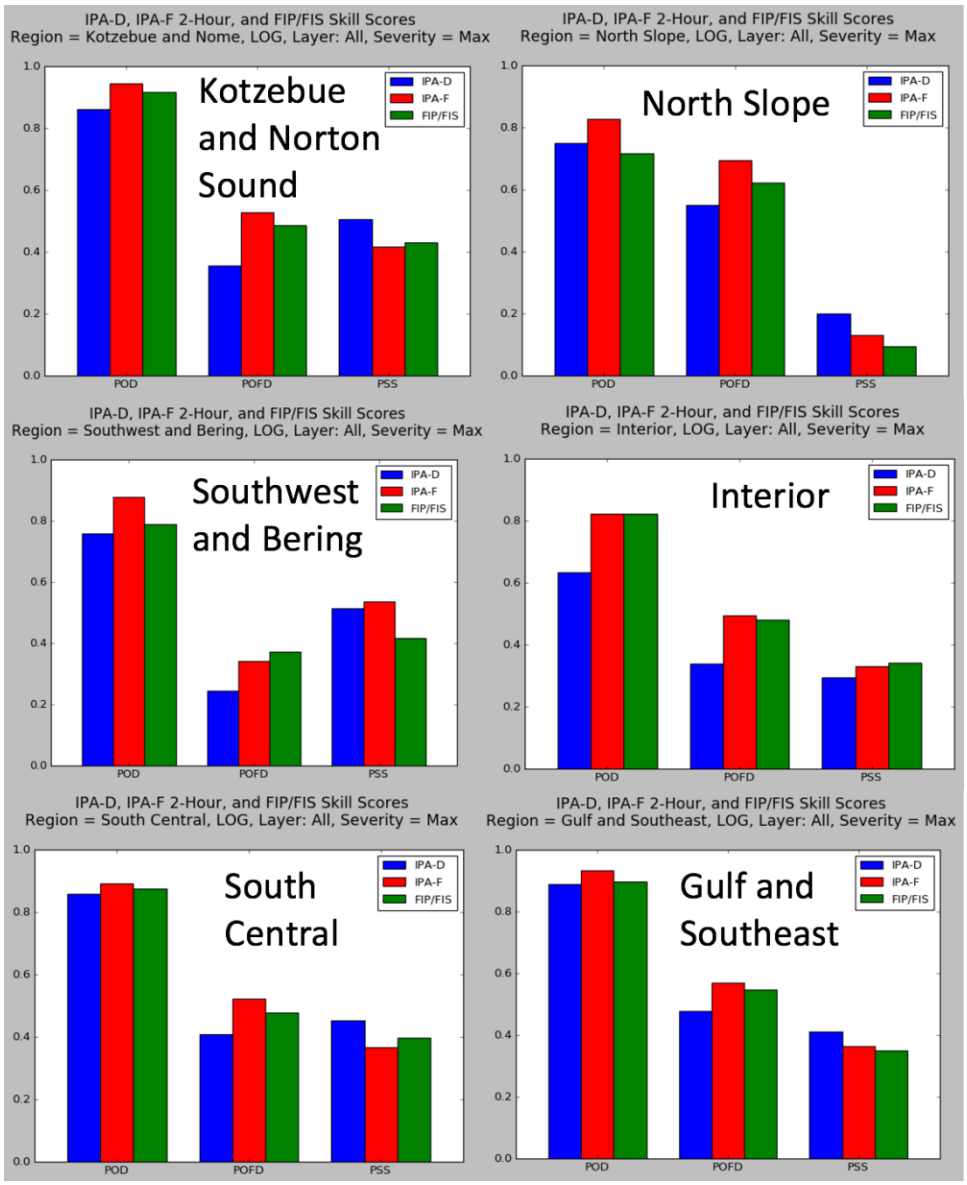

FIGURE 31. POD, POFD, AND PSS OF IPA-D, IPA-F, AND FIP/FIS IN EACH OF THE SIX REGIONS. SCORES WERE BASED ON 140 PIREPS IN KOTZEBUE AND NORTON SOUND, 121 PIREPS IN THE NORTH SLOPE, 74 PIREPS IN THE SOUTHWEST AND BERING, 145 PIREPS IN THE INTERIOR, 150 PIREPS IN THE SOUTH CENTRAL, AND 110 PIREPS IN THE GULF AND SOUTHEAST REGION.

TAMDAR-based PODs for each region (not shown) showed similar distributions among the three algorithms. In particular, the TAMDAR-based PODs reinforced the fact that the IPA-D POD lagged behind the IPA-F POD more in the Interior, Southwest, and Bering than in the other regions.

6.4 Analysis by Altitude

6.4.1 Geographic Distributions

The relative distribution of IPA-D compared to IPA-F also varied as a function of altitude. Figure 32 shows the geographic distribution of the differences between IPA-D and IPA-F. Similar to the geographic distribution analysis by region in Section 6.3, IPA-F had higher counts of icing in most areas, with the exceptions being mountainous regions. The differences between the two algorithms reached a maximum in the low layer (Figure 32) and decreased at altitudes above and below that level.

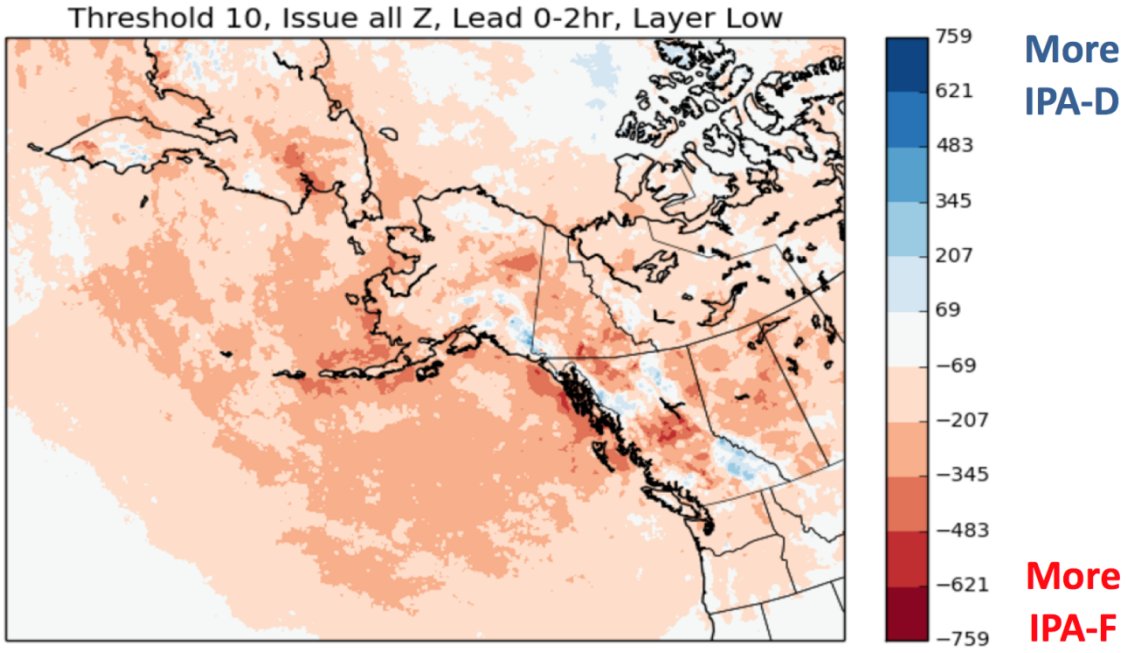

FIGURE 32. GEOGRAPHIC DISTRIBUTION OF IPA-D MINUS IPA-F FOR THE LOW LAYER ALTITUDES FROM 6,500 – 12,000 FT.

