NOAA Technical Memorandum OAR GSD-58
https://doi.org/10.7289/V5/TM-OAR-GSD-58

# Icing Product Alaska (IPA) Follow-On Assessment: Final Results

**August 2016**

Matthew S. Wandishin
Laura Paulik
Joan E. Hart
Melissa A. Petty

Earth System Research Laboratory
Global System Division
Boulder, Colorado
August 2016

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

NOAA Technical Memorandum OAR GSD-58

# Icing Product Alaska (IPA) Follow-On Assessment: Final Results

Matthew S. Wandishin[1,2]
Laura Paulik[1,2]
Joan E. Hart[1,2]
Melissa A. Petty[1,3]

[1] National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD)

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

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

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

## EXECUTIVE SUMMARY

The QA PDT was tasked to complete a follow-on quality assessment of the Icing Product Alaska Forecast (IPA-F) algorithm developed by the In-Flight Icing (IFI) PDT at the National Center for Atmospheric Research. The purpose of this second assessment was to compare the performance of the IPA-F to the current suite of Alaska icing forecast products, including the Forecast of Icing Probability (FIP), the Forecast of Icing Severity (FIS), and the gridded Alaska Aviation Weather Unit (AAWU) icing product. The results of this study are intended to assist the Technical Review Panel in determining the readiness of IPA-F to be transitioned into operations. The study was conducted on data from the autumn of 2016, with particular emphasis on the area of responsibility for the AAWU. The forecasts were compared to each other as well as to observations, including pilot reports (PIREPs), aviation routine weather reports (METARs), and upper-air soundings.

In general, the assessment findings include:

- IPA-F outperformed FIP/FIS and the gridded AAWU forecast
- IPA-F improvement was the greatest for higher probability thresholds and higher severities of icing (moderate or greater)
- IPA-F captured more moderate or greater (MOG) icing events than the gridded AAWU forecasts while forecasting that severity over a smaller area
- IPA-F forecasted larger areas of high probability and zero probability of icing than FIP; FIP forecasted a larger area of low probability of icing than IPA-F
- IPA-F forecasted a larger area of MOG icing than FIS

Overall, IPA-F outperformed FIP/FIS and the gridded AAWU forecasts when compared against observations. In particular, IPA-F shows an increase in skill when forecasting MOG icing, compared to FIP/FIS and gridded AAWU forecasts. First, IPA-F had a higher probability of detection (POD) than FIP on three of the five probability thresholds for light icing and on all five of the thresholds for moderate icing. This resulted in higher Pierce Skill Scores (PSS) for IPA-F than the FIP on four of the five probability thresholds for light icing and all five of the thresholds for moderate icing. Next, IPA-F was compared against the AAWU forecasts, where IPA-F had a higher POD at all thresholds for both light and moderate icing, leading to higher PSS for IPA-F on four of the five probability thresholds for light icing and all five thresholds for moderate icing.

IPA-F performed particularly well against the gridded AAWU forecasts. IPA-F captured more MOG icing events than the gridded AAWU forecast while forecasting the MOG intensity threshold over a smaller area. Compared to the FIP, IPA-F forecasted a greater area of high probabilities, but also a larger area of zero probabilities of icing. FIP forecasted a larger area of low probabilities for icing. Given the stronger performance of IPA-F at higher probability thresholds, IPA-F's increased forecast certainty yielded more accurate forecasts as evaluated by the PSS metric.

## TABLE OF CONTENTS

# 1 INTRODUCTION

This document outlines the Quality Assessment (QA) Product Development Team (PDT) assessment of the Icing Product Alaska (IPA) algorithm developed by the In-Flight Icing (IFI) PDT at the National Center for Atmospheric Research. Currently, the IPA consists only of a forecast component "IPA-F"; an analysis component by the IFI PDT is forthcoming.

The QA PDT performed an assessment (Assessment 1) of the IPA–F, the results of which were presented to a Technical Review Panel (TRP) to determine the readiness of IPA-F for transition to operations. The TRP responded with several recommendations for reassessment in order to determine readiness:

1. *An assessment period of Sept – Dec.* The first IPA-F assessment was Feb – May, when icing is observed less often
2. *More detail over Alaska.* One focus of the previous assessment was the region over which IPA-F and the (CONUS) FIP overlap, which is the Pacific Northwest, less important to the Alaska Aviation Weather Unit (AAWU)
3. *A comparison to the full gridded field from which the AIRMETs are generated,* rather than a comparison to the AIRMETs
4. *A comparison to the FIP/FIS (the existing icing product for Alaska being run at the AAWU),* rather than a comparison to the FIP over their overlap region
5. *Inclusion of Canadian PIREPs.* The density of PIREPs over the period of the first IPA-F assessment shows a clear 'hole' in Canada.

In response to the TRP recommendations, the new aspects in Assessment 2 included:

- An assessment period of 15 Sept 2015 – 15 Jan 2016
- A focus area of Alaska (rather than the Pacific NW overlap region), and subregions within
- A comparison of IPA-F to the FIP/FIS for Alaska, not the CONUS FIP, as the baseline product
- A comparison of IPA-F to the gridded AAWU-generated icing product, rather than the AAWU AIRMETs

This second assessment will incorporate forecast output from IPA-F, the FIP and FIS products running at the AAWU, and the AAWU-produced gridded icing forecast, as well as observation data from METARs, PIREPs, soundings, and satellite observations, to establish a performance baseline.

The assessment has six main areas of investigation, which are similar to the first IPA-F assessment:

1. Characteristics of the product fields, including climatological maps and distributions
2. Consistency of IPA forecasts between the various forecast issues and leads
3. Overall performance and meteorological accuracy of IPA-F
4. Performance of the IPA-F as compared to the FIP/FIS, where IPA-F and FIP domains overlap
5. Performance of the IPA-F as compared to the AAWU AIRMETs
6. Performance of IPA-F when used in conjunction with the AAWU AIRMETs.

This independent assessment of the IPA-F product is being performed as part of its transition into operations. The assessment will include findings from an evaluation of the product's skill, as well as a characterization of the product's strengths and weaknesses. The information from the assessment will not only serve as a baseline of performance for future versions of the product, but will also be provided to the Technical Review Panel to determine readiness of IPA-F for transition to operations.

# 2 DATA

This section describes the forecast and observation data included in the assessment, along with the principal stratifications used. The time period for this study is 15 September 2015 through 15 January 2016.

## 2.1 FORECASTS

### 2.1.1 IPA-F

The output from the grid-based IPA-F algorithms includes calibrated icing probability, icing severity, and potential for supercooled large drops (SLD), including freezing drizzle and freezing rain. The methodology is based upon the logic included in the FIP that runs over the CONUS, references which can be found in McDonough et al. (2003), Brown and Bernstein (2006), and Wolff et al. (2008). IPA-F is produced from fields output from the Rapid Refresh (RAP) model. The RAP is on an 11.25-km grid (NCEP GRIB Grid 242, shown in Figure 2.1), with new forecasts produced hourly.

### 2.1.2 FORECAST ICING POTENTIAL (FIP)/FORECAST ICING SEVERITY (FIS)

The AAWU is presently running a FIP and FIS product for their domain of interest. This product is a predecessor of the IPA-F, based on the algorithm as it existed in 2005. In addition to being an older version of the algorithm, the FIP/FIS is also produced using the North American Mesoscale (NAM) model, rather than the RAP model. The NAM, as used in FIP/FIS, is on a 40-km grid (which is then mapped onto a 6-km grid), with new forecasts produced every 6 hours.

The spatial and temporal attributes of the IPA, FIP/FIS, and gridded AWUU forecasts are outlined in Table 2.1.

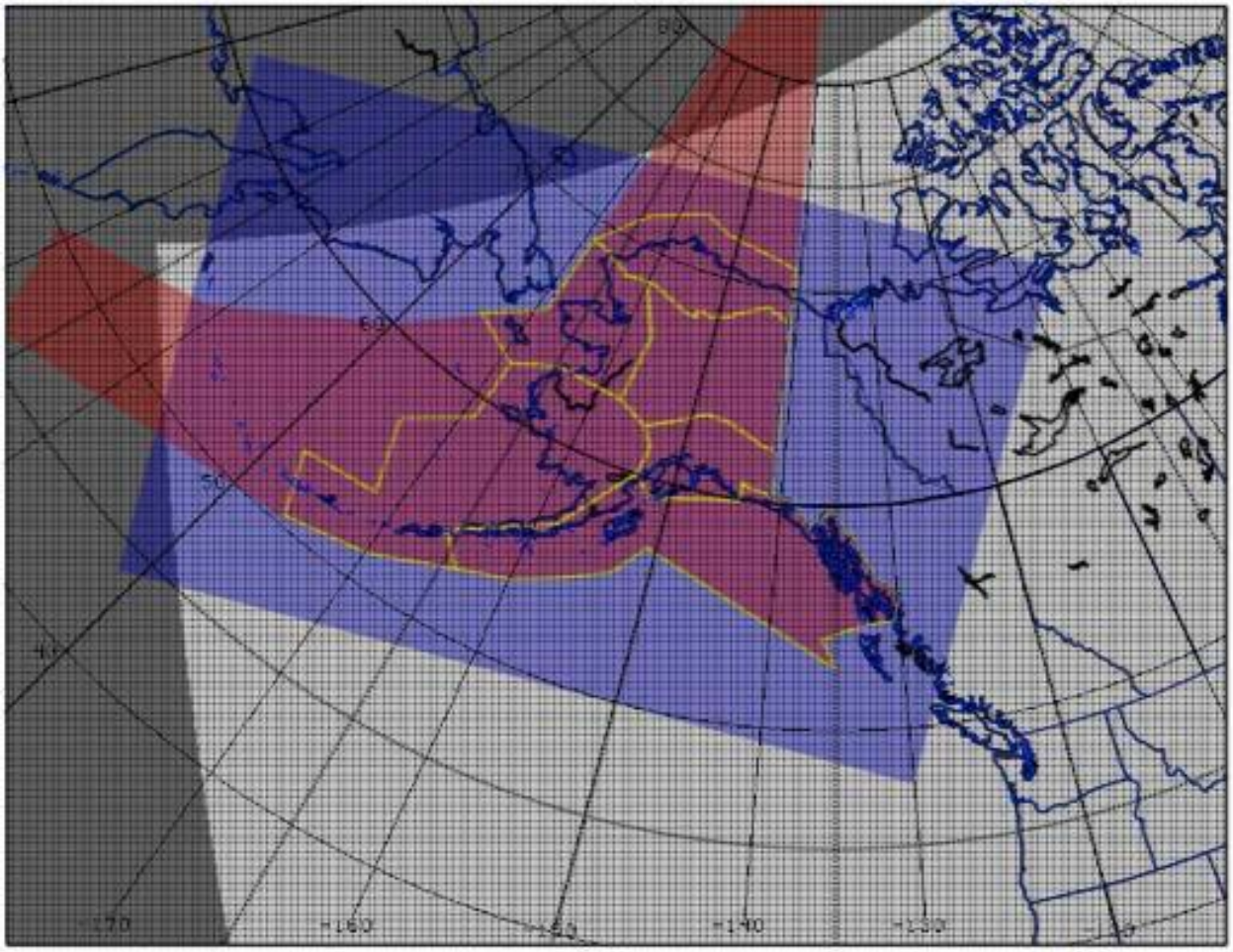

**Figure 2.1: From http://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html#GRID242, The domain of NCEP Grid 242. Area covered by the FIP/FIS as run by the AAWU is noted in blue. Area for which the AAWU produces gridded icing forecasts shown in red. The area of the 242 grid for which there is IPA-F is in white, with the grey being the area in Grid 242 for which there is no IPA-F data.**

**Table 2.1: Attributes of IPA-F, AAWU FIP/FIS, and the gridded AAWU forecasts.**

| | IPA-F | FIP/FIS | AAWU Gridded Forecasts |
|---|---|---|---|
| **Issues** | Every hour | 0000, 0600, 1200, and 1800 UTC | NA |
| **Probability and Severity Fields** | Probability of Icing<br><br>Icing Severity<br>• None/Trace<br>• Light<br>• Moderate<br>• Severe | Icing Potential (FIP) {10, 30, 50, 70, 90}; converted to probability; adjusted for lead time<br><br>Icing Severity (FIS) PDT/AAWU<br>0.000/0.000 = None/Trace<br>0.175/0.250 = Light<br>0.375/0.400 = Moderate<br>0.700/0.600 = Severe | Icing Severity<br>• Isolated Moderate Icing<br>• Occasional to Constant Moderate Icing<br>• Moderate with Isolated Severe Icing |
| **Supercooled Large Drops** | Present | Not present | Not present |
| **Leads** | 1, 2, 3,6,9,12, 15, 18 | 0, 3,6,9,12,15,18,21,24 | 0–11, 1–12, or 2–13 depending on issue (see Table 2.2) |
| **Horizontal Resolution** | 11.25km | 6km (from 40km) | 6km |
| **Altitudes** | 500–30,000 ft | 1,000–30,000 ft | Sfc-cloud top |
| **Vertical Resolution** | 500-ft increments | 1,000-ft increments at and below 6,000 ft<br>3,000-ft increments above 6,000 ft | 1,000-ft increments |

As noted previously, FIP/FIS provides an uncalibrated value of icing potential, between zero and one, as well as a severity value between zero and one. The conversion of FIS severity value to IPA category is done two ways, using two sets of thresholds. One set is from the IFI PDT, as applied to the IPA continuous severity output to produce the severity categories. The other set is from the AAWU, and correspond to FIS thresholds used in the AAWU forecast process. The relationship between potential and probability, as defined by the IFI PDT (Kuchera et al. 2007), is

$$\text{probability} = \text{potential} * (-0.033*\text{fcst\_lead\_seconds}/3600+0.84).$$

### 2.1.3 AAWU Gridded Icing Forecast

The AAWU icing forecast is issued every 8 hours to designate any areas of significant icing. The forecasts are issued at approximately 4:00 a.m., 12:00 p.m., and 8:00 p.m. local time each day (though

the exact time of issuance will vary from day to day, depending upon a number of factors impacting the forecasters that produce the forecast, Table 2.2). Areas of isolated moderate, occasional moderate, or isolated severe icing are displayed by shaded regions of yellow, orange, and red, respectively. Additionally, the base- and top-height values are provided. The AAWU icing forecast can be viewed at http://newaawu.arh.noaa.gov/index.php?tab=2&subtab=2 and an example is shown below. These grids support the generation of the AAWU AIRMET.

**Table 2.2: Expansion of gridded AAWU forecast issuances and leads converted to UTC**

| **05Z File** | | | |
|---|---|---|---|
| **Daylight Time: Issue time 8:00 p.m. AKDT = 04Z** | | **Standard Time: Issue time 8:00 p.m. AKST = 05Z** | |
| **Valid** | **Leads** | **Valid** | **Leads** |
| 05Z | 1–3 | 06Z | 1-3 |
| 08Z | 4-6 | 09Z | 4-6 |
| 11Z | 7-9 | 12Z | 7-9 |
| 14Z | 10-12 | 15Z | 10-12 |
| **13Z File** | | | |
| **Daylight Time: Issue time 4am AKDT = 12Z** | | **Standard Time: Issue time 4am AKST = 13Z** | |
| **Valid** | **Leads** | **Valid** | **Leads** |
| 14Z | 2–4 | 15Z | 2–4 |
| 17Z | 5–7 | 18Z | 5–7 |
| 20Z | 8–10 | 21Z | 8–10 |
| 23Z | 11–13 | 00Z | 11–13 |
| **21Z File** | | | |
| **Daylight Time: Issue time 12pm AKDT = 20Z** | | **Standard Time: Issue time 12pm AKST = 21Z** | |
| **Valid** | **Leads** | **Valid** | **Leads** |
| 20Z | 0–2 | 21Z | 0–2 |
| 23Z | 3–5 | 00Z | 3–5 |
| 02Z | 6–8 | 03Z | 6–8 |
| 05Z | 9–11 | 06Z | 9–11 |

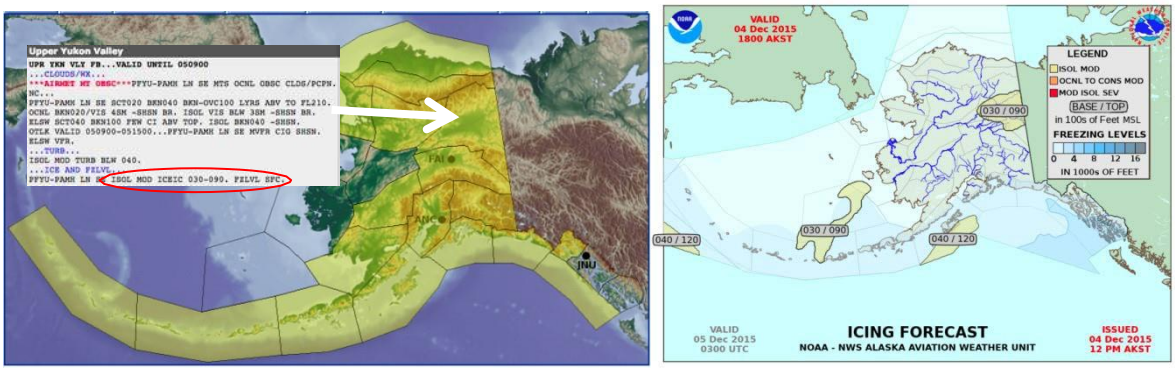


**Figure 2.2: Domain covered by AAWU AIRMETs (left) and a sample AAWU gridded icing forecast (right) both valid at 0300 UTC on 5 December 2015. AIRMETs are based on gridded forecasts from the forecast offices in Fairbanks (FAI), Anchorage (ANC), and Juneau (JNU).**

All forecast products will be evaluated on the IPA grid. The FIP/FIS data on the 6-km grid will be mapped onto the IPA grid using a weighted average of all 6-km gridboxes contained within an IPA (~11 km) gridbox. Vertical regridding will be done by linear interpolation between vertical levels for which a product exists.