6.4.2 Performance Metrics by Altitude

PIREP and TAMDAR icing reports ranged in altitude from the surface to 30,000 feet during the assessment period. The vertical range of these reports allowed them to be binned by altitude and analyzed, the results of which are displayed in Figure 33. Here again, IPA-D trailed behind the other two algorithms in POD at every level, but also had the lowest POFDs for all but the high layer. IPA-D attained the highest PSSs for the near surface and low layers. The results using TAMDAR observations were similar, with IPA-D lagging behind IPA-F and FIP/FIS in POD at all but the high layer.

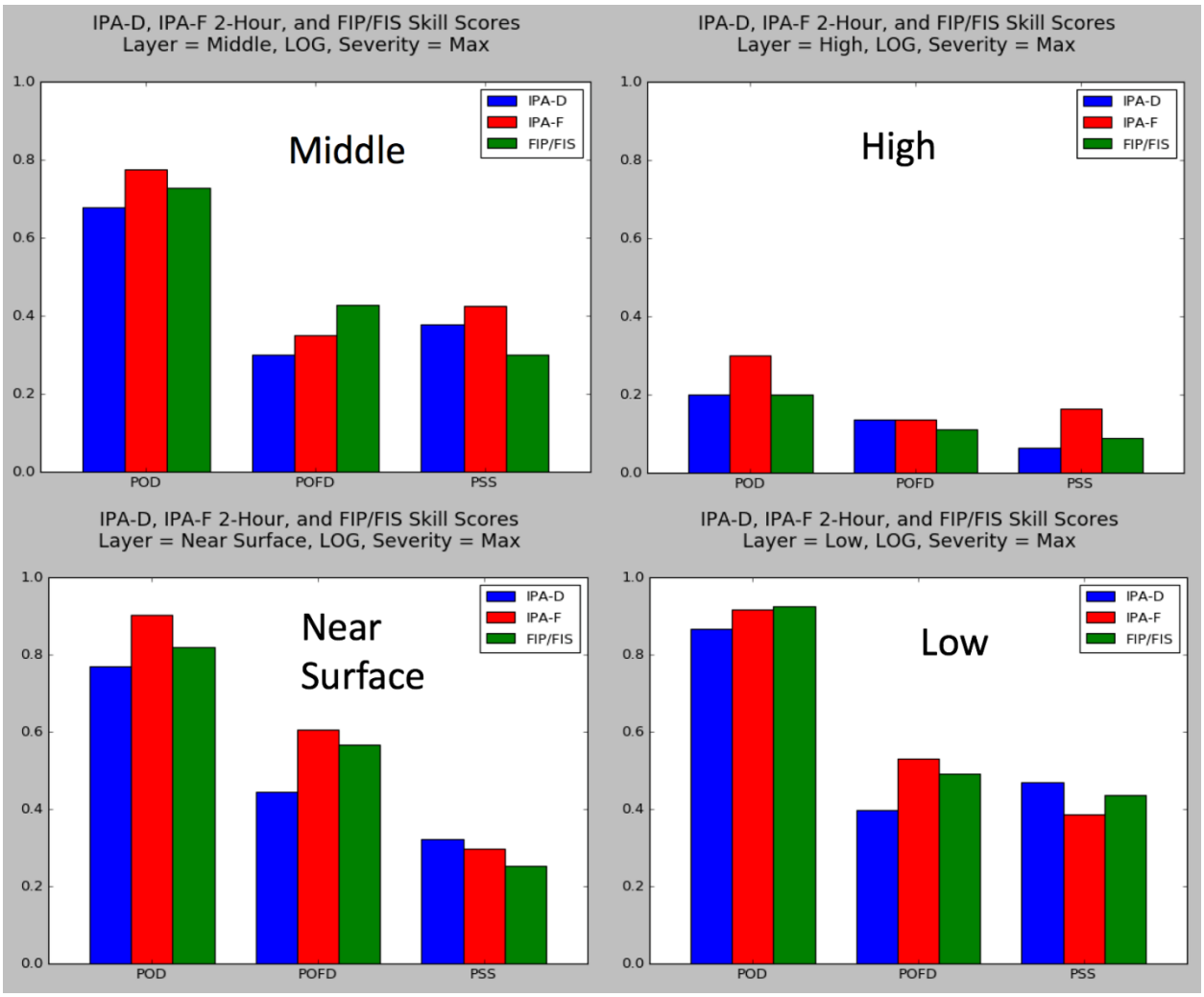

FIGURE 33. POD, POFD, AND PSS FOR IPA-D, IPA-F, AND FIP/FIS AT FOUR ALTITUDE STRATIFICATIONS. SCORES WERE BASED ON 419 PIREPS IN THE NEAR SURFACE LAYER, 233 PIREPS IN THE LOW LAYER, 51 PIREPS IN THE MIDDLE LAYER, AND 47 PIREPS IN THE HIGH LAYER.

6.5 Analysis of SLD

6.5.1 Field Distribution of SLD

SLD was rarely present in the output of IPA-D and IPA-F, occurring in 0.0053% and 0.0066% of grid points in each algorithm, respectively. Figure 34 shows the distributions of SLD for IPA-D and IPA-F, as well as the ratio between the two. Similar to the icing probability distribution in Section 6.1.2, overall IPA-D had fewer positive values of SLD than IPA-F. IPA-D accumulated most of its positive SLD values at the 0.68 value, while IPA-F had a smoother distribution of positive values.

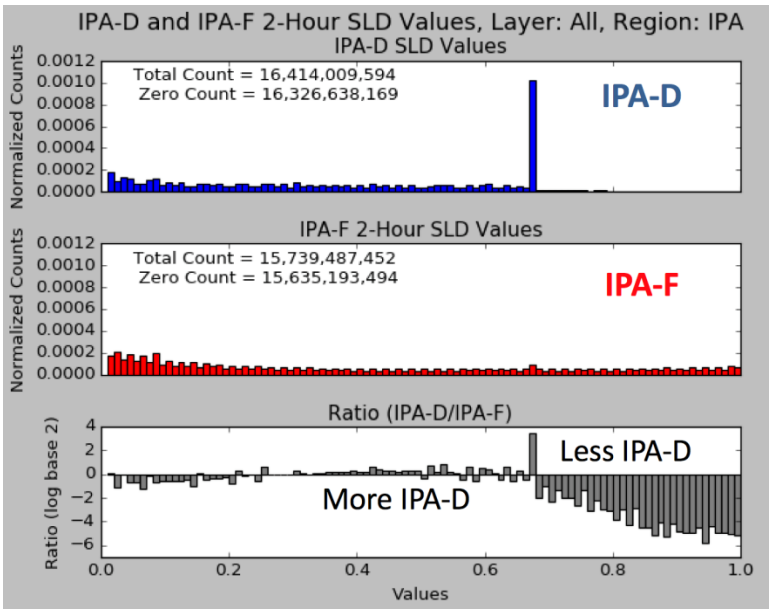

FIGURE 34. DISTRIBUTION OF PROBABILITY VALUES FOR IPA-D (TOP) AND IPA-F (MIDDLE) IN THE IPA DOMAIN. THE BOTTOM PANEL SHOWS THE LOG BASE 2 RATIO OF THE NORMALIZED COUNTS, WITH POSITIVE VALUES INDICATING MORE RELATIVE COUNTS OF IPA-D.

6.5.2 Performance Metrics for SLD

METARs were used as the observations against which SLD was verified. Freezing rain and freezing drizzle were used to identify the presence of SLD, while clear skies or snow were used to identify that SLD was absent. Figure 35 shows the skill score results across each valid hour. IPA-D and IPA-F SLD performance was very similar. The PODs for each were slightly higher during the 18 through 01Z times, resulting in improved skill scores during that period.

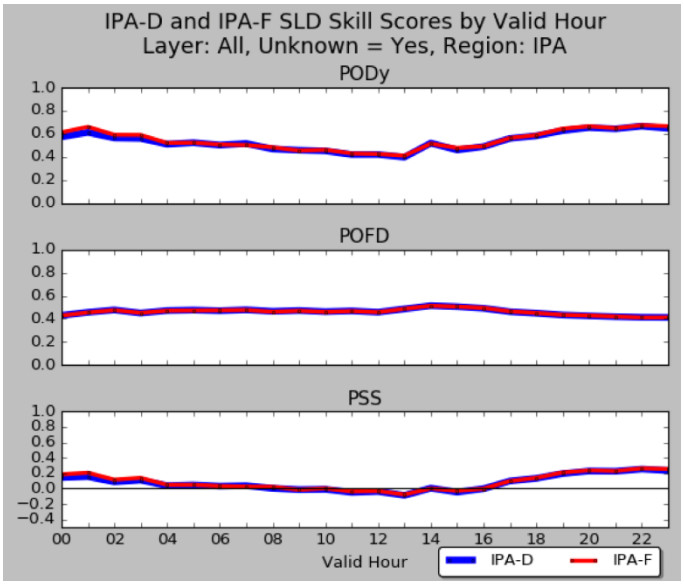

FIGURE 35. POD (TOP), POFD (MIDDLE), AND PSS (BOTTOM) FOR IPA-D AND IPA-F COMPARED AGAINST METARS.

6.6 Case Studies

A list of potential case study dates was developed with the AAWU. From that list, three periods were selected based upon the prevalence of MOG icing and the geographic and temporal proximity of the PIREPs.

6.6.1 Anchorage on 1 December 2016

On December 1, 2016, there were six PIREPs of light and five PIREPs of moderate icing in the Anchorage Bowl between 1900Z and 2359Z. The reports ranged in altitude from 1,000 ft to 9,000 ft, but the bulk of the reports were in the 5,000 – 9,000 ft layer.

The synoptic setup for this event was a decaying low-pressure system that had moved into the Prince William Sound during the icing period. As the low filled, it left a stationary boundary over the Cook Inlet stretching towards Palmer. The 2 December 00Z radiosonde from Anchorage showed a moist layer from the surface through 10,000 ft, with temperatures in this layer ranging from -2 to -15°C. METARs from Anchorage during this event reported intermittent light snow and ceilings ranging from 700 through 6,000 ft.

A snapshot of the outputs from the IPA-D, IPA-F, and FIP probability fields is shown in Figure 36. The 7,000 ft altitude layer at 21Z was selected as being representative of each algorithm's output during the time of the icing PIREPs. Of the three algorithms, IPA-D had the highest probabilities in the Anchorage Bowl region that accurately reflected the potential for icing in this scenario. It was noteworthy that IPA-D also had fill values located in close proximity to where the moderate icing reports occurred.

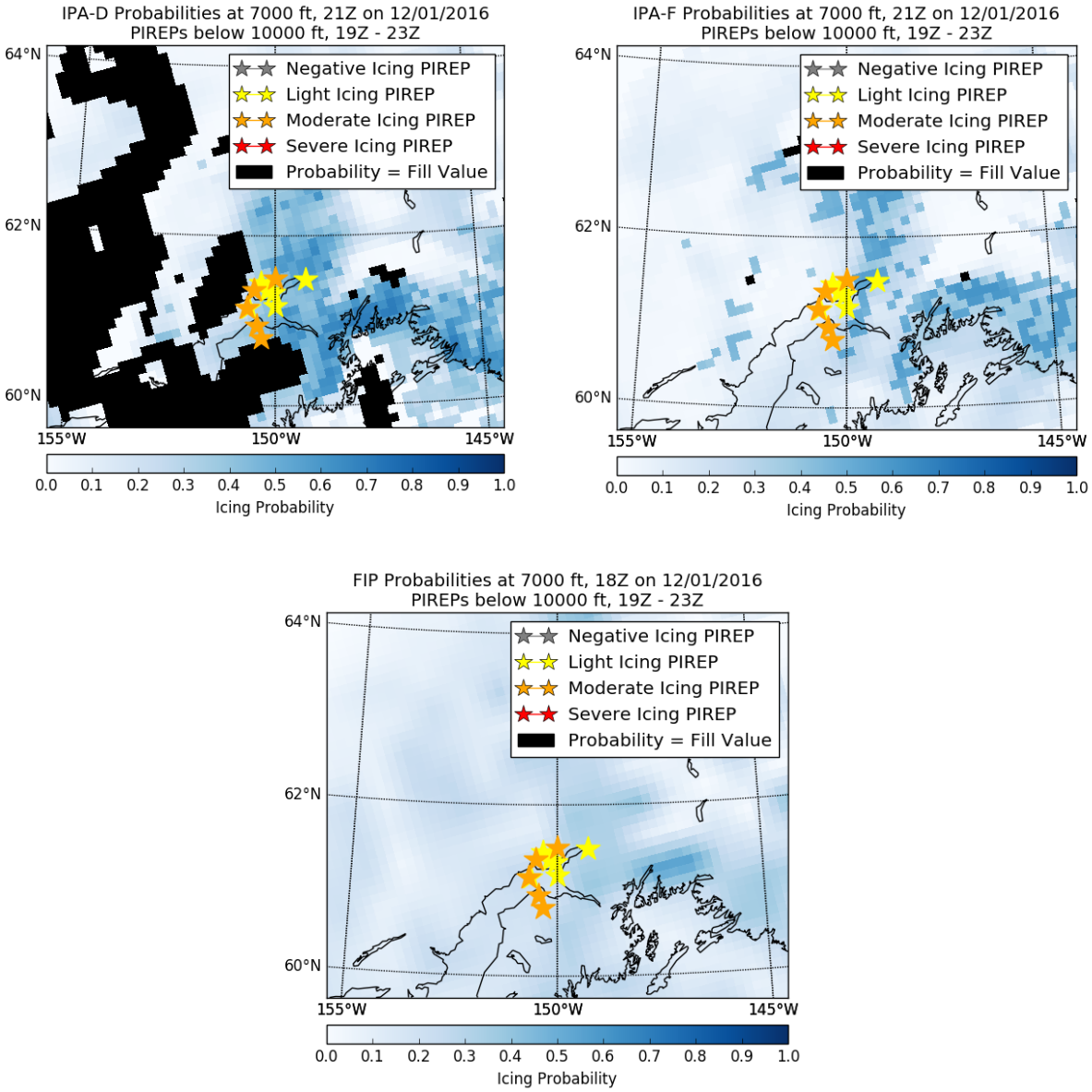

FIGURE 36. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIP (BOTTOM) PROBABILITY FIELDS OVER THE ANCHORAGE BOWL ON 1 DECEMBER 2016 AT 21Z AND FOR THE 7,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 1900Z AND 2359Z AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

Outputs from the severity fields of IPA-D, IPA-F, and FIS are displayed in Figure 37. IPA-D and IPA-F both identified moderate and severe icing in a corridor west of the Chugach mountains from Anchorage north to Palmer. IPA-F was more accurate than IPA-D with its placement of moderate icing south of the Turnagain Arm, which matched two PIREPs that occurred there. Both IPA-D and IPA-F missed the moderate icing west of Anchorage, instead placing only light icing there. FIS blanketed the region with moderate icing and low probabilities, providing the least utility of the three algorithms.