## 2.2 OBSERVATIONS

### 2.2.1 VOICE PILOT REPORTS (PIREPS)

PIREPs are reported irregularly at the pilot's discretion and include a subjective assessment of many meteorological variables including the existence/absence of icing and a subjective measure of the icing intensity. Included in the icing reports are the location, altitude or range of altitudes, type of aircraft, air temperature, intensity, and type of icing (NWS 2007). The full range of intensity values will be used (listed below), as forecasts of 'moderate or greater' (MOG) imply the need for the full range. The 'clear' type is used to indicate the possibility of SLD.

*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 (over one hour).
2. Light: The rate of accumulation may create a problem if flight is prolonged in this environment (over one 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.

*Icing types*

1. Rime: Rough, milky, opaque ice formed by the instantaneous freezing of small supercooled water droplets.

2. Clear: A glossy, clear or translucent ice formed by the relatively slow freezing of large supercooled water droplets.
3. Mixed: This is a combination of rime and clear.

In Alaska, most data sources are sparser than over CONUS, and PIREPs are no exception. In the IPA-F domain in September through December of 2015, there were 5366 PIREPs that reported icing (Figure 2.3). For the full CONUS, there were 22618 icing PIREPs, or about four times the number that were within the IPA-F domain in the same period.

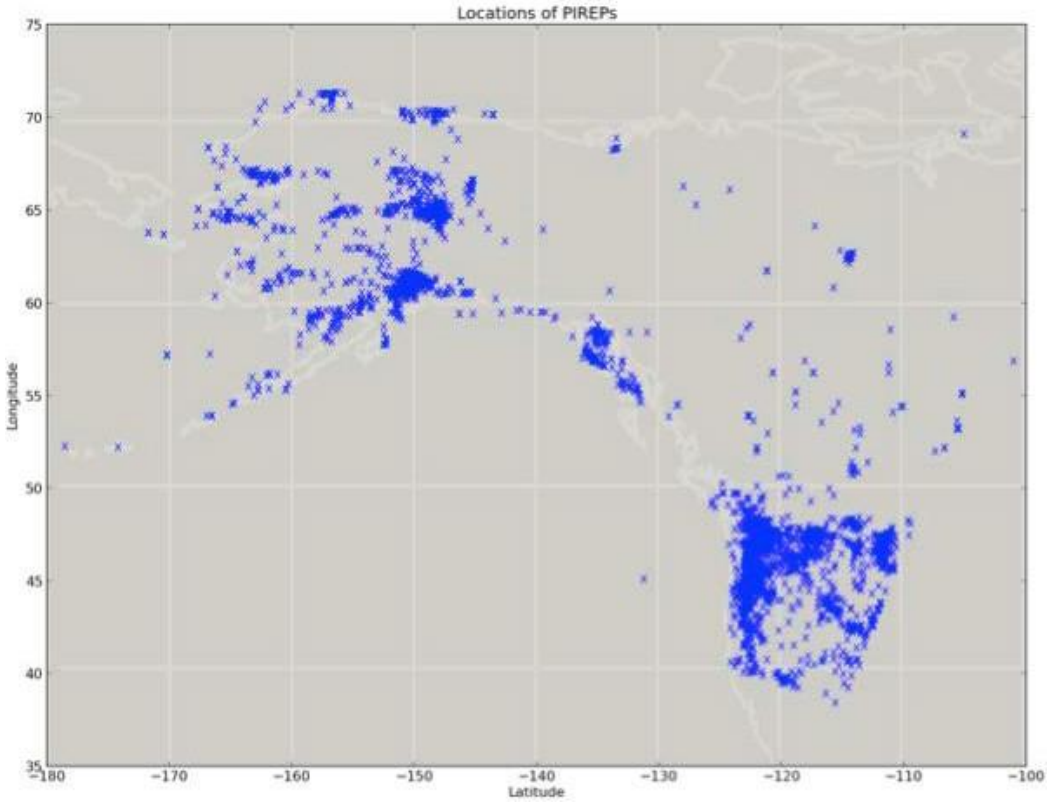


**Figure 2.3: Location of pilot reports for September – December 2015**

Note that we only use PIREPs that report icing (including "Null" severity). Though PIREPs that have no icing information (for example, turbulence reports) could be used as implicit "No-Icing" reports, they are not used this way in this assessment—only explicit icing PIREPs are used.

### 2.2.2 METAR Observations

Routine surface report (METAR) data are used to provide observations of icing conditions at the surface and to infer SLD events between the surface and the cloud ceiling. For instance, when freezing rain or freezing drizzle is recorded in the METAR, an SLD event is then inferred to exist between the surface and the cloud base (lowest cloud layer of at least "broken" coverage) (Madine 2008). This information is used to assess the quality of the IPA-F SLD parameter.

Similar to PIREPS, the density of METAR observations over the IPA-F region is less than that over the CONUS. As shown in Figure 2.4, in the IPA-F domain there are a total of 810 locations, whereas in the

CONUS domain there are 4147 such sites. Thus, the IPA-F domain has about 1/6 the number of METAR sites as the FIP domain, which is similar to the ratio of PIREPs within the IPA-F domain to those in the CONUS domain (1/4, as noted in the previous section).

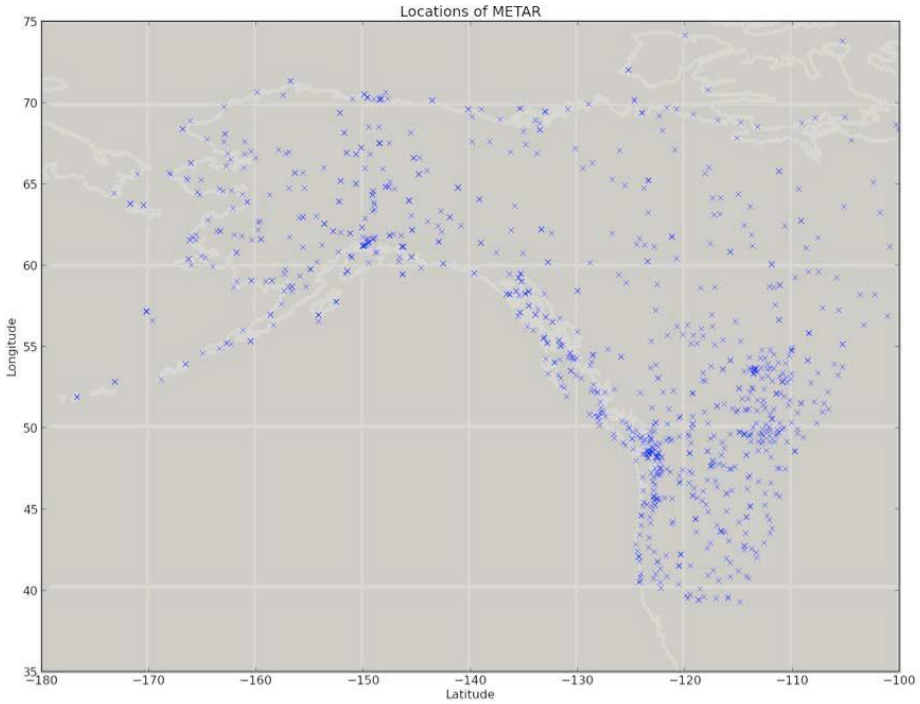


**Figure 2.4: As in Figure 2.3, but for METARs**

### 2.2.3 SOUNDINGS

Given the sparseness of in-situ observations in the IPA-F forecast area, soundings will be used for verification of icing conditions. Figure 2.5 shows the locations of sounding sites that are potentially within the IPA-F domain. There are 43 sites total, with 22 in the United States (13 of those in the state of Alaska), 13 in Canada, and 8 in Russia. Soundings are generally taken twice daily, at 0000 UTC and 1200 UTC, though at times there are special soundings at off-hours.

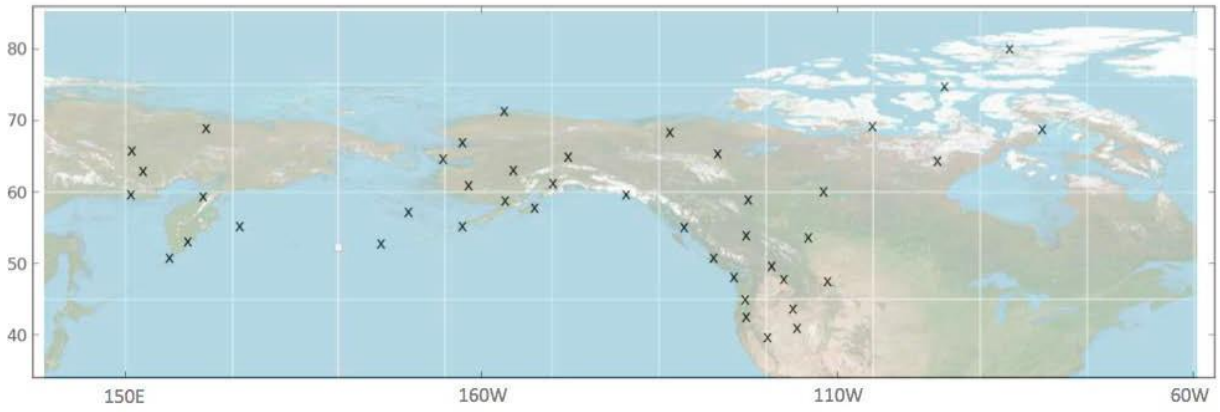


**Figure 2.5: Location of sounding sites within the IPA-F domain. Of the 43 sites in the IPA-F domain, 22 are in the United States, 13 in Canada, and 8 in Russia.**

Following the classification criteria in Schultz and Politovich (1992), and Bernstein et al. (2007), relative humidity and temperature from the sounding are combined to verify the IPA-F severity field. Details of this approach are in section 3.2 of this document.

### 2.2.4 SATELLITE DATA: CLOUDSAT/CALIPSO

Data products from the satellites CloudSat and CALIPSO combined with a temperature field provide a way to measure 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).

In previous studies, data from the radar from CloudSat and the lidar from CALIPSO have been combined to provide bounds on forecast performance, and it was planned to do so again in this assessment. However, two major problems in using this data for Alaska were revealed. Firstly, the orbital paths are such that there is a strong connection between the time of day and the area of the domain that is traversed. In addition, technical issues with CloudSat make that satellite inoperable at night. With the change of the assessment period from spring to autumn and winter, the available periods of sunlight diminish substantially. As a result, the combined CloudSat/CALIPSO data is obtainable for only a few particular swaths, each of which are tied to a particular hour of the day. Consequently, the satellite data are not used in this assessment.

## 2.3 STRATIFICATIONS

Results are stratified spatially, temporally, and according to certain icing intensity thresholds as follows.

### ICING PROBABILITY STRATIFICATIONS

As IPA-F is not yet included in the Aviation Digital Data Service, the probability stratifications used for IPA-F will be consistent with what is commonly used at the AAWU. For FIP/FIS, the severity values will be masked using the FIP thresholds of 10, 30, 50, 70, and 90, respectively. IPA-F icing severity is masked using the same thresholds, but after they are converted from potential to probability using equation 1, given in section 2.1. The FIS severity values will be mapped onto categories according to both the IFI PDT and AAWU thresholds shown in Table 2.3.

### SLD STRATIFICATIONS

Following the visualization of SLD in ADDS, potential values are grouped into three categories: < 0% (unknown), between 0% and 5%(no SLD), and ≥ 5% (SLD present).

### ALTITUDE BINS

Results are 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 |

### TEMPORAL STRATIFICATION

Forecast performance is stratified by forecast issue and lead times.

### INTENSITY STRATIFICATION

The majority of the focus of the evaluation is on the Moderate-or-Greater (MOG) level of icing severity, but all IPA-F categories are assessed. However, PIREPs, IPA, FIS, and gridded AAWU icing forecasts all use different measures of icing severity. Table 2.3 shows how intensity values are related between the three data sets.

**Table 2.3: Mappings of icing severity categories**

| (ADDS) PIREP | IPA-F Category | FIS (PDT) | FIS (AAWU) | AAWU-Gridded Icing Forecast |
|---|---|---|---|---|
| **Neg**<br>**Neg-Clr**<br>**Trace** | None/Trace | 0.00–0.174 | 0.00–0.24 | |
| **Trace-Light**<br>**Light** | Light | 0.175–0.374 | 0.25–0.39 | |
| **Light-Mod** | Moderate | 0.375–0.699 | 0.40–0.59 | Isolated Moderate Icing |
| **Moderate** | Moderate | 0.375–0.699 | 0.40–0.59 | Occasional to Constant Moderate Icing |
| **Mod-Severe**<br>**Severe** | Severe | 0.700–1.000 | 0.60–1.00 | Moderate with Isolated Severe Icing |

## GEOGRAPHIC STRATIFICATION

As noted in the response to the initial assessment, there is a desire to see results stratified by subregions of the Alaska area. Following McDonough et al. (2010), Alaska is divided into the following six subregions: North Slope, Kotzebue and Norton Sound, Interior, Southwest and Bering, South Central, and Southeast and Gulf Coast (Figure 2.6). Observations that fall outside these subregions are categorized simply as 'outside'.

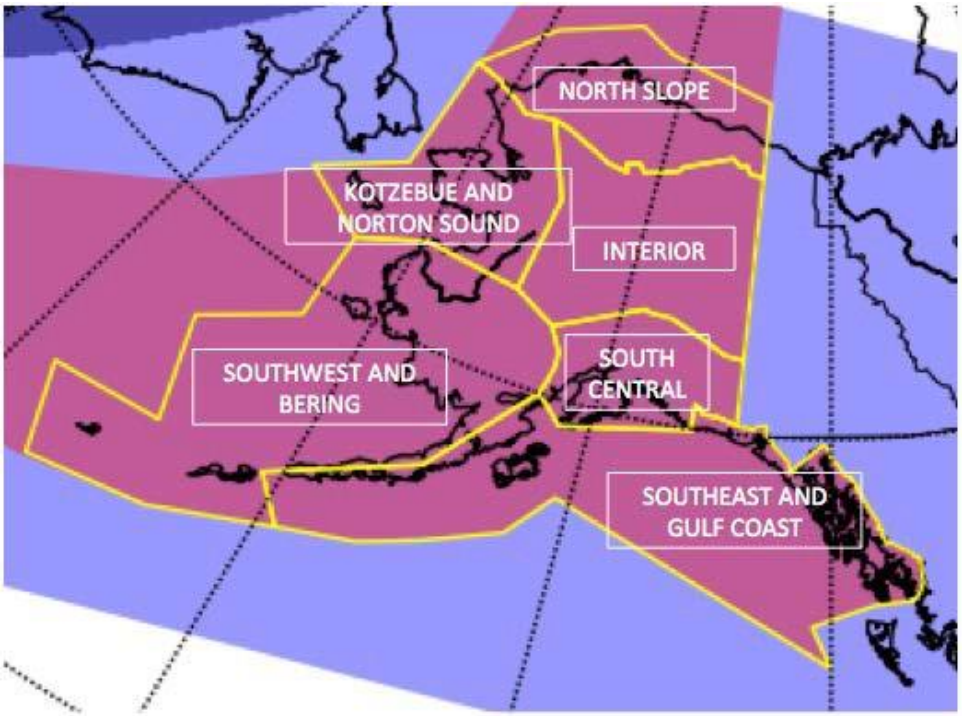


**Figure 2.6: The six geographic subregions used for stratification of results. Observations that are not located in any of these regions are classified as 'outside'.**

# 3 METHODS

A variety of verification approaches are employed in this assessment and are described in the following subsections. The mechanics of the assessment include: 1) an evaluation of field characteristics of IPA-F and FIP/FIS, 2) an evaluation of IPA-F forecast consistency, 3) a neighborhood-based approach for verifying probability/severity and SLD forecasts using PIREP observations, 4) a verification of probability/severity forecasts using sounding observations, 5) an evaluation of IPA-F compared to gridded AAWU icing forecast, 6) an evaluation of IPA-F as a supplement to AAWU gridded icing forecast, 7) a verification of SLD forecasts using METAR observations, and 8) an investigation of specific case studies identified by the AAWU.

To provide information on the performance of IPA-F relative to that of FIP/FIS, IPA-F probability and severity and FIP/FIS are compared in the area in which the two overlap. Event equalization is applied in all comparisons to account for the differences in issues and lead times of the two products, such that only data for dates, issues, and leads for which both products exist is included in the comparsion. Note that though FIP/FIS does produce an SLD field (SIP) the AAWU does not make use of this, and thus, there will be no comparisons of IPA-F SLD with any other SLD product.

Additionally, in all techniques that involve the use of temperature data, locations in which the temperature is greater than 5°C or less than -40°C will be excluded, as those locations should never have icing conditions, and thus are not appropriate for product assessment.

Table 3.1 contains terminology and score definitions for reference in the subsequent sections.