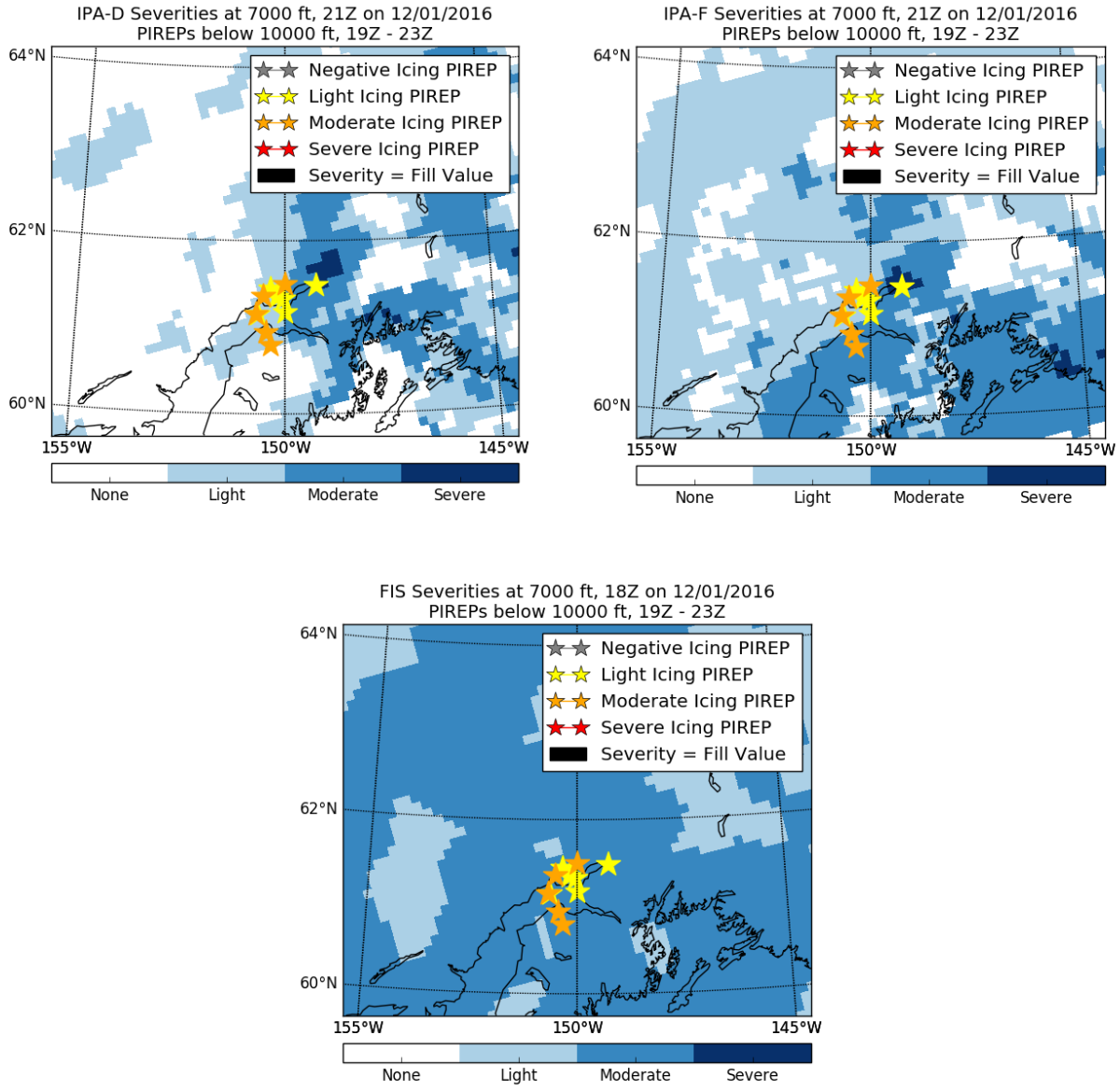

FIGURE 37. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIS (BOTTOM) SEVERITY FIELDS OVER THE ANCHORAGE BOWL ON 1 DECEMBER 2016 AT 21Z AND FOR THE 7,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 1900Z AND 2359Z, AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

6.6.2 Western Alaska on 22 December 2016

On December 22, 2016, there were two cases of icing in Western Alaska. The first case occurred earlier in the day, between 1500Z and 2100Z and consisted of four PIREPs of light and four PIREPs of moderate icing that occurred mostly near Kotzebue. The reports ranged in altitude from 2,000 ft to 11,000 ft, with the moderate icing confined to 3,000 – 6,000 ft. Later in the day, another grouping of icing reports occurred between Unalakleet and Cold Bay between 2100Z and 2359Z. This second case consisted of four reports of moderate icing and one report of moderate to severe icing that occurred between 500-ft to 4,000-ft.

The synoptic setup for this event was a weak low-pressure system moving east through the Bering Sea. A warm front associated with the low made landfall over the Yukon Delta region near 2100Z triggering the icing. The 23 December 00Z radiosonde over Nome showed a moist layer from the surface through 5,000-ft, with temperatures in this layer ranging from -7 to -17°C. METARs from Kotzebue during the first part of the event reported intermittent light snow and ceilings ranging from 1,600-ft through 3,700-ft. Closer to the second case, the 23 December 00Z radiosonde from Bethel reported moist layers between the surface and 2,500-ft and between 7,000-ft and 10,000-ft, with temperatures between ranging between -2 and -8°C in the lower layer. METARs from Cape Romanzof had ceilings varying from 300-ft up to 1,800-feet during the event.

The probability fields from IPA-D, IPA-F, and FIP during the first event in Western Alaska are shown in Figure 38. IPA-D and IPA-F had low probabilities of icing near Kotzebue during this event while FIP had probabilities near zero.