**Table 3.1 Terminology**

| | |
|---|---|
| **LOG:** | **Light or Greater** |
| **MOG:** | Moderate or Greater |
| **LTMod:** | Less Than Moderate Icing |
| **POD (= PODy):** | Proportion of all observed events that are correctly forecast to occur, in this case, of detecting icing at a specific threshold |
| **POFD (= 1 – PODn):** | Proportion of all observed nonevents 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) |

## 3.1 FIELD CHARACTERISTICS

### 3.1.1 FIELD DISTRIBUTIONS

The makeup of the IPA-F fields is evaluated using value-based distributions. Distributions are generated for each field: bins for IPA-F severity are generated per severity category, and the probability and SLD values are binned from 0 to 1 using a bin size of 0.01. FIP distributions will also be generated matching the IPA probability bin sizes and FIS values will be recalculated to severity

categories, as per Table 2.1, and binned similarly to IPA severity. Additionally, the probability and severity plots will contain ratio plots that better identify the differences between IPA-F and FIP/FIS.

### 3.1.2 Climatological Maps

Further evaluation of the characteristics of IPA-F will be performed using climatology maps, for IPA-F and FIP/FIS, as well as differences between IPA-F and FIP/FIS. Spatial distributions are derived by aggregating counts of IPA-F field values exceeding a threshold (e.g., 50 for potential, MOG for severity) over a date range, issue and lead times, and vertical layers as defined in section 2.3, for each grid point. For differences between IPA-F and FIP/FIS, the aggregation will be applied to the remapped FIP/FIS, and difference maps of the counts (IPA-FIP or IPA-FIS) are generated where positive values (blue) indicate a higher IPA-F count and negative values (red) indicate a higher FIP/FIS count.

## 3.2 Consistency

To assess the consistency of the IPA-F, auto-lag correlation skills are computed on forecasts of adjacent lead times to identify if there are any sudden changes in the correlation between forecasts at a particular lead time. For example,

- IPA-F 3-hour forecast is compared to the IPA-F 2-hour forecast valid at the same time
- IPA-F 2-hour forecast is compared to the IPA-F 1-hour forecast valid at the same time.

From this type of comparison, we expect to see similar hour-to-hour changes in IPA-F. What would be unexpected is if the difference between two consecutive hours, say the 4-hour and 3-hour forecasts, were significantly different than all other hour-to-hour changes.

## 3.3 Probability/Severity and SLD Verification using PIREPs

PIREPs are used to determine product skill in detecting IPA-F and FIP/FIS severity as well as SLD for IPA-F; however, 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 IPA-F product to PIREPs yields the following contingency table:

**Table 3.2 PIREP 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 equals or exceeds a given threshold, and 'no' signifies that the forecast or observed value is less than the threshold. POD, POFD, and PSS are computed from the contingency table.

### 3.3.1 Forecast and Observation Pairing Techniques

To enable forecast comparisons and evaluation of quality, forecasts and observations are matched spatially and temporally using the mechanics described in the following subsections. The FIS severity field is masked using FIP potential values of 10, 30, 50, 70, and 90, respectively. In all techniques, the IPA-F severity field is masked using those same potential values, after being converted into probability values.

#### 3.3.1.1 IPA-F Probability/ Severity, FIP/FIS, and AAWU-Gridded Icing Forecast to PIREP Severity

It is known that PIREPs have location errors. Pearson and Sharman (2013) reported that the median horizontal error of turbulence PIREPs was 35 km, and the average vertical error 20 ft. To account for PIREP location uncertainty, a neighborhood approach using forecast grid points around the location of the PIREP is used to match the forecast to the PIREP. A horizontal neighborhood, consisting of all grid points contained within a circle of radius approximately 35 km centered at the grid point closest to the PIREP location (Figure 3.1), is included at each flight level. The severity value within the neighborhood that best matches the PIREP intensity is taken as the associated forecast value. If there is not a perfect match (an IPA-F intensity directly matching the PIREP intensity), the closest match is determined by first identifying the nearest intensity greater than that observed, then, if no forecast intensities greater than observed are found, the nearest intensity less than that observed is identified. The choice to use the exact match followed by highest intensities first was done because high intensities have impact and are thus considered more important. Note that Figure 3.1 is only an example of how a neighborhood could be made; the variable size of IPA grid points on the NCEP 242 Grid, ranging from ~9.5 km to ~12.5 km, implies that some 35-km radius neighborhoods will have a greater number of IPA grid points than others.

The selection of neighborhood size is a balance between accounting for PIREP location uncertainty and weakly representing the resolution of information for a product. The choice of 35 km is based on the median error, and as such it is expected that only half of the PIREPs will actually be in their respective neighborhood. However, enlarging the neighborhood, combined with the use of the "best match" approach for selecting the grid value to compare to a PIREP, could increase the occurrences where the forecast is falsely matched to a PIREP and awarded with a hit.

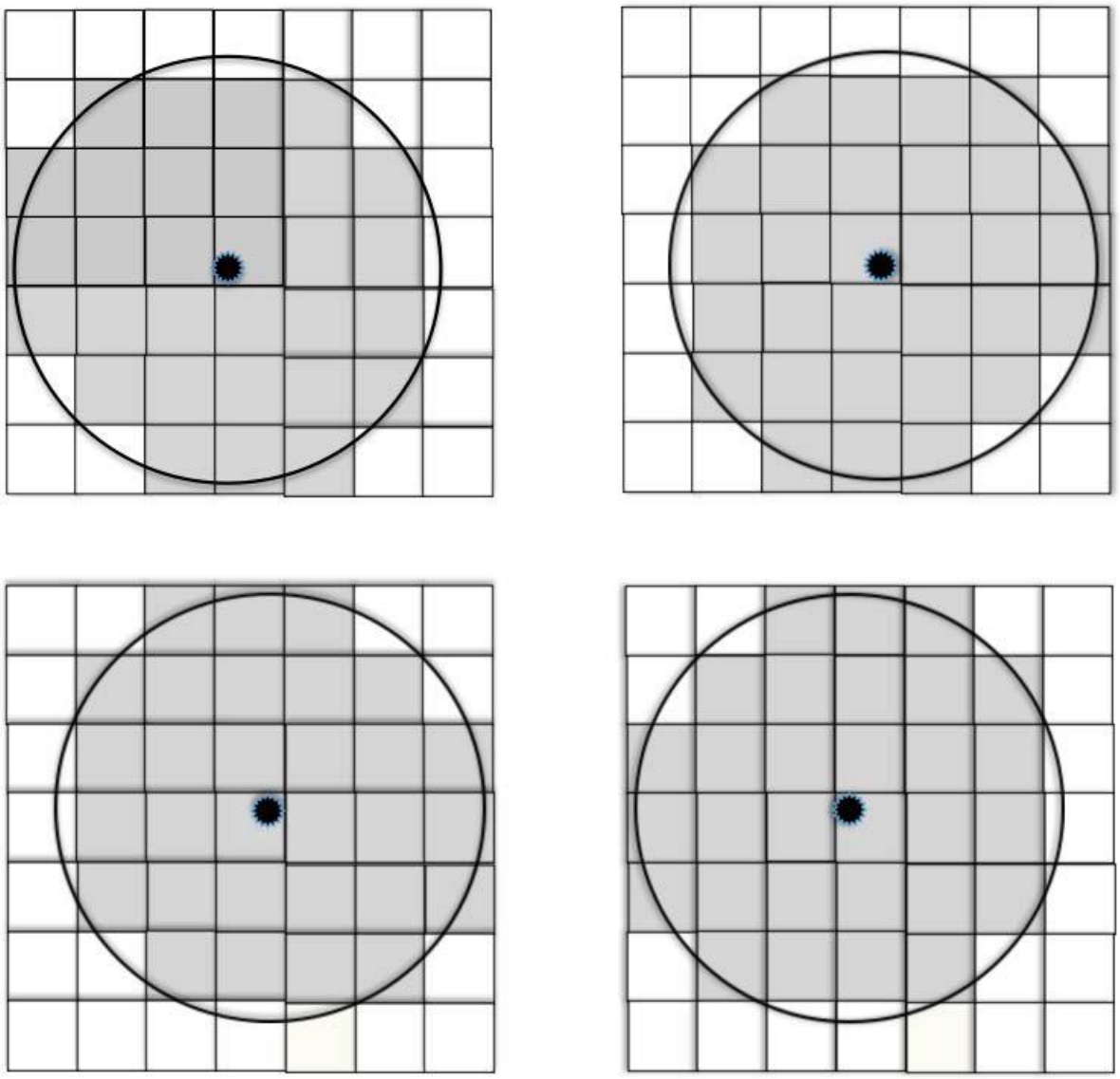

**Figure 3.1: A schematic of four possible IPA-F (35-km radius) neighborhoods surrounding a PIREP. Note that the actual number of grid points incorporated will vary.**

For PIREPs in which only a single level of icing is reported, the neighborhood is defined by including the grid points within 35 km of the PIREP at the IPA-F level closest to the PIREP flight level, along with the level above and below (+/- 500 feet), resulting in three vertical levels around the PIREP (Figure 3.2, left). Though this is a greater vertical extent than the value from Pearson and Sharman (2013), that work investigated turbulence, not icing, PIREPs. Given that icing is more likely on ascent/descent than at cruise altitude, we expect greater vertical uncertainty in icing PIREPs than in turbulence PIREPS.

For PIREPs in which a top and base of icing is reported, the neighborhood is defined by including the IPA-F grid points within 35 km of the PIREP from the IPA-F level above the level closest to the PIREP's reported 'icing top' to the IPA-F level below the level closest to the PIREP's reported 'icing base'

(Figure 3.2, center). For PIREPs that report no icing over a layer, in which the top layer is 'unlimited', the neighborhood extends from one layer above the icing base layer to the top layer of IPA (Figure 3.2, right).

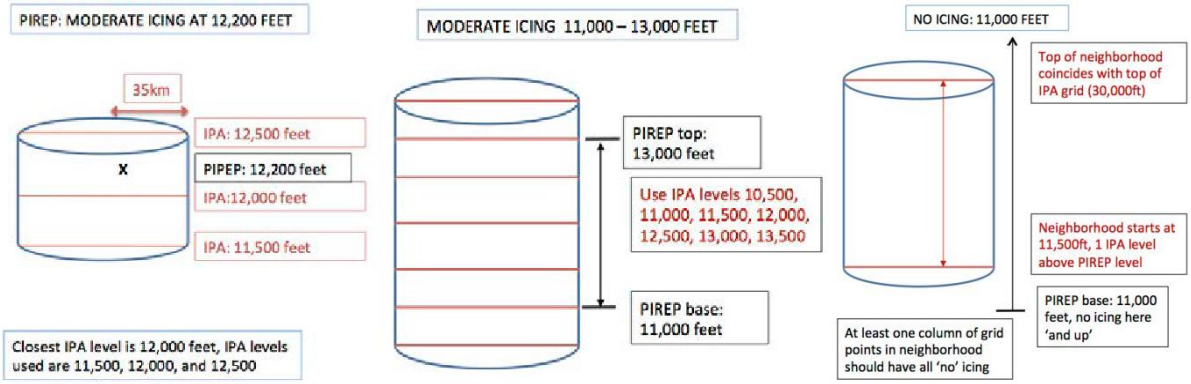


**Figure 3.2: A schematic of three possible neighborhoods surrounding a PIREP: The neighborhood surrounding a report for a single vertical level (left), the neighborhood for a report with a top and base for the icing conditions (center), and or no icing at and above a given altitude (right).**

On the boundaries of the IPA-F grid, the subset of points available in the neighborhood is used for a best match. Grid points located below the model surface elevation are also excluded.

For temporal matching, all PIREPs within a time window of [-30, 30) minutes around the forecast valid time are used for verification.

Given that FIP/FIS and the AAWU Gridded Icing Forecasts are mapped onto the IPA grid, comparison of icing PIREPs to those two products shall be done in the same manner as comparisons of PIREPs to IPA-F.

#### 3.3.1.2 IPA-F SLD TO SLD PIREPS

A PIREP is considered an observation of SLD conditions if it indicates severe icing intensity and clear icing type. PIREPS of all other types and intensities are considered negative reports of SLD, with one exception: PIREPs include a 30-character "Weather String" that provides information on the weather observed. In this string, it is possible to have reports of freezing rain or freezing drizzle. If such a report exists within the Weather String, that PIREP is considered to be a positive report of SLD, regardless of what icing type and severity was reported.

PIREPs are matched to the IPA-F forecast grid using the same spatial and temporal neighborhoods as for severity. The SLD forecast field is categorized into 'yes,' 'no,' and 'unknown' using the thresholds indicated in the in section 2.3.

As the SLD field known as SIP, part of the FIP/FIS being run at the AAWU, is not used by the AAWU, there will be no comparisons of the SIP nor the gridded AAWU icing product, to SLD PIREPs.

## 3.4 PROBABILITY/SEVERITY VERIFICATION USING SOUNDINGS

In the classification criteria of Schultz and Politovich (1992), icing is considered possible for layers of a sounding in which the temperature is between 0°C and -20°C and relative humidity is greater than 50%. Layers that are outside of those bounds are considered areas in which icing is highly unlikely. In this assessment, the bottom end of the icing-is-possible area is extended from -20 to - 25 °C to be consistent with that used for other datasets (namely satellite); this layer is referred to as "extended Class 1". Layers with temperatures warmer than 0°C, colder than -25 °C, or drier than 50% RH are layers where icing is **not** expected.

There is a second, more stringent categorization called "Class 2", in which temperatures are between -2°C and -15°C, and the relative humidity is greater than 65%. These "Class 2" criteria were developed for model soundings and are not suitable for use with real data. Bernstein et al. (2007) indicates that 87% relative humidity is a better indicator of icing for observed soundings. Thus, layers that fit the criteria of relative humidity greater than 87% and temperatures between -2°C and -15°C are considered layers in which icing is expected. These criteria are used for verification of the occurrence of icing in the corresponding layers at each of the sounding locations within the domain (Figure 3.3).

The IPA-F grid box that contains the sounding site is compared to the sounding (as such, the drift of the sonde during ascent is not accounted for) from the lowest vertical level up to 30,000 ft. For sounding layers that indicate no icing (T>0°C and T<-25°C, RH<50%; outside of extended Class 1), it is expected that all grid boxes above the site in the no-icing layers will be free of icing. For the layers that are in the "Icing Likely" category (temperature in the range from -2 °C to -15 °C, relative humidity greater than 87%; Class 2; e.g., Figure 3.3), at least one of the vertical levels in the column of IPA-F grid boxes in the chosen layer should contain icing (effectively, a 'best match' approach).

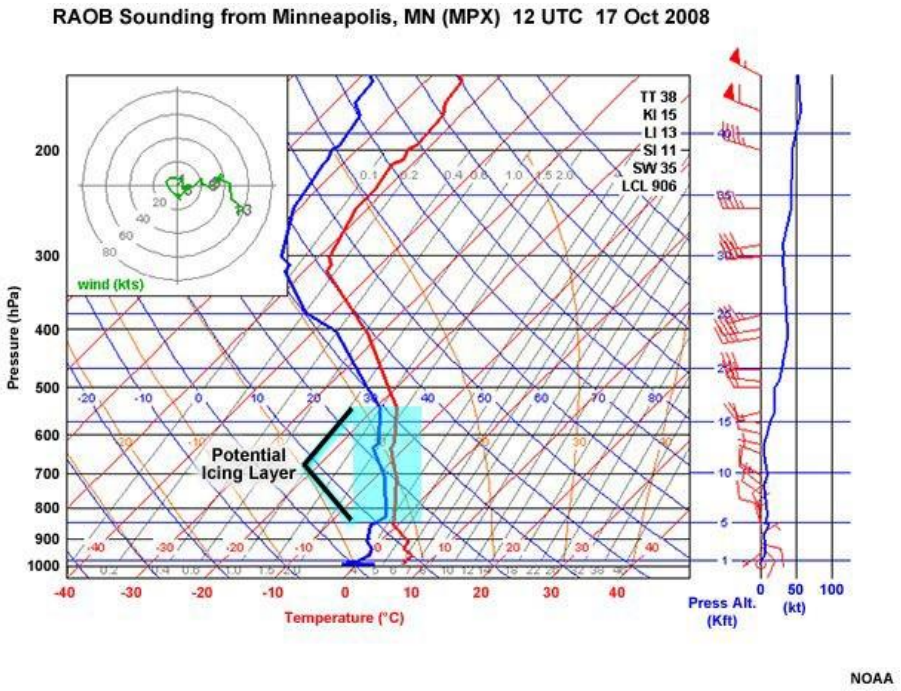


**Figure 3.3: Example of the sounding verification methodology. The levels within the blue box meet the 'icing-likely' criteria.**

The extended Class 1 and the Class 2 fields of Schultz and Politovich (1992) provide an upper and lower bound, respectively, for areas in which icing is possible. The extended Class 1 area (temperature between 0°C and -25°C, relative humidity greater than 50%) defines the regions where IPA-F and FIP/FIS could identify icing potential; the area outside the Class 1 area defines the regions where IPA-F and FIP/FIS should not identify icing potential. The adjusted-Class 2 area, (temperature between -2°C and -15°C, relative humidity greater than 87%) is where icing conditions should be found. Statistics are computed for IPA-F MOG severity using the five potential masks (converted to probability, as per equation 1 in section 2.1). POD is computed for the Class 2 region, and POFD is computed for the area outside the extended Class 1 area, which represents the fraction of grid cells for which IPA-F or FIP/FIS (incorrectly) identifies icing potential. Because these scores are computed over different parts of the domain, they will not be combined to compute PSS.

## 3.5 IPA-F Compared to AAWU-Gridded Icing Forecast

When comparing IPA-F and gridded AAWU icing forecast fields, the discrepancy between the valid times of these two forecasts (instantaneous vs. valid period, respectively) will be addressed. The AAWU product will be treated as if it were issued at a specific time and the valid period will be separated into multiple leads. For example, the forecast issued at 4 a.m. AKDT (12Z) with a valid time of 14Z will be evaluated as a forecast issued at 12Z with leads 2, 3, and 4 valid at 14Z, 15Z, and 16Z.

As noted in Table 2.1 and Table 2.3, there are three categories in the gridded AAWU icing forecasts:

- Isolated Moderate Icing
- Occasional to Constant Moderate Icing
- Moderate with Isolated Severe Icing

The 'isolated moderate icing' category will be considered as a light forecast in order to be consistent with its use at the AAWU. The areas of 'occasional to constant moderate icing' and 'moderate with isolated severe icing' represent AAWU MOG forecasts. PIREPs of LOG intensity will be used to evaluate the Isolated Moderate Icing category and PIREPs of MOG intensity will be used to evaluate the AAWU categories considered as moderate. Ideally, we expect to see MOG PIREPs within these volumes, and not find MOG PIREPs outside of them.

The same caveats for PIREPs (due to their non-systematic nature) listed in section 3.3.1.1 also hold for gridded AAWU icing forecast comparisons. The values of POD, POFD, and forecast volume from IPA-F and gridded AAWU icing forecast are compared. In addition to the MOG IPA-F thresholds, thresholds at other intensities will also be included in the comparison to assess how other IPA-F intensities perform in comparison to the gridded AAWU icing forecast.

The gridded AAWU icing forecast gives volumes where icing is to occur. Therefore, the contingency table is defined in Table 3.3, where the gridded AAWU icing forecast area is taken to mean the AAWU MOG forecasts, as defined above.

**Table 3.3 Gridded AAWU contingency table**

| **Hit:** | **MOG PIREP inside an AAWU gridded icing forecast area** |
|---|---|
| **False alarm:** | Less Than Moderate PIREP inside an gridded AAWU icing forecast area |
| **Miss:** | MOG PIREP outside an AAWU gridded icing forecast area |
| **Correct no:** | Less Than Moderate PIREP outside an gridded AAWU icing forecast area |

The gridded AAWU icing forecast contingency-table statistics POD, POFD, and PSS are then compared to the IPA-F contingency-table statistics as determined above, using the IPA-F issuances 1 and 3 hours prior to the AAWU issue.

## 3.6 IPA-F AS A SUPPLEMENT TO AAWU-GRIDDED ICING FORECAST

In this study we provide a complementary view of IPA-F performance by considering its contribution as a supplement to gridded AAWU icing forecasts. Inside a gridded AAWU icing forecast area, where icing is predicted, IPA-F disagreement can potentially lower false-alarm rates by reducing forecast volume. Outside a gridded AAWU icing forecast area, where icing is thus not predicted, IPA-F disagreement can potentially reduce the likelihood of encountering an unexpected icing event without drastically increasing forecast volume.

When comparing the IPA-F forecast to a gridded icing forecast, the same method for accounting for the valid time of the gridded icing forecast used in section 3.3.1.1 will be used here.

POD, POFD, and PSS will be computed separately for the cases when the IPA-F field is inside a gridded icing forecast area and when it is outside.

As mentioned in section 3.3.1.1, when making comparisons to PIREPs, the neighborhood approach is used for the IPA-F algorithms, but in comparing gridded AAWU icing forecast to PIREPs, the 'in or out' metric described previously (is a MOG PIREP inside or outside of the AAWU AIRMET volume) is used. The higher two AAWU gridded icing forecast categories are again taken to be a 'yes' for MOG icing.

## 3.7 SLD VERIFICATION USING METARS

METARs are included as an observation set for verification of SLD. Positive SLD observations are established using reports of freezing rain (FZRA) or freezing drizzle (FZDZ) that also reported a cloud layer of at least "broken". The ceiling value from the METAR is used to estimate the depth of the observed SLD layer, with the ceiling value identified as the top of the SLD layer and the ground being the bottom. A METAR is considered a report of negative SLD if it reports clear skies or snowfall. For METARs that indicate SLD, SLD is assumed present from the ground to cloud base (Figure 3.4). For METARs that indicate no-SLD, the observation is assumed valid from the ground to either cloud base if the METAR indicates snow, or to 30,000 ft if the METAR indicates clear skies.

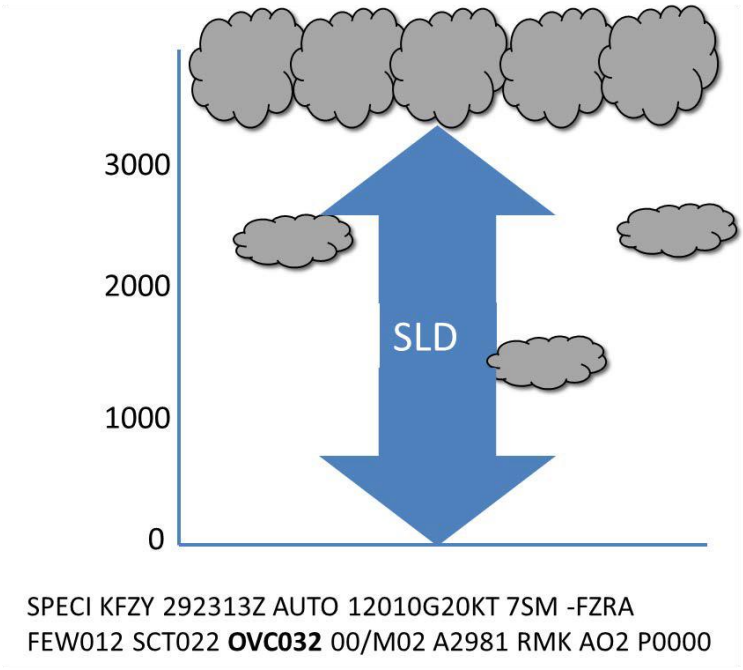


**Figure 3.4: Schematic representing verification approach using METAR reports.**

The grid box that contains the METAR location, from the lowest vertical level up to the chosen top level, is compared to the METAR report. For METARs that indicate SLD, at least one of the vertical levels in the column of forecast grid boxes above the METAR site is expected to contain SLD. For the METARs that indicate no SLD, it is expected that all grid boxes above the site, up to the chosen top, will not contain SLD.

## 3.8 Case Studies

Case studies for specific events as identified by the AAWU are also considered. Analysis includes an investigation of the different product behaviors for specific events, and possible reasons for differences, such as product inputs.

# 4 Results

## 4.1 Product Characteristics

Product characteristics are investigated using the techniques described in section 3.1. Characteristics for various stratifications are described in the subsequent sections.

### 4.1.1 Probability/Potential

#### 4.1.1.1 As a Function of Layer

Figure 4.1 contains climatological maps highlighting counts where IPA-F and FIP are greater than 50 potential for both the Near-Surface (left) and Low layer (right). In general, both products show a consistent signal in the Near Surface layer: high counts that are well distributed across the Bering Sea, spreading into the North Pacific and Gulf of Alaska. However, there are some subtle difference between FIP and IPA-F. FIP has several focused areas of higher counts (for example, on the south side of the Brooks Range and just north of Anchorage) that IPA lacks, and the FIP maximum counts in the Bering Sea are shifted slightly further north compared to IPA. The Near Surface difference map highlights that IPA tends to have more everywhere, except central Alaska.

On the other hand, the Low layer generally contains less over the oceans and has a high concentration over the southern coast of Alaska and down along the Canadian coast. While the regional pattern shifts, the maximum counts for the Low and Near-Surface layers are comparable for FIP (~9000) but increase for IPA-F (from 12610 to 16030). One notable difference between FIP and IPA is the finer detail available in the IPA. Additionally, the difference map highlights that, again, IPA tends to have more high probabilities compared to FIP.

In general, the frequency of icing decreases in moving to the Middle and High layers (not shown). FIP/IPA-F differences in the Middle layer look similar to the Low layer with slightly lower maximum counts. Little difference is seen when comparing IPA-F and FIP in the High layer as there is very little icing in this layer.

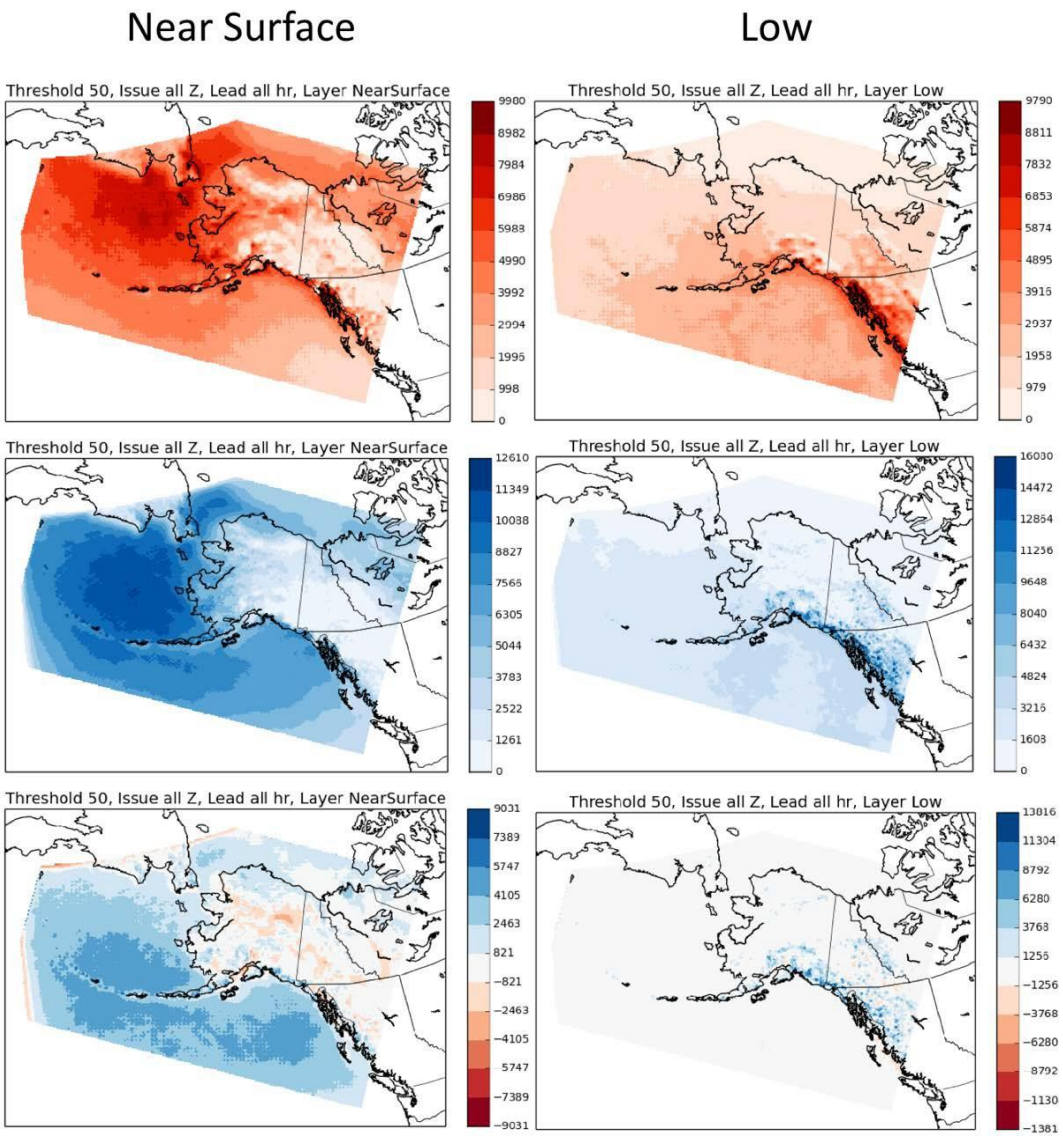


**Figure 4.1 FIP (top row) and IPA-F (middle row) climatological maps of probability greater than or equal to 50 potential for the Near-Surface (left column) and Low layer (right column) in the IPA-F and FIP/FIS overlap domain. IPA-FIP difference maps are shown for the respective layers (bottom row). In the difference maps, blue colors correspond to where IPA-F is greater than FIP and red colors correspond to where FIP is greater than IPA.**

#### 4.1.1.2 As a Function of Month

Figure 4.2 and Figure 4.3 show the FIP and IPA-F monthly climatological maps of counts for potential greater than or equal to 70 for the Near-Surface layer. These plots show a general tendency for high counts to shift southward from September to December. Both IPA and FIP highlight the Arctic and North Slope in September, shifting to the Bering Sea in October. The products diverge in November, as IPA is centered on the Aleutians and FIP remains strong in the Bering Sea. Both place a greater concentration of higher potential over the Kuskokwin Mountains, but IPA tends to have higher counts in the Gulf of Alaska. Also note, IPA contains a more diffuse area of higher potential west of Anchorage and near the Inside Passage.

FIP:

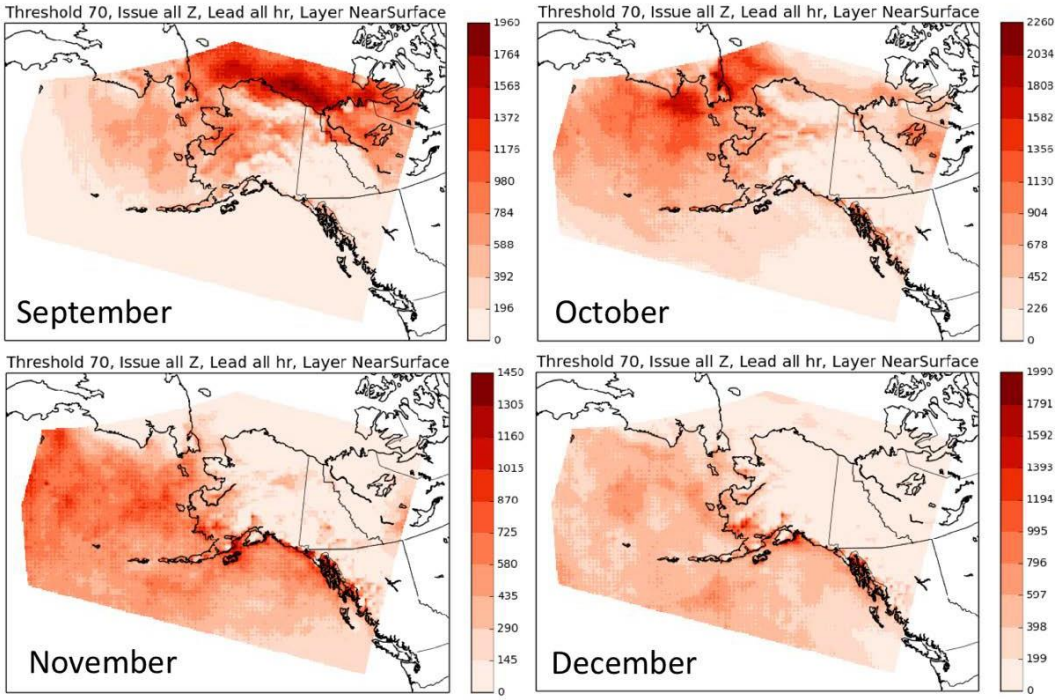


**Figure 4.2 Counts of FIP potential greater than or equal to 70 in the Near-Surface layer, stratified by month.**

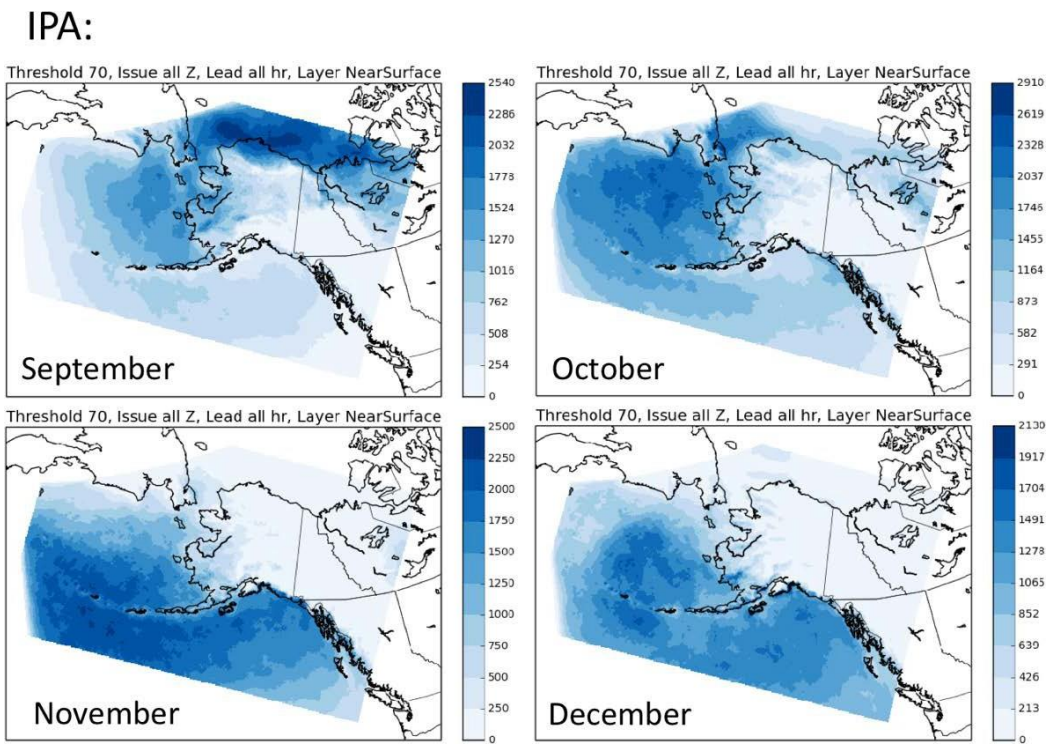


**Figure 4.3 Same as Figure 4.2, but for IPA-F.**

#### 4.1.1.3 As a Function of Threshold

Figure 4.4 displays difference maps for the 10, 30, 50, and 70 thresholds of potential in the Near-Surface layer. It shows that FIP has more low potential (red regions; see the potential-mask of 10 and 30), except along the Alaska Range, the Brooks Range, MacKenzie Mountains, and Gulf of Alaska. On the other hand, there is more high potential in IPA-F (blue regions; see the potential mask of 50 and 70), except on the south side of the Brooks Range and into the Yukon Territory.