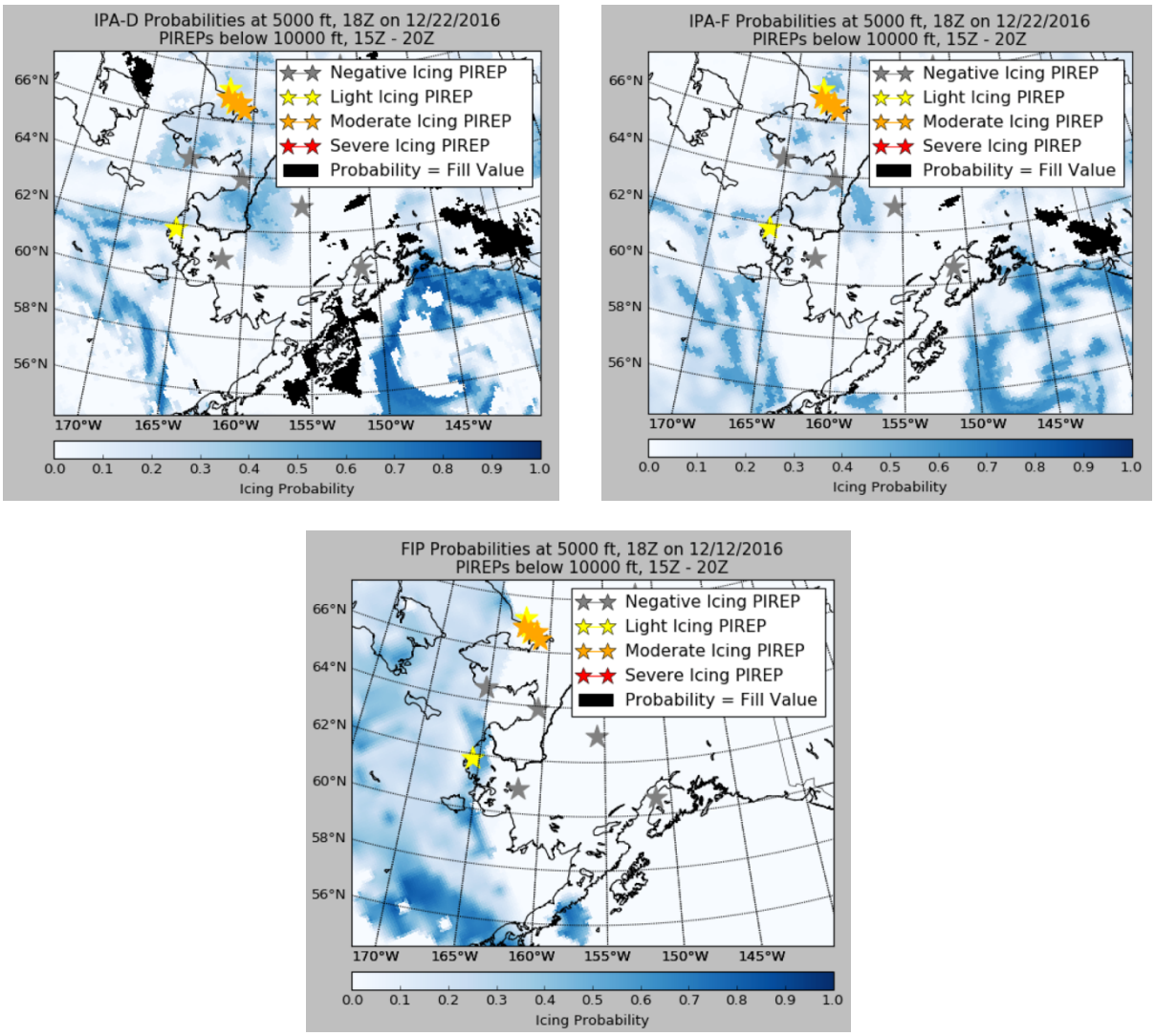

FIGURE 38. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIP (BOTTOM) PROBABILITY FIELDS OVER WESTERN ALASKA ON 22 DECEMBER 2016 AT 18Z AND FOR THE 5,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 1500Z AND 2100Z, AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

The severity fields for the first case are shown in Figure 39. All three algorithms identified light icing near Kotzebue where the moderate icing PIREPs occurred. Both IPA-D and IPA-F had some areas of moderate icing on the Seward Peninsula and IPA-F had the greater extent of moderate between the two algorithms. FIS had most of its moderate icing located offshore.

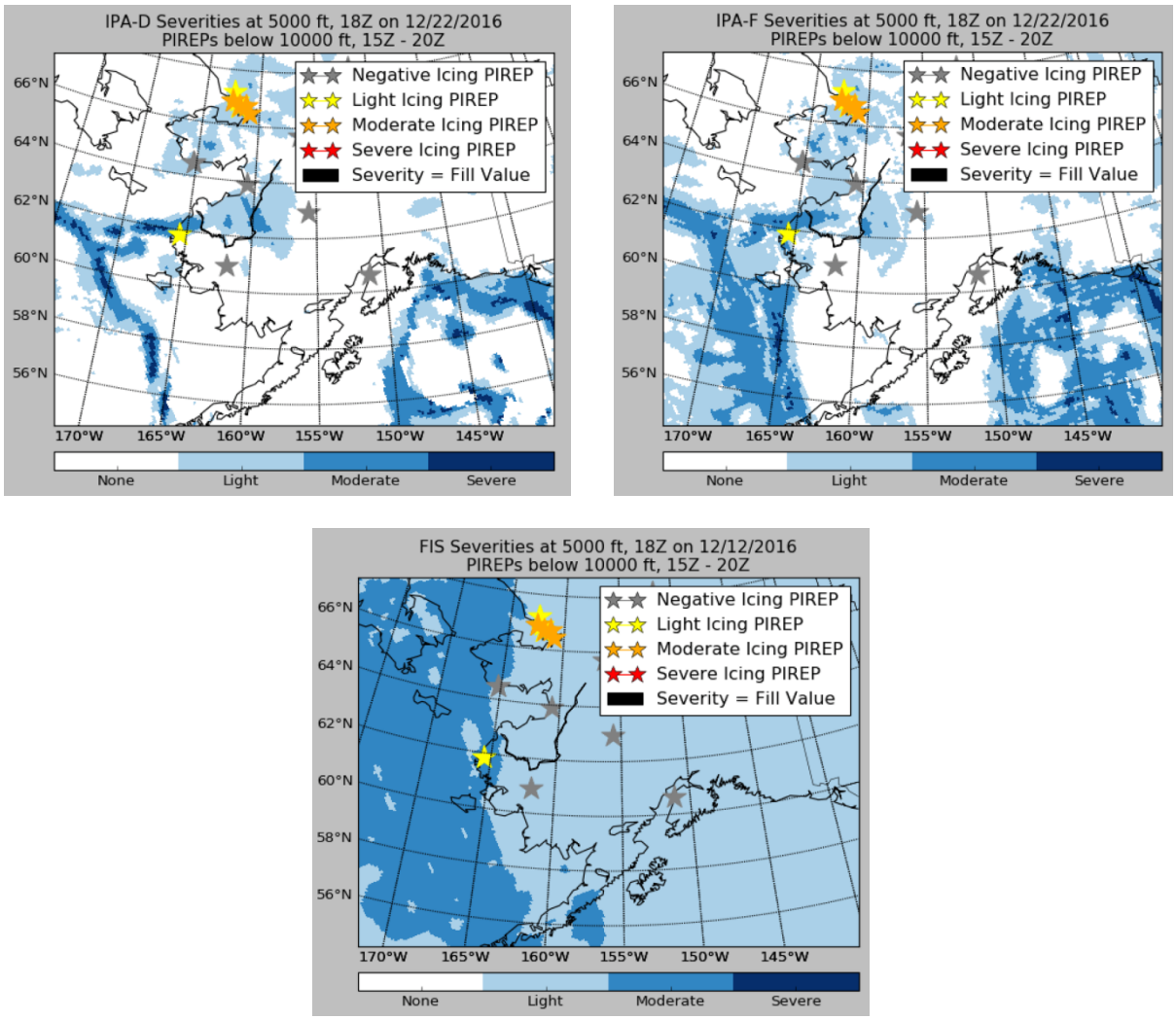

FIGURE 39. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIS (BOTTOM) SEVERITY FIELDS OVER WESTERN ALASKA ON 22 DECEMBER 2016 AT 18Z AND FOR THE 5,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 1500Z AND 2100Z, AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

The probabilities from the second case are presented in Figure 40. Again, IPA-D and IPA-F have relatively low probabilities of icing near the PIREP locations, especially near the moderate to severe report near Unalakleet. The algorithms handled the icing near Cape Romanzof best, with the highest probabilities located just offshore from the PIREP. IPA-F and IPA-D also did well with the PIREP near Cold Bay; however, IPA-D did have significant areas of fill values present along the Alaska Peninsula and over Katmai National Park. FIP probabilities were more diffuse, but were at their maximum over the Yukon Delta region.

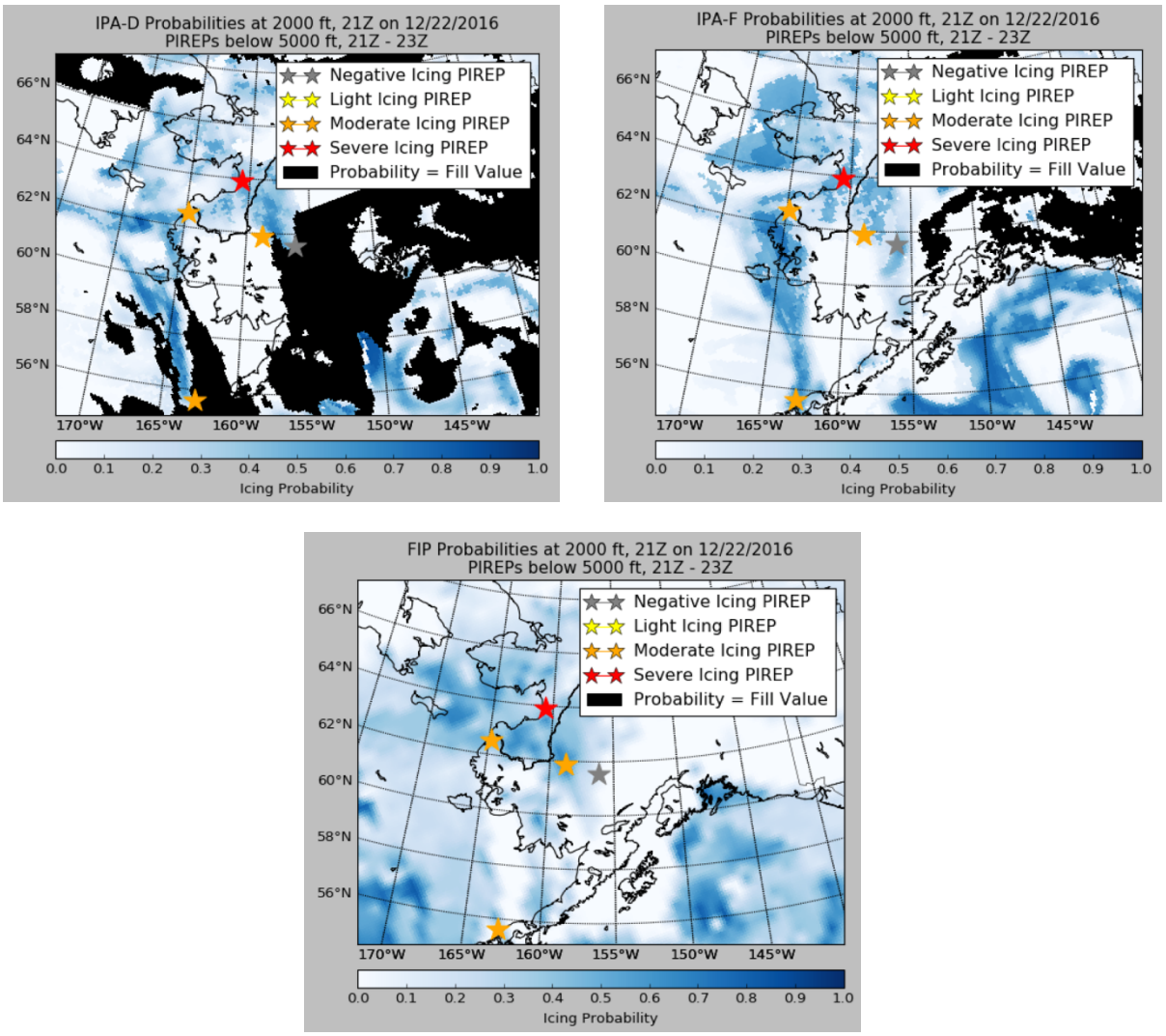

FIGURE 40. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIP (BOTTOM) PROBABILITY FIELDS OVER WESTERN ALASKA ON 22 DECEMBER 2016 AT 21Z AND FOR THE 2,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 2100Z AND 2359Z, AT ALTITUDES LESS THAN 5,000 FT AGL ARE OVERLAID.

The severities from the second case are shown in Figure 41. All three algorithms correctly captured the moderate icing severity near Cape Romanzof, but the results were mixed for the other PIREPs. Both IPA-D and IPA-F missed the moderate icing severity near Unalakleet, although IPA-F was closer with its placement of moderate icing than IPA-D. IPA-D did better than IPA-F with its placement of moderate icing close to the moderate PIREP at the bend in the Yukon River. In general, IPA-D's icing areas are tailored to a smaller spatial extent than those of IPA-F. FIS was able to capture the moderate intensity of all three events in the Yukon Delta and performed best in this section of the scenario.

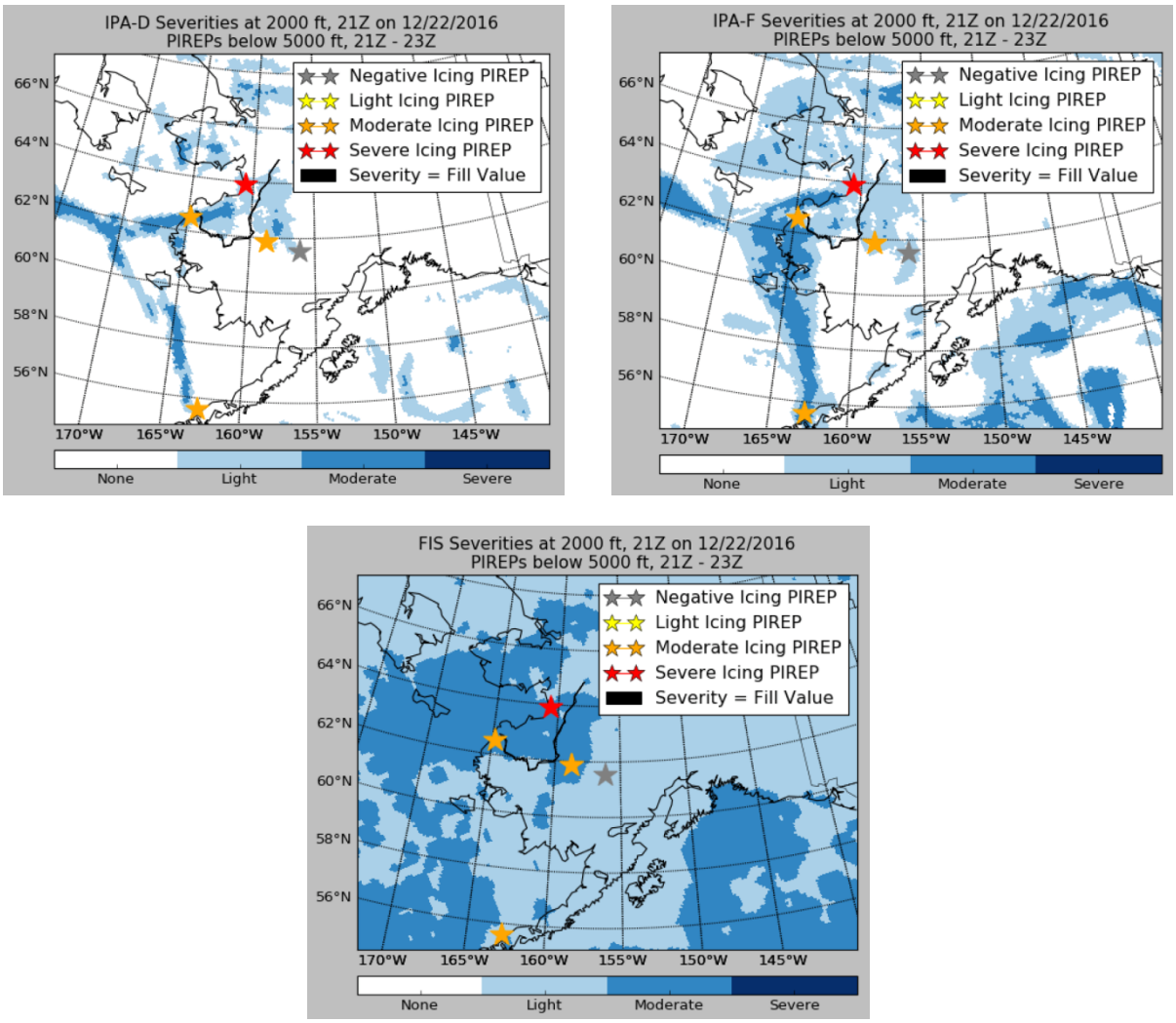

FIGURE 41. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIS (BOTTOM) SEVERITY FIELDS OVER WESTERN ALASKA ON 22 DECEMBER 2016 AT 21Z AND FOR THE 2,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 2100Z AND 2359Z, AT ALTITUDES LESS THAN 5,000 FT AGL ARE OVERLAID.