Figure 4.5 shows similar plots but for the Low layer. Overall, the pattern is similar, but there is more low potential in FIP nearly everywhere, and there is more high potential in IPA-F confined to Southcentral, Southeast, and British Columbia regions.

This phenomenon can also be seen in the distribution plots. Figure 4.6 show the distributions of FIP and IPA probabilities (top and middle, respectively), and a corresponding ratio of the two (bottom). The negative ratio values for low probabilities and positive ratio values for high probabilities support the results of the climatological maps. Note that while the absolute value of the ratio is smaller for the low probabilities compared to the high probabilities, the actual number of occurrences for the lower probabilities is greater; hence, the higher counts for the 10 and 30 thresholds of potential in Figure 4.4 and 4.5. Interestingly, the distribution plots also show that IPA-F has more zero probabilities when compared to FIP, as indicated by the lower count in non-zero valued grid points.

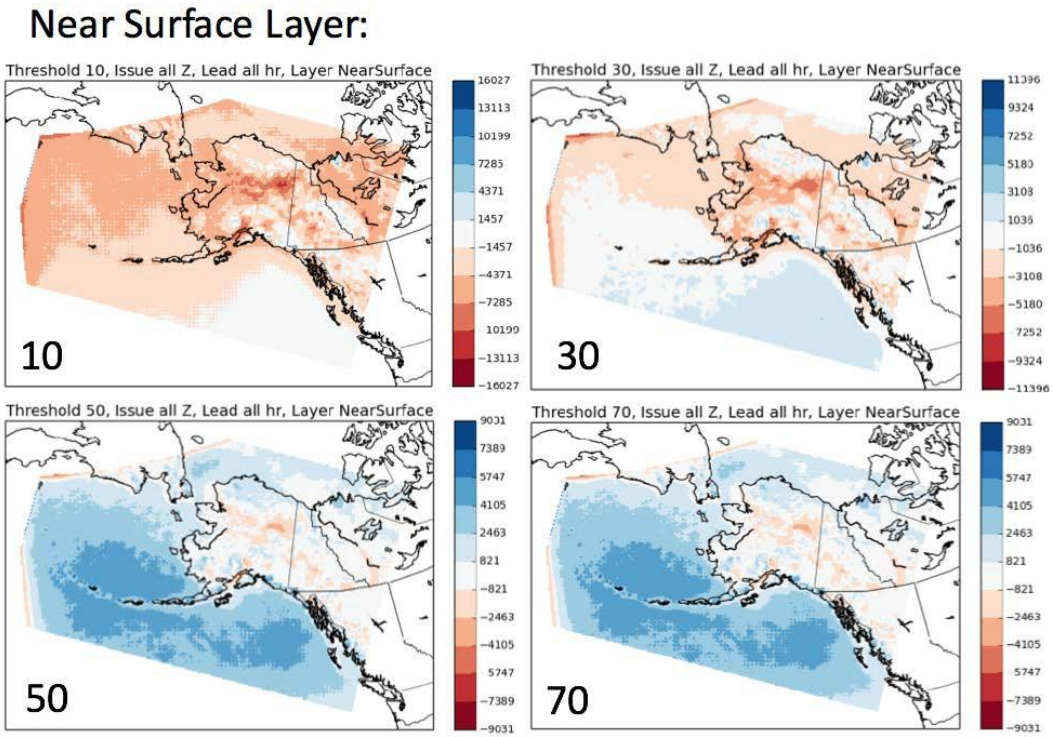


**Figure 4.4 Difference maps (IPA-FIP) showing the Near-Surface layer for various potential thresholds. In the difference maps, blue colors correspond to where IPA-F is greater than FIP and red colors correspond to where FIP is greater than IPA-F.**

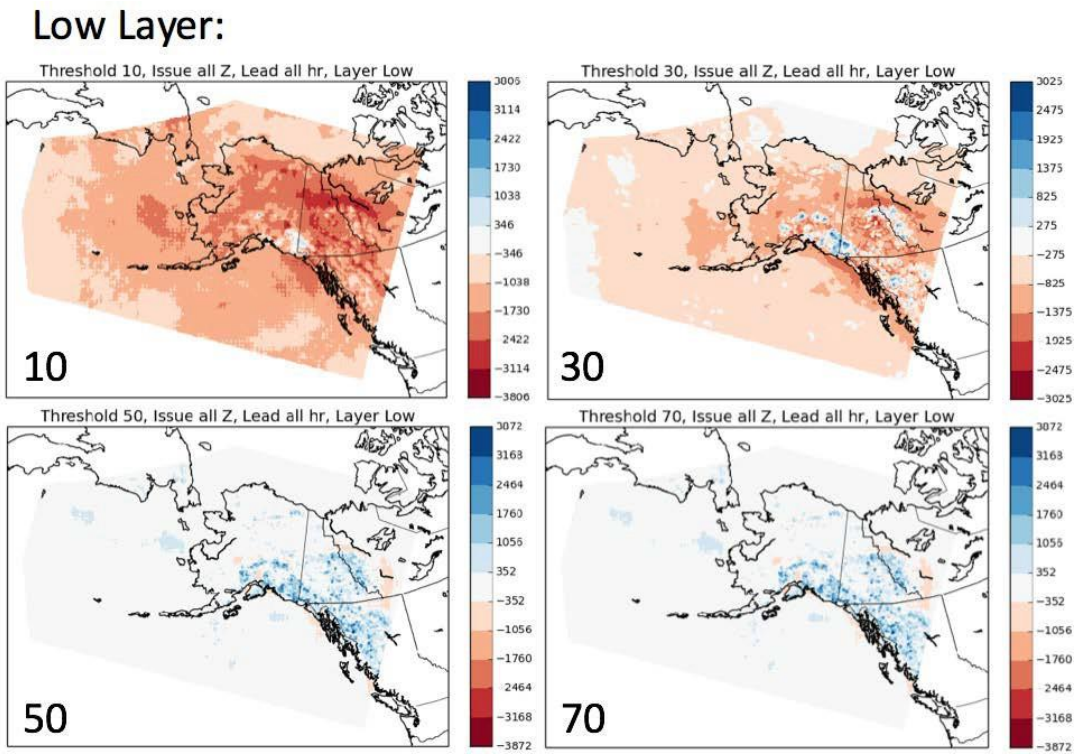


**Figure 4.5 Same as Figure 4.4, but for the Low layer.**

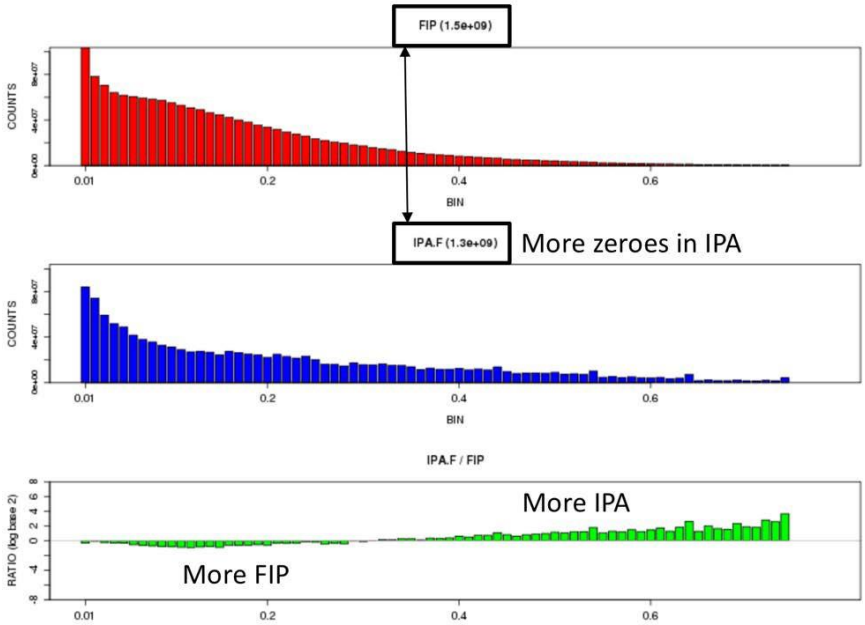


**Figure 4.6 Probability distribution for FIP (top), IPA-F (middle), and a ratio plot (bottom) for the Near-Surface layer.**

#### 4.1.1.4 IPA Full vs Overlap Domain

Section 4.1.1.1 noted that the occurrence of high icing potential decreases in the Middle and High layers, when looking at the IPA/FIP overlap domain. Interestingly, the full IPA-F domain contains a relative maximum outside the overlap domain in the Middle and High layers. Figure 4.7 demonstrates that the full IPA-F domain contains a higher contribution from lower latitudes. Nonetheless, the maximum counts are less than what is seen in the Low and Near-Surface layers.

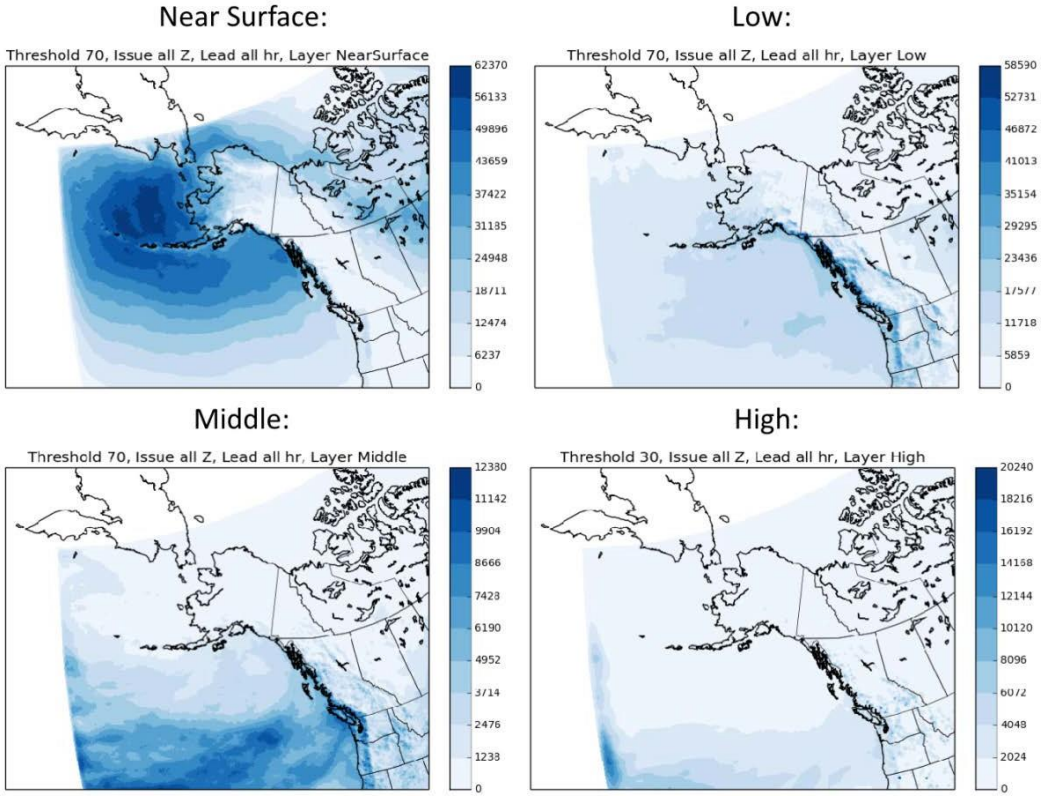


**Figure 4.7 IPA-F full domain climatological maps of probability greater than or equal to a potential of 70 FOR each layer.**

### 4.1.2 Severity

The FIS is investigated using both the PDT and AAWU severity thresholds. The climatological maps show a similar overall pattern for the AAWU and PDT thresholds, just slightly different counts that reflect the differences in Table 2.1. For conciseness, only climatological maps using PDT thresholds are shown, while the supporting text contains information regarding the signatures of the AAWU thresholds.

#### 4.1.2.1 As a Function of Layer

Figure 4.8 shows a comparison of MOG icing in Near-Surface and Low layers for FIS PDT and IPA-F. FIS PDT Near-Surface MOG is more prevalent over water, with splotches of high MOG counts over a large region. The AAWU threshold (not shown) looks similar but with lower counts given its higher moderate threshold (0.4 for AAWU vs. 0.375 for PDT). Overall, there is a ~10% increase in peak occurrence count in FIS using PDT thresholds as compared to the AAWU thresholds.

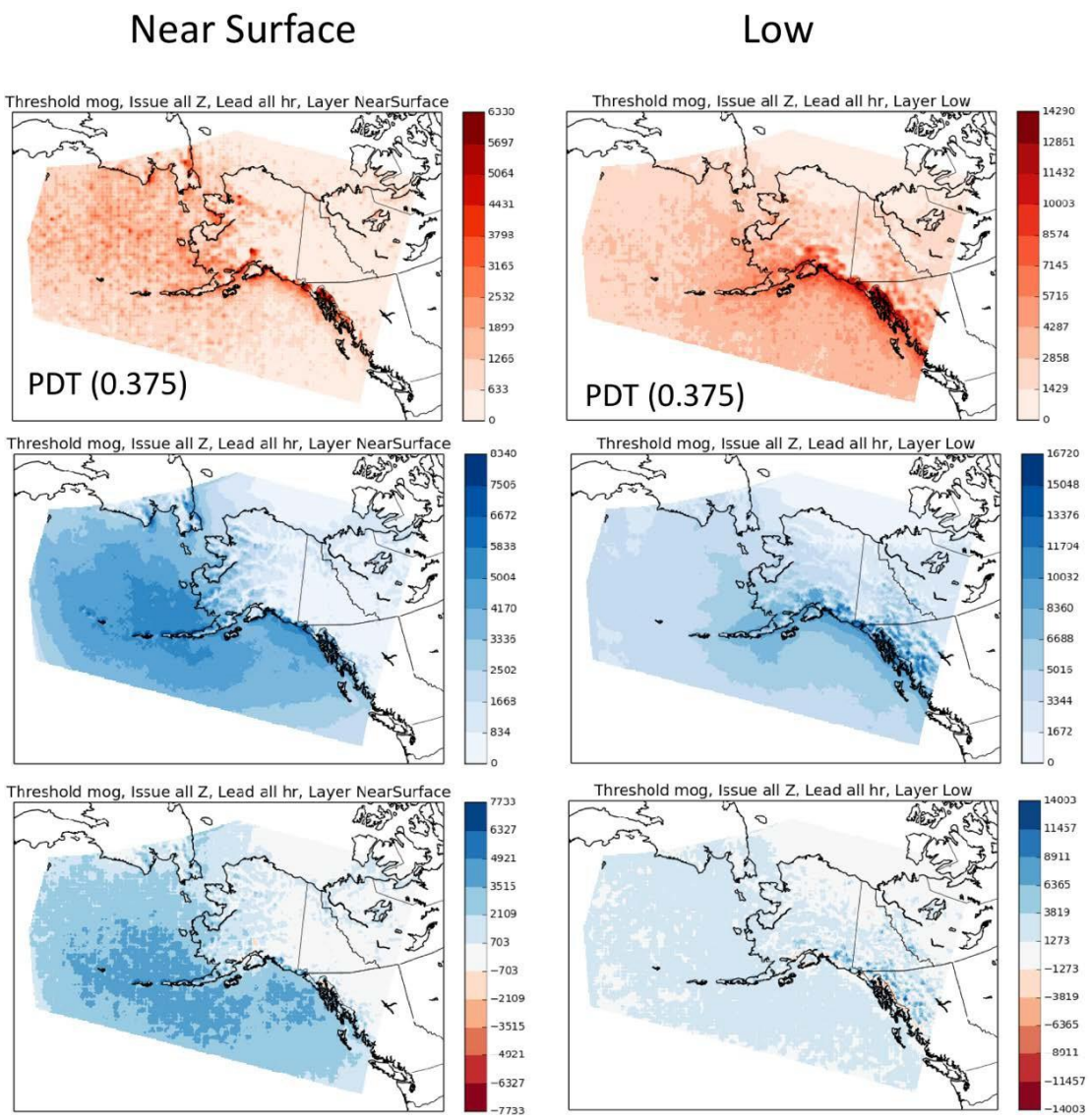


**Figure 4.8 Same as Figure 4.1, but comparing MOG icing.**

On the other hand, IPA-F Near-Surface MOG is concentrated over the Aleutian Islands and the Gulf of Alaska. Interestingly, the highest occurrence frequency of IPA-F MOG is located slightly further south compared to the maximum counts of potential greater than 50 (see Figure 4.1).