6.6.3 KETCHIKAN ON 13 JANUARY 2017

The final case study occurred on January 13, 2017, when there was one PIREP of light, two PIREPs of moderate, one PIREP of moderate to severe, and one PIREP of severe icing near Ketchikan between 2300Z and 2359Z. The reports ranged in altitude from 4,000-ft to 8,000-ft, with MOG icing found between 6,000-ft and 8,000-ft.

The synoptic setup for this event was a deep low-pressure system that had moved into the Gulf of Alaska during the icing period. An occluded front made landfall in the Alaska Panhandle, bringing icing conditions to the region. The 14 January 00Z radiosonde from Annette Island showed a moist layer from the 1,000-ft through 12,000-ft, with temperatures in this layer ranging from 4 to -10°C. METARs from Ketchikan during this event reported intermittent rain and a ceiling at 1,000-ft.

Figure 42 shows the probability fields for each algorithm during the icing event. Both IPA-F and FIP had the strongest icing probabilities offshore at this time, while IPA-D had a band of high probabilities well offshore and some higher probabilities near Ketchikan. Here, IPA-D did a better job of capturing the icing in the Panhandle region.

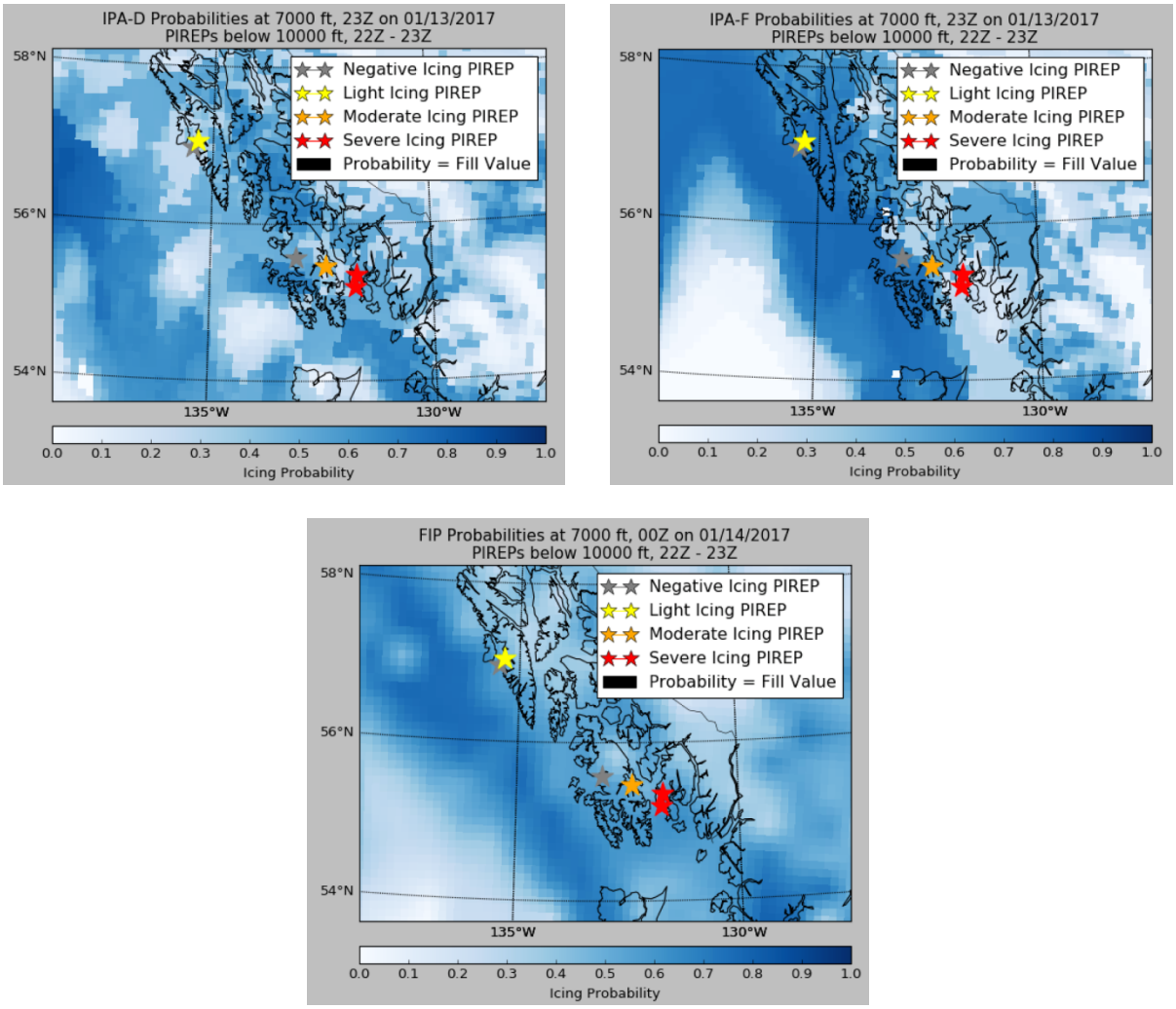

FIGURE 42. IPA-D (TOP TEFT), IPA-F (TOP RIGHT), AND FIP (BOTTOM) PROBABILITY FIELDS NEAR KETCHIKAN ON 13 JANUARY 2017 AT 23Z AND FOR THE 7,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 2300Z AND 2359Z, AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

The corresponding severity fields are displayed in Figure 43. Both IPA-D and IPA-F correctly placed severe icing near the severe reports around Ketchikan. Again, IPA-D has more tailored icing areas than IPA-F, which matched well with the PIREPs. FIS blanketed the area with moderate icing and did not provide nearly as much fidelity as the other two algorithms.