Moving to the Low layer, it is evident that the spatial pattern behaves similarly to that of the potential, where both FIS and FIP focus MOG on the southern coast of Alaska and down along the Canadian coast (see Figure 4.1). Notice that the maximum counts in the Low layer are double those in the Near-Surface for both the FIS and IPA-F. For the Middle layer (not shown), the maximum counts are similar to the Near-Surface layer, and decrease substantially in the High layer.

Overall, the difference plots show that IPA-F has more MOG than FIS for both the Low and Near-Surface layer.

#### 4.1.2.2 AS A FUNCTION OF THRESHOLD

Figure 4.9 shows difference maps for LOG icing, which can be compared to MOG difference maps in Figure 4.8 (bottom). While there is more IPA-F MOG compared to FIS PDT MOG everywhere for both the Near-Surface and Low layer, the LOG threshold looks slightly different. There tends to be more FIS PDT LOG compared to IPA-F in the Low layer, except over the northern part of the domain and over the Wrangell and St. Elias Mountains. In the Near-Surface layer, there is more IPA-F LOG over the Gulf of Alaska and more FIS PDT LOG over the Brooks Range.

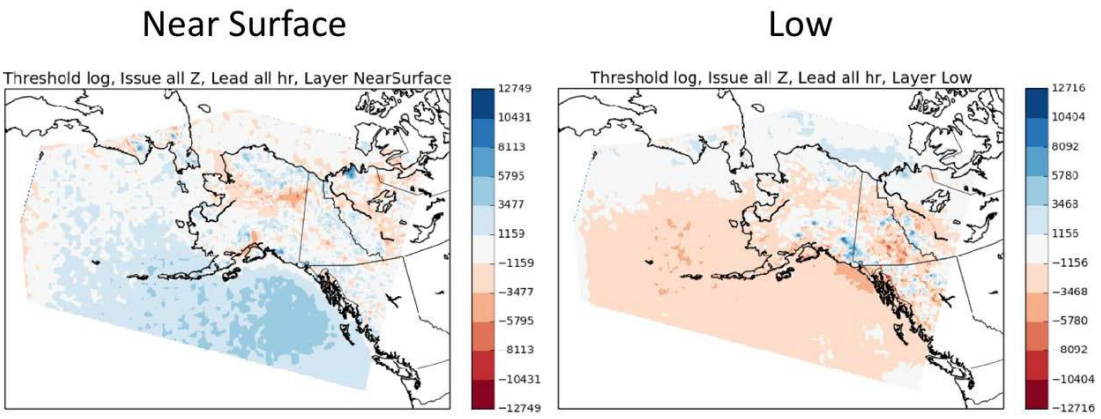


**Figure 4.9 Climatological maps showing count differences (IPA-FIS PDT) of LOG in the Near-Surface (left) and Low (right) layer.**

Distribution plots in Figure 4.10 and Figure 4.11 confirm the conclusions from the climatological maps. IPA-F and FIP PDT have comparable or slightly less Trace and Light counts in the Near-Surface layer, while IPA-F tends to have more Moderate and Severe. In the Low layer, a bigger difference in the Light category is evident where FIP tends to have more Light. Again, IPA-F has notably more Severe.

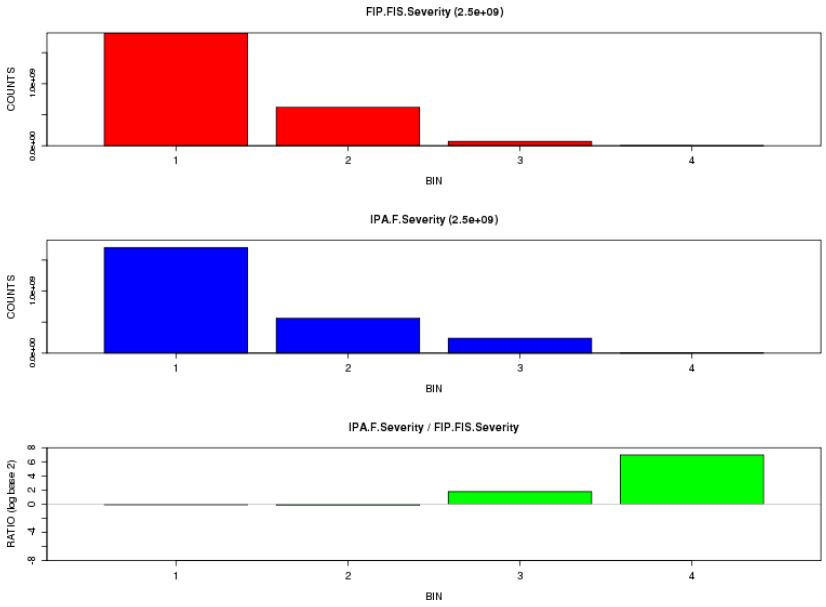


**Figure 4.10 Distribution of FIS PDT (top) and IPA-F (middle) severity thresholds, and a ratio plot of IPA-F to FIS PDT (bottom) for the Near-Surface layer. The ratios are plotted on a log2 scale, so that 1 represents a doubling of the counts and -1 represents a halving of the counts.**

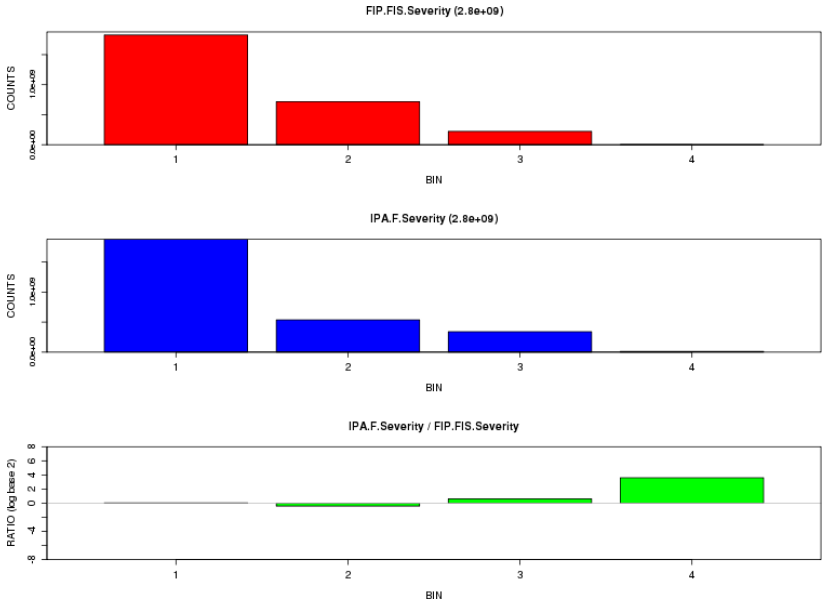


**Figure 4.11 Same as Figure 4.10, but for the Low layer.**

#### 4.1.2.3 IPA Full vs Overlap Domain

Severity in the full IPA domain behaves similarly to probability, having maxima outside the Overlap domain for the Middle and High layers. Figure 4.12 shows distribution ratios comparing severity in the Overlap domain to the Full domain, broken down by layer. Note that the Overlap region is a subset of the Full domain, and so differences between the two reflect the behavior outside of the Overlap region. Further, note that counts have been normalized within each region by dividing the counts

within each bin by the total number of grid points in the respective regions. A distinct pattern emerges that higher severities are more prevalent in the Overlap region for the Near-Surface and Low layers, while the opposite is true for the Middle and High layers.

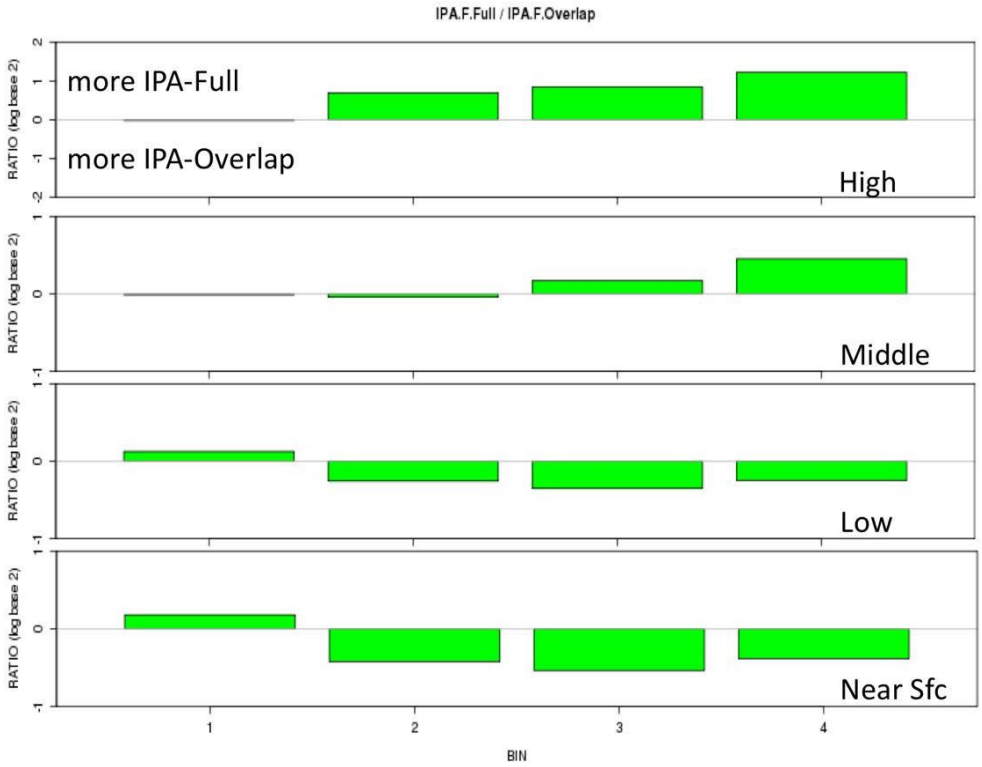


**Figure 4.12 Ratio plot comparing severity counts in the full IPA domain and IPA-F overlap domain for each layer. Counts are first normalized within each region, respectively, to account for the difference in the number of grid points in the two regions. The ratios are plotted on a log2 scale, so that 1 represents a doubling of the counts and -1 represents a halving of the counts.**

### 4.1.3 SLD

#### 4.1.3.1 As a Function of Lead

Figure 4.13 shows the climatological map of SLD counts greater than 5% probability as a function of lead in the Near-Surface layer. Note that there is less SLD at the 3-h lead (max pixel = 240 vs. 300–340 for later leads). Also, the maxima become more diffuse for longer leads. Figure 4.14 highlights the small scale SLD features that appear along the Coast Mountains.

In the Low layer (not shown), there is little SLD over Alaska which is confined to small bullseyes along the Coast Mountains, similar to the features in Figure 4.14. While a lead signature was noted for the Near-Surface layer, there is no apparent dependence on lead time for Low layer.

SLD is not forecast in either the FIP/FIS product or the gridded AAWU product.

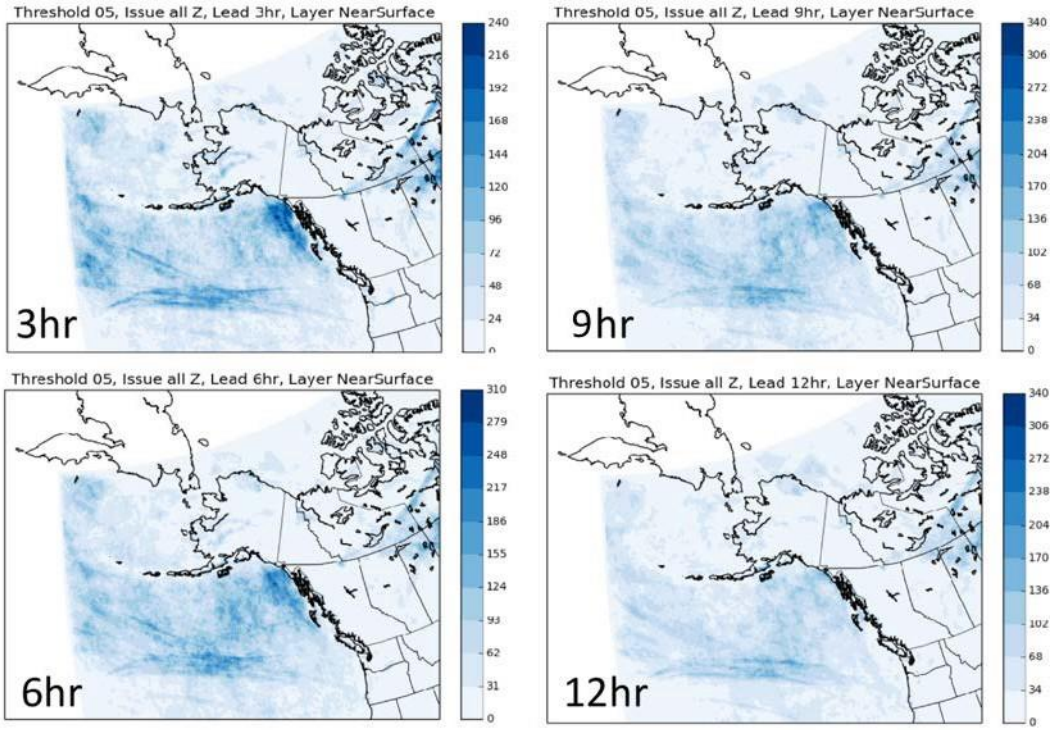


**Figure 4.13 IPA-F SLD for the 5% threshold in the Near -Surface layer as a function of lead.**

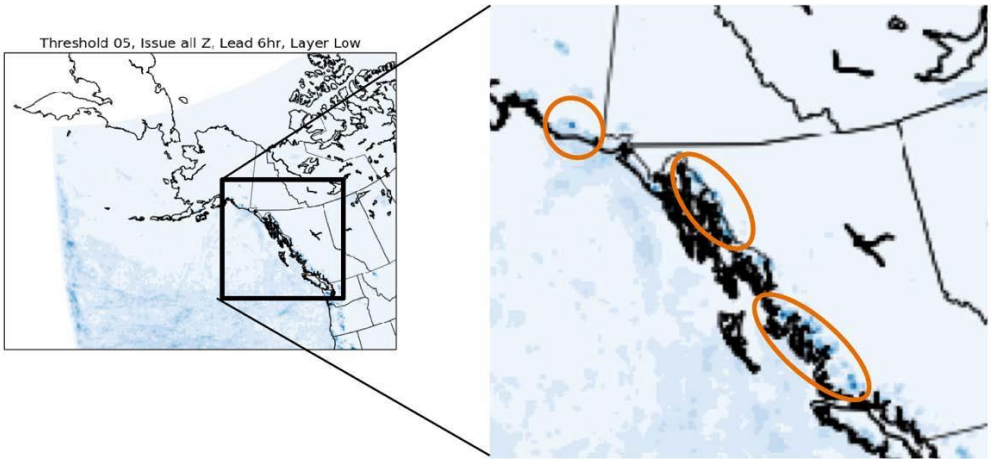


**Figure 4.14 IPA-F SLD for the 5% threshold in the Low layer along with a zoom of the region indicated, highlighting small scale SLD features along the Alaska and Canadian coasts.**

## 4.2 PROBABILITY/SEVERITY PERFORMANCE

### 4.2.1 CONSISTENCY

Figure 4.15 shows the consistency of LOG and MOG IPA-F forecasts for consecutive leads valid at the same time. Interestingly, the consistency for the MOG icing fields changes little out through 12 hours, then begins to drop off. LOG icing shows a similar decline in consistency after 12 hours, along with a

drop after the 3-hr lead (when the difference in leads changes from 1 to 3 hours). Field consistency may be due to good model consistency for fields such as temperature that are inputs to IPA-F algorithm.

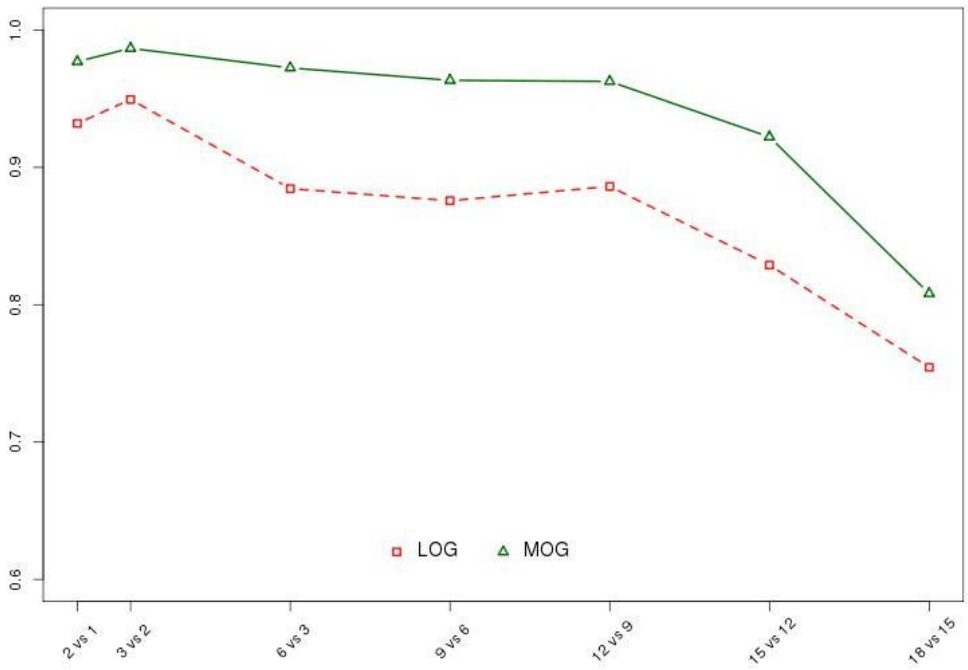


**Figure 4.15 Consistency of IPA-F LOG and MOG forecasts with the previous lead valid at the same time.**

### 4.2.2 VERIFICATION USING PIREPS

Figure 4.16 shows IPA-F POD, POFD, and PSS plotted as a function of lead time. It suggests that there is only a slight decline in performance by lead, but there is a large decrease in performance as the probability threshold increases. The lack of a decline in skill as a function of lead time is congruent with the forecast consistency shown in Figure 4.15. In addition, the neighborhood best-match approach likely contributes to the maintenance of performance across leads. Greater skill is typically easier at larger scales.

Lowering the forecast threshold increases the forecast volume, which in turn, tends to increase both the number of hits and the number of false alarms. However, in rare-event scenarios, the benefit of more hits outweighs the penalty of more false alarms, thus explaining substantial increase in skill as the forecast threshold is lowered.

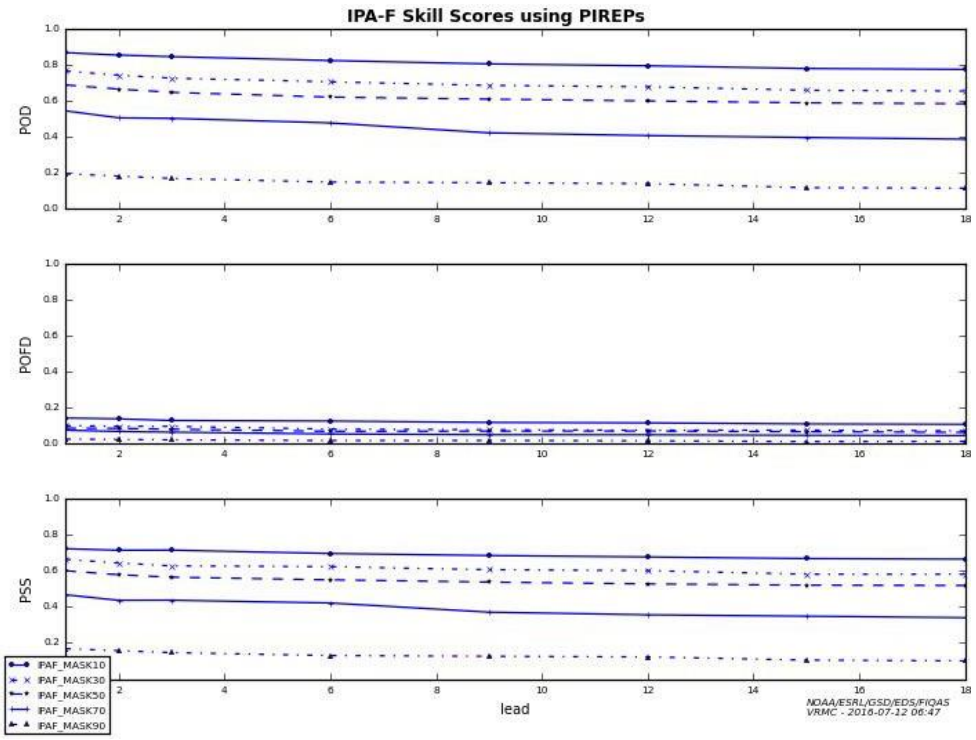


**Figure 4.16 POD (top), POFD (middle), and PSS (bottom) plotted as a function of lead for various potential masks.**

Figure 4.17 shows the POD, POFD, and PSS comparing IPA-F and FIP/FIS PDT thresholds for LOG icing combining all lead times. For LOG, IPA-F is comparable to or more skillful than FIP (PDT) except where the potential is greater or equal to 30. Figure 4.18 shows the same plot but for MOG. It shows IPA-F has higher POD than FIS PDT at all thresholds, but also slightly higher POFD. The high POD results in substantially better skill (PSS) for IPA-F compared to FIP/FIS PDT for MOG.

The FIP/FIS AAWU thresholds were also investigated and found to have lower POD and slightly lower POFD than results using the PDT thresholds (not shown). As a result, IPA-F skill is again generally better than FIP/FIS AAWU, with the exception of when potential is greater than or equal to 30, LOG.

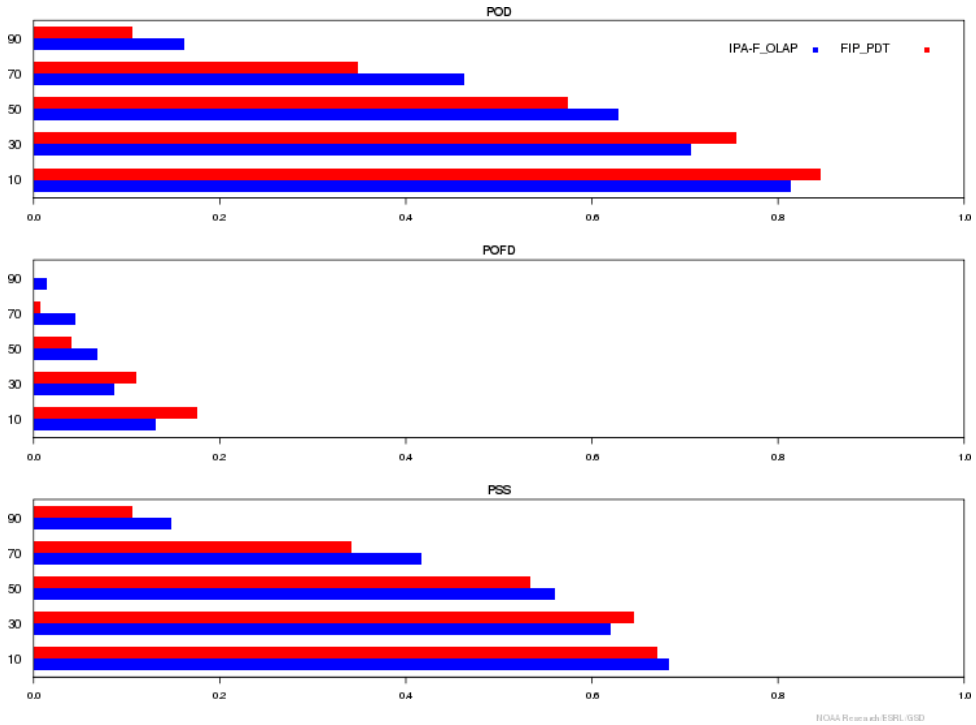


**Figure 4.17 POD (TOP), POFD (middle), and PSS (Bottom) for the thresholds of potential: 10, 30, 50, 70 and 90, for LOG severity. This plot uses PDT thresholds and contains all layers.**

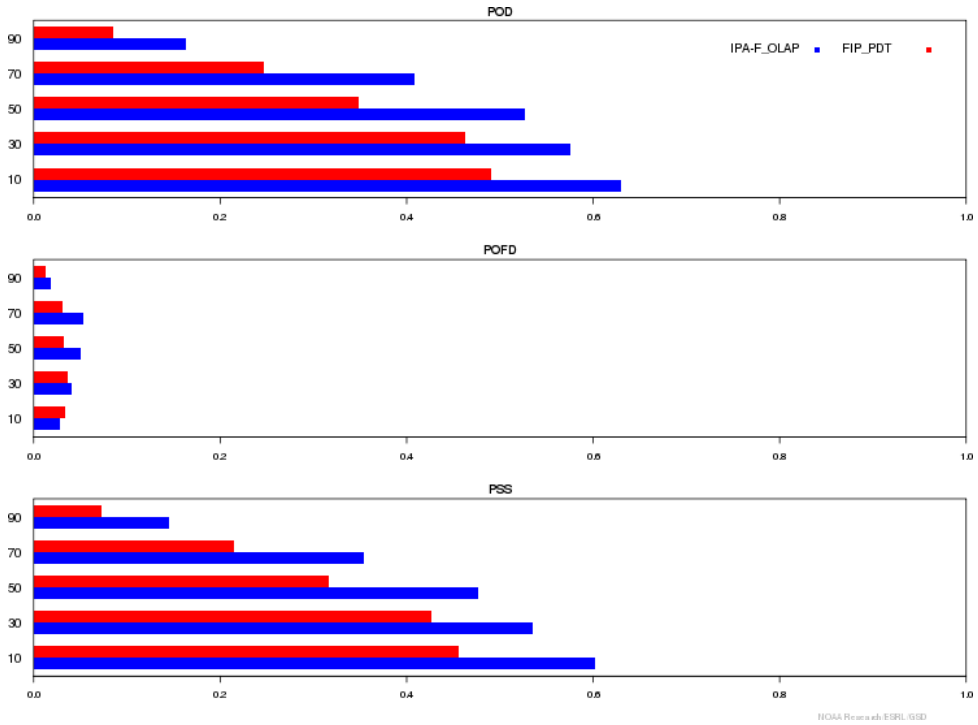


**Figure 4.18 same as Figure 4.17, but for MOG severity.**

Figure 4.19, Figure 4.20, and Figure 4.21 demonstrate regional exceptions where FIP/FIS PDT tends to outperform IPA-F. This occurs only for LOG, in the Interior, South Central, and the Southwest and Bering regions.

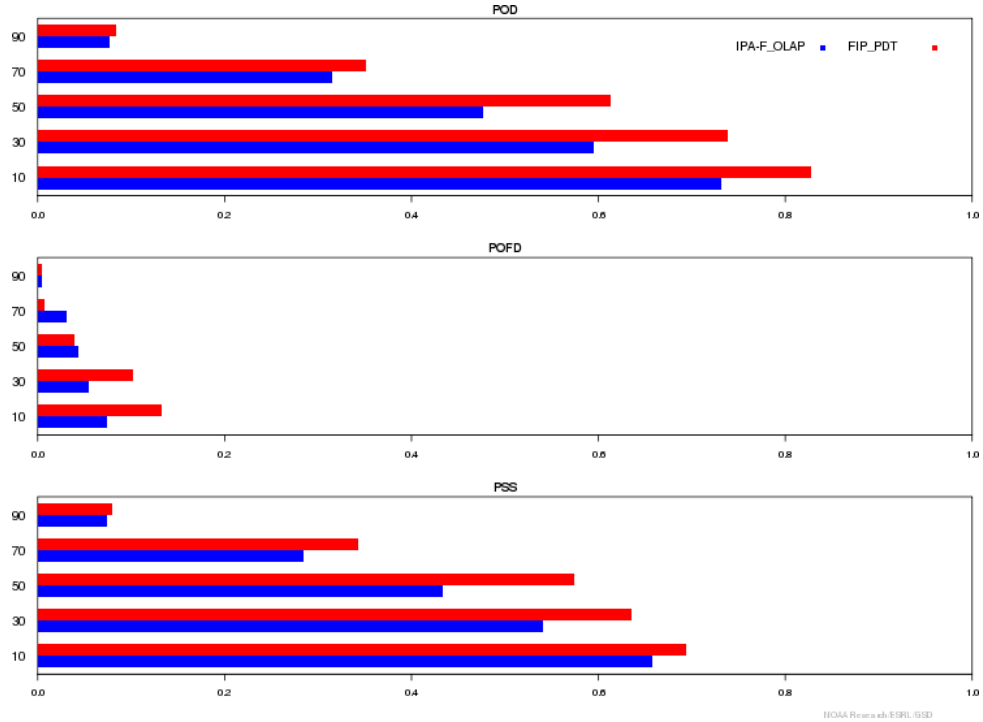


**Figure 4.19 Same as Figure 4.17, but for LOG severity, the Near-Surface layer, and the Interior region.**

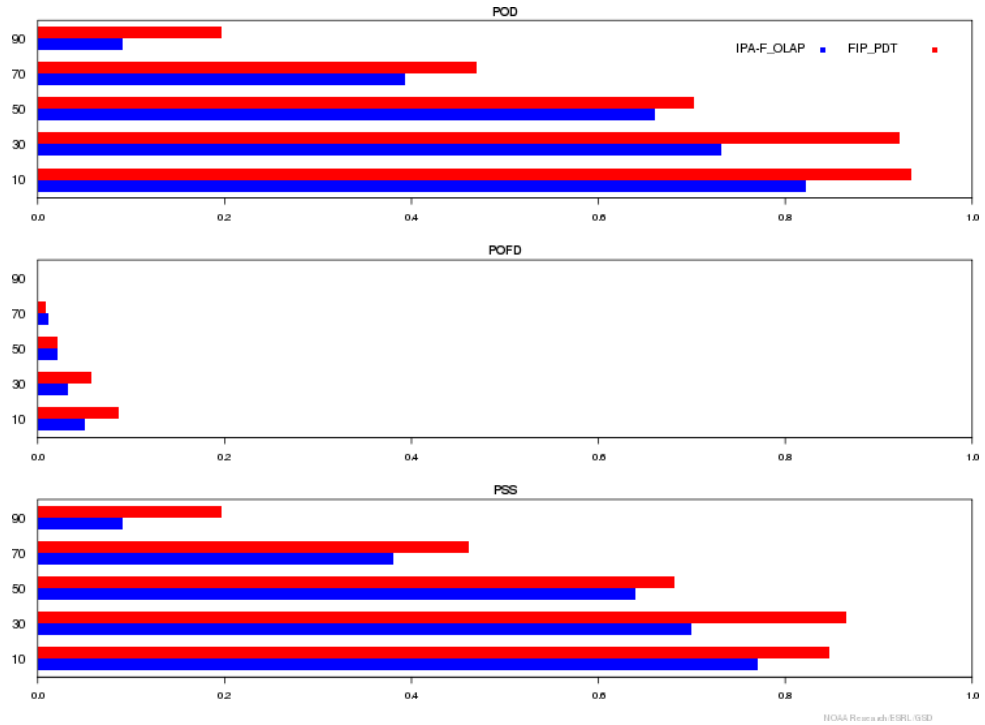


**Figure 4.20 Same as Figure 4.17, but for LOG severity, the Near-Surface layer, and the South-Central region.**

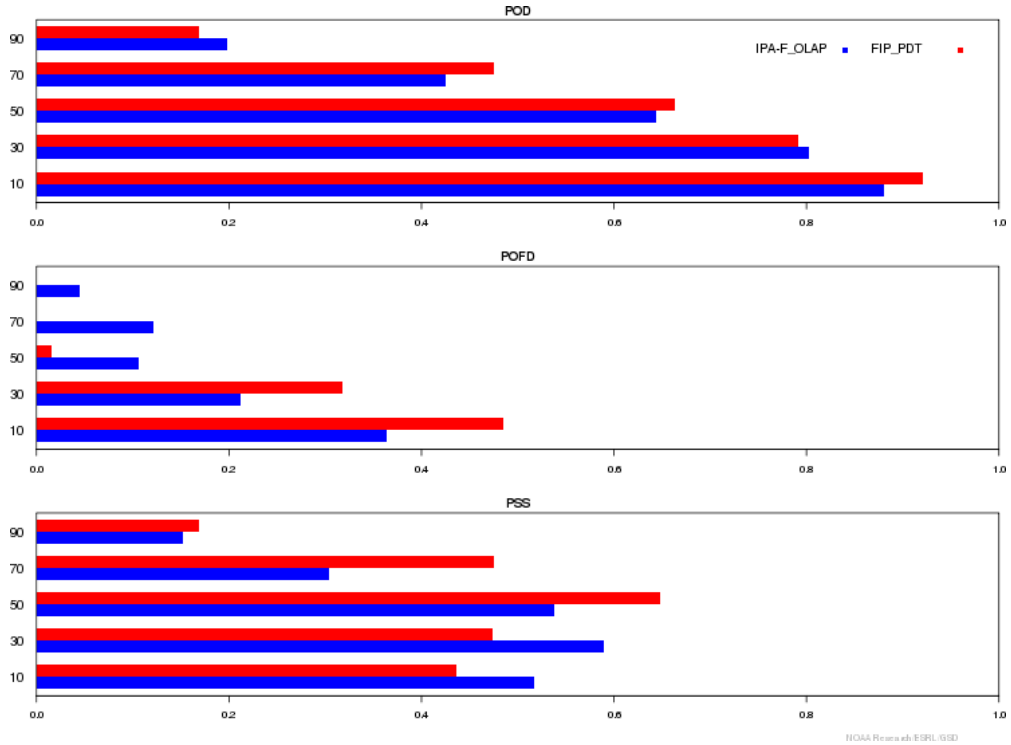


**Figure 4.21 Same as Figure 4.17, but for LOG -severity, Low layer, and the Southwest and Bering region.**

In section 4.1.1.4 and section 4.1.2.3, differences were noted between IPA-F in the Full domain and in the Overlap region, with high severities more prevalent in the Full domain in the higher layers but

more prevalent in the Overlap for lower layers. Relatedly, IPA-F is slightly more skillful in the Overlap region compared to the Full domain for low levels, but is more skillful in the Full domain for higher levels (not shown).