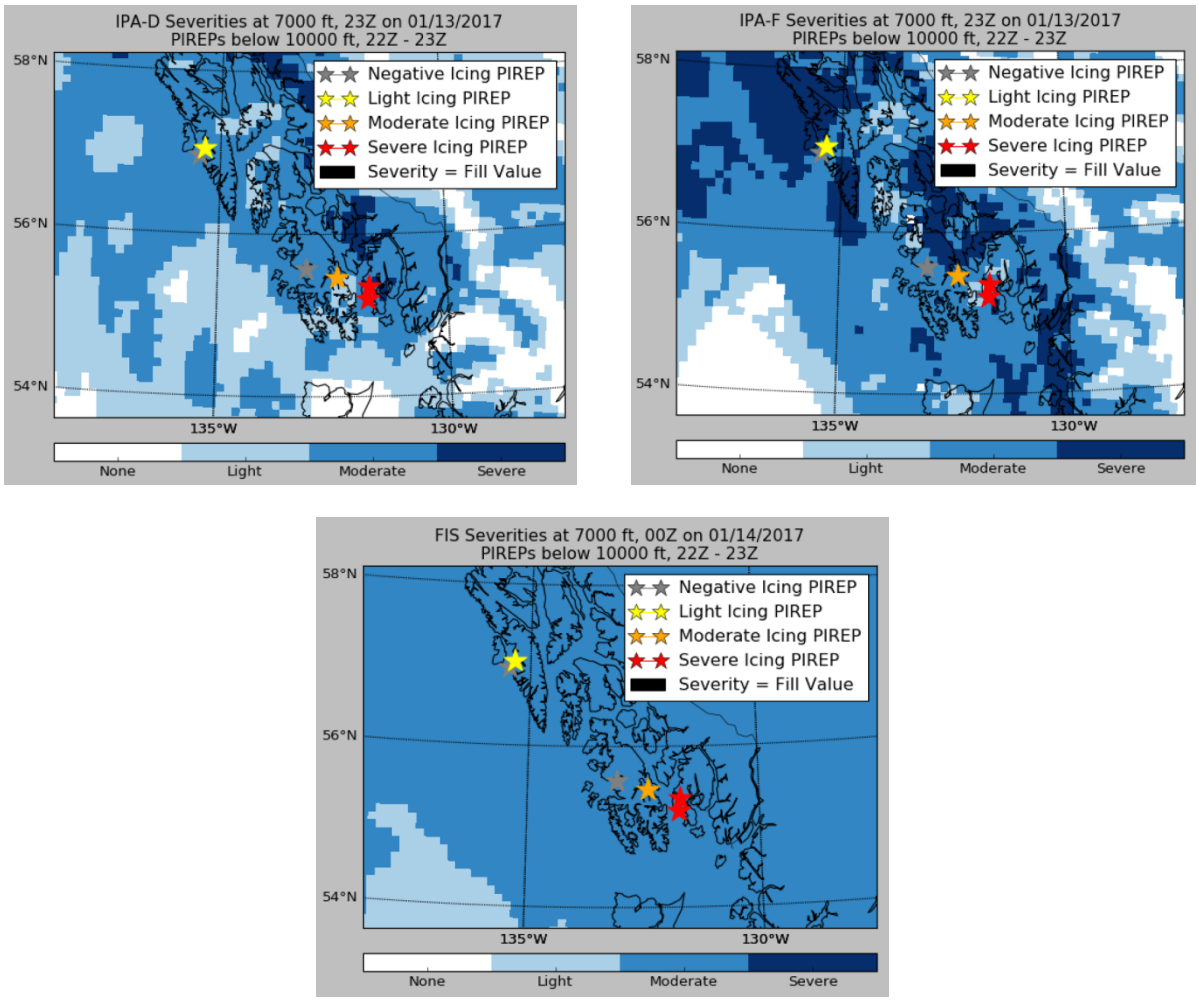

FIGURE 43. IPA-D (TOP LEFT), IPA-F (TOP RIGHT), AND FIS (BOTTOM) SEVERITY FIELDS NEAR KETCHIKAN ON 13 JANUARY 2017 AT 23Z AND FOR THE 7,000 FT LAYER. PIREPS OF ICING THAT OCCURRED BETWEEN 2300Z AND 2359Z, AT ALTITUDES LESS THAN 10,000 FT AGL ARE OVERLAID.

6.7 CONSISTENCY ANALYSIS

The consistency of IPA-D was evaluated by comparing adjacent valid times of IPA-D as well as comparing IPA-F 2 and 3-Hour forecasts to IPA-D (Figure 44). Overall, IPA-D was highly consistent with itself and with IPA-F, and the consistency behavior was similar between the LOG and MOG thresholds. The consistency decreased with increasing probability thresholds, which was expected because as the higher probabilities were more sparse than the lower probabilities. Additionally, the consistency was better at lower altitudes than higher altitudes, again due to the density of icing counts in the lower layers.

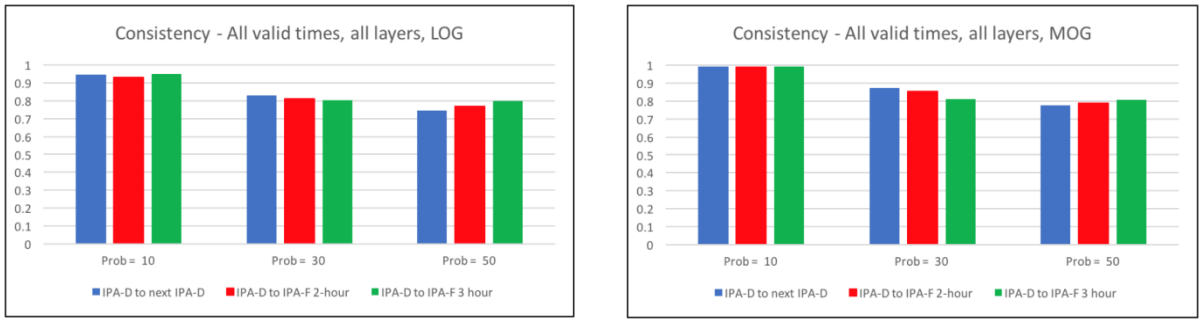

FIGURE 44. CONSISTENCY OF IPA-D AT THE LOG THRESHOLD (LEFT) AND THE MOG THRESHOLD (RIGHT).

7 CONCLUSIONS

In summary, IPA-D, IPA-F, and FIP/FIS were evaluated over a period of 15 September 2016 through 15 January 2017 and compared against observations from PIREPs, TAMDARs, radiosondes, satellites, and METARs. The results showed that with the differences between the algorithms come trade-offs to between icing detection and false alarms. On the one hand, IPA-D identified more available airspace than IPA-F, but this came at the cost of more missed icing events. One exception to the trade-off was that FIP/FIS was able to achieve both a higher POD than IPA-D over a smaller forecast volume at the lowest potential threshold evaluated. IPA-D experienced a greater decrease in POD than IPA-F at higher severities, such as the difference between the LOG and MOG detection rates. In general, IPA-D had lower PODs and POFDs than IPA-F and FIP/FIS, both of which resulted from IPA-D identifying icing in a smaller volume of airspace than the other two algorithms. The skill scores balanced out such that the three algorithms had similar results, within the 95% confidence level of one another. The results for POD and POFD were corroborated by all of the observation platforms evaluated.

The differences in skill observed among the three algorithms were consistent with the behavior in their relative distributions. Overall, IPA-D identified less icing than IPA-F and FIP, and less LOG icing than IPA-F and FIS. One geographic exception to this occurred in mountain regions, where IPA-D had higher relative counts of icing. In addition, the IPA-D probability field had a high number of fill values, which were more prevalent in the lower altitudes and in the earlier part of the assessment period. The IFI PDT is aware of the fill value issue and is working to address it.

The SLD distribution and skill scores between IPA-D and IPA-F were found to be similar and the positive SLD composed a small percentage of each algorithm's output field.

Overall, IPA-D exhibited similar skill to the IPA-F 2-Hour and FIP/FIS 0-Hour algorithms which was accomplished through an increased amount of available airspace in IPA-D.

8 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.

9 REFERENCES

Bernstein, B. C., C. A. Wolff and F. McDonough, 2007: An Inferred Climatology of Icing Conditions Aloft, Including Supercooled Large Drops. Part I: Canada and the Continental United States. *Journal of Applied Meteorology and Climatology*, 46, 1857-1878. DOI: 10.1175/2007JAMC1607.1

Brown B.G. and B.C. Bernstein, 2006: A calibration approach for CIP based on forecast performance measures. Report submitted to the FAA Aviation Weather Research Board. Available from B. Brown (bgb@ucar.edu).

Madine, S., S. A. Lack, S. A. Early, M. Chapman, J. K. Henderson, J. E. Hart, and J. L. Mahoney, 2008: Quality Assessment Report: Forecast Icing Product (FIP).

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.

Pearson, J. M., and R. Sharman, 2013: Calibration of in situe eddy dissipation rate (EDR) severity thresholds based on comparisons to turbulence pilot reports (PIREPSs). *16th Conference on Aviation, Range, and Aerospace Meteorology*. 5-10 January 2013, Austin, TC, Amer. Met. Soc.

Schultz, P., and M. K. Politovich, 1992: Toward the Improvement of Aircraft-Icing Forecasts for the Continental United States. *Wea. Forecasting*, **7**, 491–500.

Seidel, D. J., B. Sun, M. Pettey, and A. Reale, 2011. Global radiosonde balloon drift statistics. *Journal of Geophysical Research*, Vol. 116. Dio:10.1029/2010JD014891.

Wolff, C.A, F. McDonough, M.K. Politovich, B.C. Bernstein, and G.M. Cunning, 2008: FIP severity technical document. Prepared for the Aviation Weather Technology Transfer Technical Review Panel. Report available from C. Wolff (cwolff@ucar.edu).