### 4.2.3 Verification Using Soundings

Figure 4.22 shows the POD and POFD values for IPA-F and FIP/FIS PDT MOG icing where potential is greater than or equal to 10, 30, and 50, as a function of lead time, verified using sounding data. Plots show IPA-F consistently outperforms FIP/FIS out to 18 hours, but accuracy declines with lead time more rapidly for IPA-F.

Note that the POD values shown in the sounding plots are much lower than POD values when verified against PIREPs. This is due to differing methodologies; against PIREPs, the best-match within a neighborhood is used, while against soundings, only the nearest grid point is used. The decrease in accuracy occurs due to this restricted spatial scale.

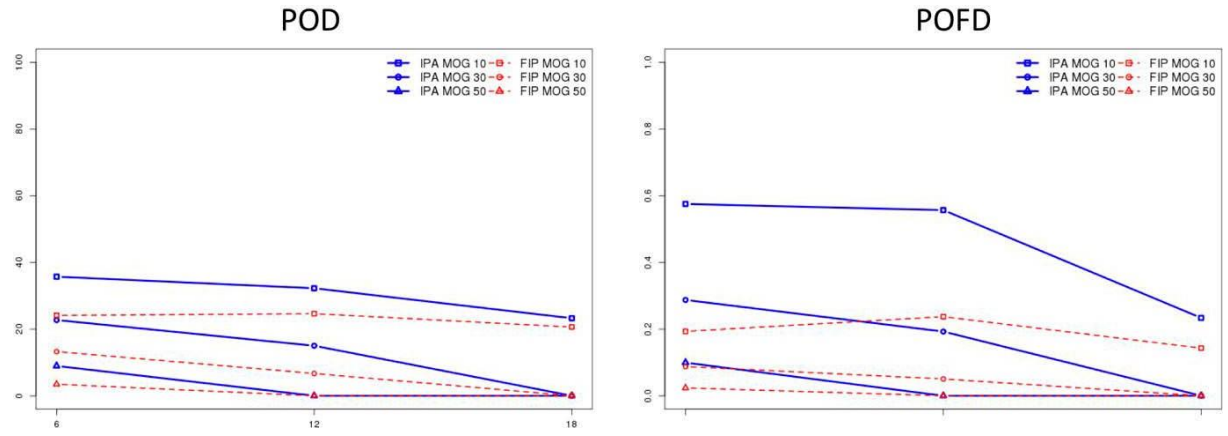


**Figure 4.22 POD (left) and POFD (right) of IPA-F (Blue) and FIP/FIS PDT (red) as a function of lead (6, 12, 18-hr) for various MOG thresholds (10, 30, and 50 potential)**

### 4.2.4 IPA-F Compared With Gridded AAWU

Figure 4.23 shows POD as a function of volume for IPA-F compared to gridded AAWU product, in each layer. For LOG forecasts (left), IPA-F yields a substantially higher POD than the gridded AAWU product for each layer, and the volume of IPA-F forecast is significantly less than that of gridded AAWU product, except for the Near-Surface layer. For the Near-Surface layer, the doubling of the POD for IPA-F compared with the gridded AAWU product comes at the expense of a near doubling in the volume. While IPA-F volume decreases with height, the gridded AAWU forecast volume is largest in the Low layer.

For MOG forecasts (right), the gridded AAWU forecasts cover very little volume, and consequently, PODs are less than 10% for any layer. Only the IPA-F High layer has a smaller forecast volume than the gridded AAWU MOG forecasts, but manages a POD of 50%. Interestingly, IPA-F performance is lowest for the Middle layer, while the gridded AAWU product performs best in this layer.

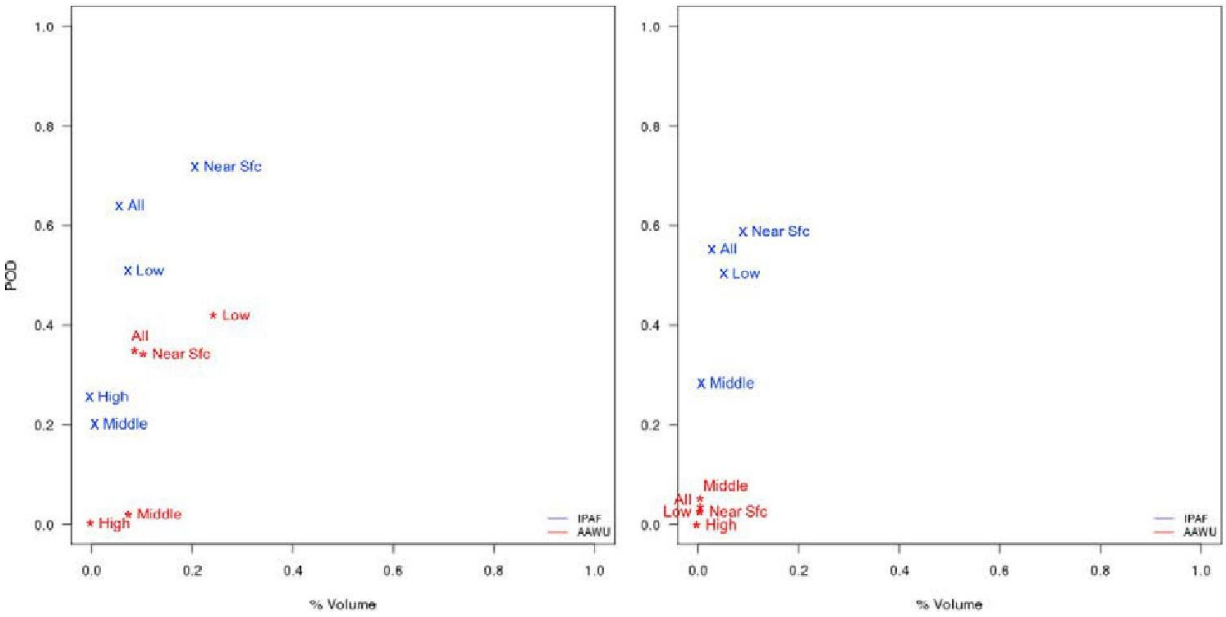


**Figure 4.23 Percent volume vs. POD for IPA-F (blue) and gridded AAWU product (red) for each layer for LOG (left) and MOG (right) forecasts.**

Figure 4.24 shows a similar plot of POD as a function of volume, but for different thresholds of potential. The gridded AAWU MOG volumes are similar to the IPA-F SEV forecasts. Using a potential-mask of 10, IPA-F MOG volume is slightly less than the volume of the gridded AAWU LOG forecast, with nearly twice the POD. IPA-F MOG volume and POD decrease as the potential mask increases, but the POD remains higher than for the gridded AAWU LOG forecast up through the potential mask of 70. Using 3-hr IPA-F offset rather than the 1-hr offset shown here does not change the results (not shown).

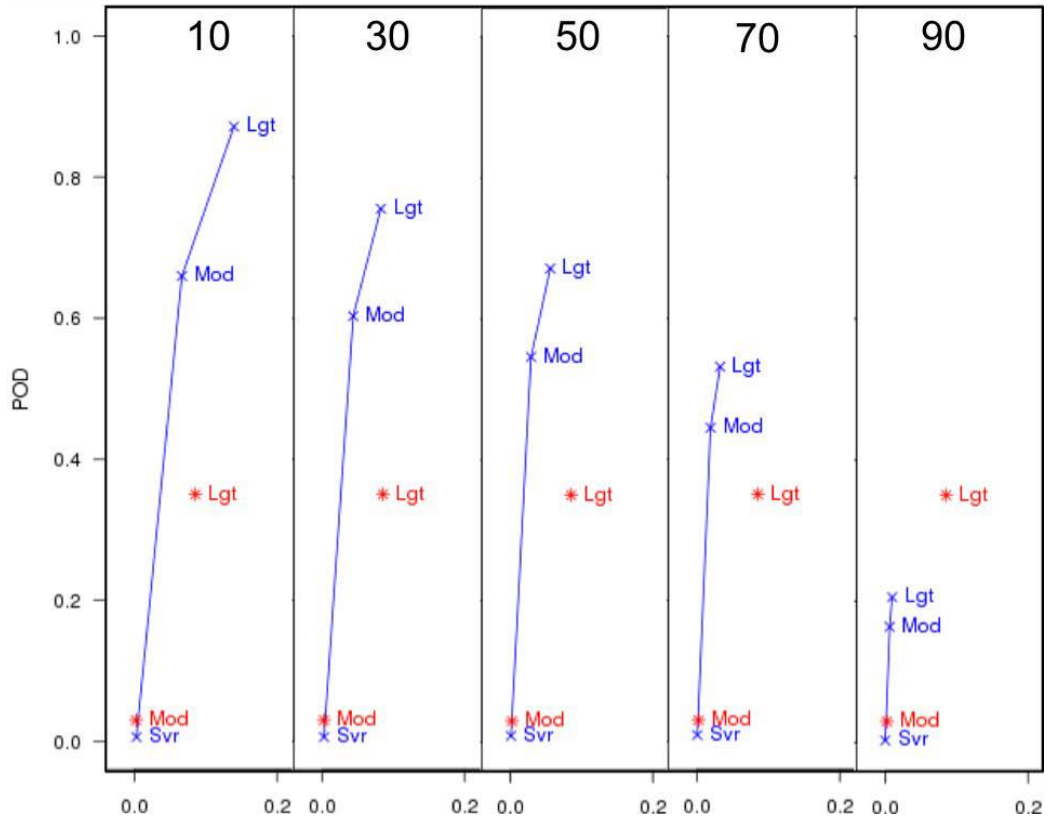


**Figure 4.24 Percent volume vs. POD for IPA-F (blue) and gridded AAWU product (red) for varying potential thresholds.**

### 4.2.5 IPA-F IN CONJUNCTION WITH GRIDDED AAWU

Figure 4.25 and Figure 4.26 show POD, POFD, and PSS for IPA-F in and out of the gridded AAWU using a potential-mask of 10 and 50, respectively. "In" refers to the area for which the gridded AAWU product is forecasting MOG icing. "Out" refers to the rest of the region common to both the gridded AAWU product and IPA-F; that is, where the gridded AAWU product is not forecasting MOG icing to occur. As a result, for In (forecast MOG = yes everywhere), the POD and POFD are equal to one, while for Out (forecast MOG = no everywhere), the POD and POFD are zero. The IPA-F scores are thus described relative to gridded AAWU forecast performance. For example, using the potential mask of 10 (Figure 4.25), IPA-F inside the gridded AAWU field excludes 20–40% of non-MOG PIREPs (POFD ~0.6-0.8) at the cost of missing very few MOG PIREPs (POD near 1.0). On the other hand, the gridded AAWU POD and POFD are equal to 0 outside the field, meaning IPA-F captures 70–90% of MOG PIREPs missed by the gridded AAWU product at the cost of a 30–40% increase in false alarms.

When using a higher probability mask (Figure 4.26), the number of non-MOG PIREPs that are excluded within the gridded AAWU product doubles, while missed MOG events increases to roughly 20%. The number of false alarms added outside of the AAWU forecast area drops to 20%, while still adding 50–70% of MOG events missed by AAWU. Increasing the potential mask to 90% (not shown), reduces the number of false alarms to near zero, both inside and outside the gridded AAWU MOG forecast area, but nearly 80% of the MOG PIREPs become missed events.

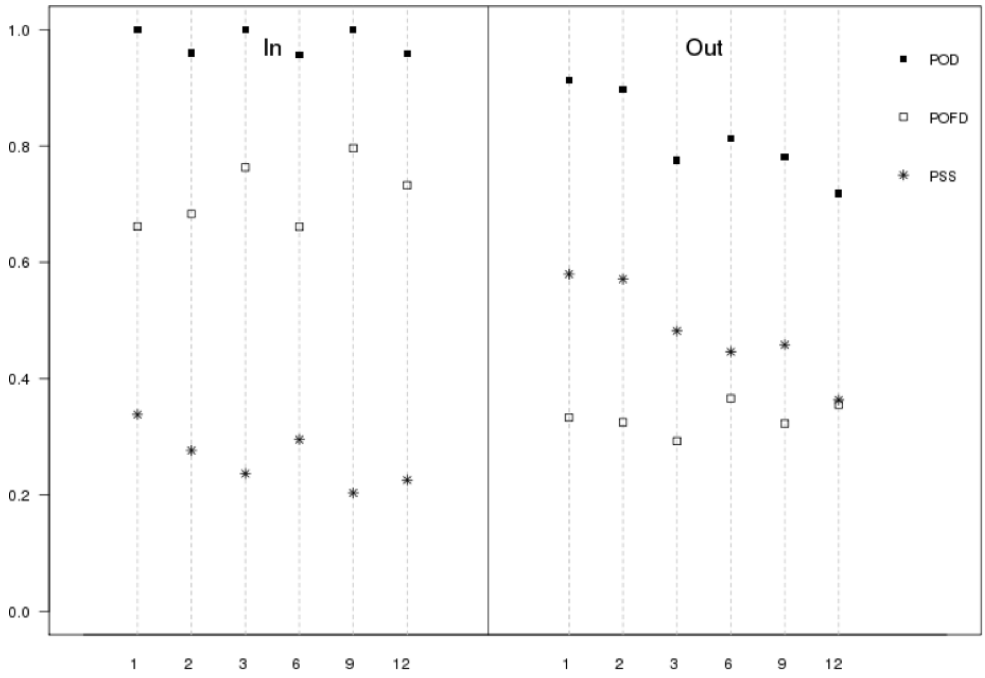


**Figure 4.25 IPA-F POD, POFD, and PSS as a function of forecast lead time, verified with PIREPs inside (left) and outside (right) of the gridded AAWU forecast, using MOG severity and potential = 10.**

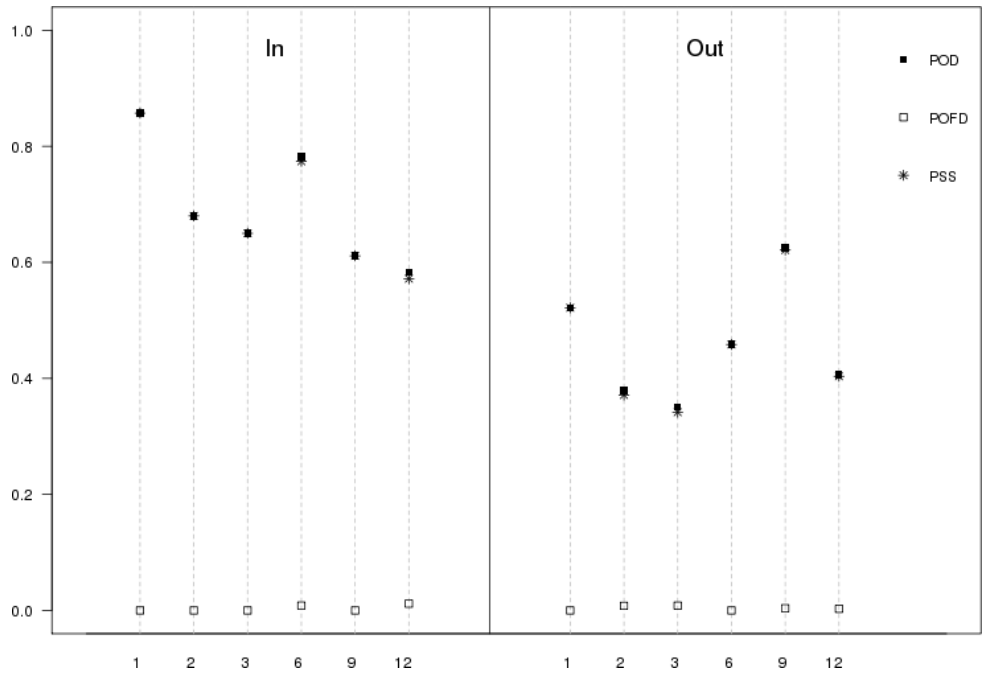


**Figure 4.26 Same as Figure 4.25, but for potential = 50.**

## 4.3 SLD PERFORMANCE

Both PIREPs and METARs are used to verify SLD. For the ADDS display of the Forecast Icing Product, SLD potential is converted to a yes/no, where all SLD potential ≥ 0.05 is defined as a 'yes' forecast of SLD. However, the SLD field has an additional category, considered 'unknown'. For the ADDS display these grid points are treated as 'no' forecasts, but the performance of the IPA-F SLD forecasts is assessed considering each of the three possible scenarios of the 'unknown' points: 1) considering the 'unknown' points as 'yes' forecasts 2) considering them as 'no' forecasts, and 3) leaving them as 'unknown'—essentially removing those points from the verification.

Figure 4.27 shows the SLD skill as a function of lead time, verified against PIREPs (gray) and METARs (blue). In this case the 'unknown' points are considered as 'yes' forecasts. The plot shows there are more events captured (higher POD), but more false alarms (higher POFD) when using PIREPs. Despite this difference, the PSS skill is comparable. The difference in POD/POFD likely arise from the different matching methods, i.e. PIREPs are matched within a neighborhood and METAR are matched within a column.

For the scenario where 'unknown' is set to 'no' or is kept as 'unknown', POD and skill drop to near zero (not shown). In other words, there are very few 'yes' SLD forecasts where observations are available. Not only the METARs (Figure 2.4) but even the PIREPs (Figure 2.3) are located over land while the IPA-F SLD forecasts are more prevalent over water (Figure 4.13).

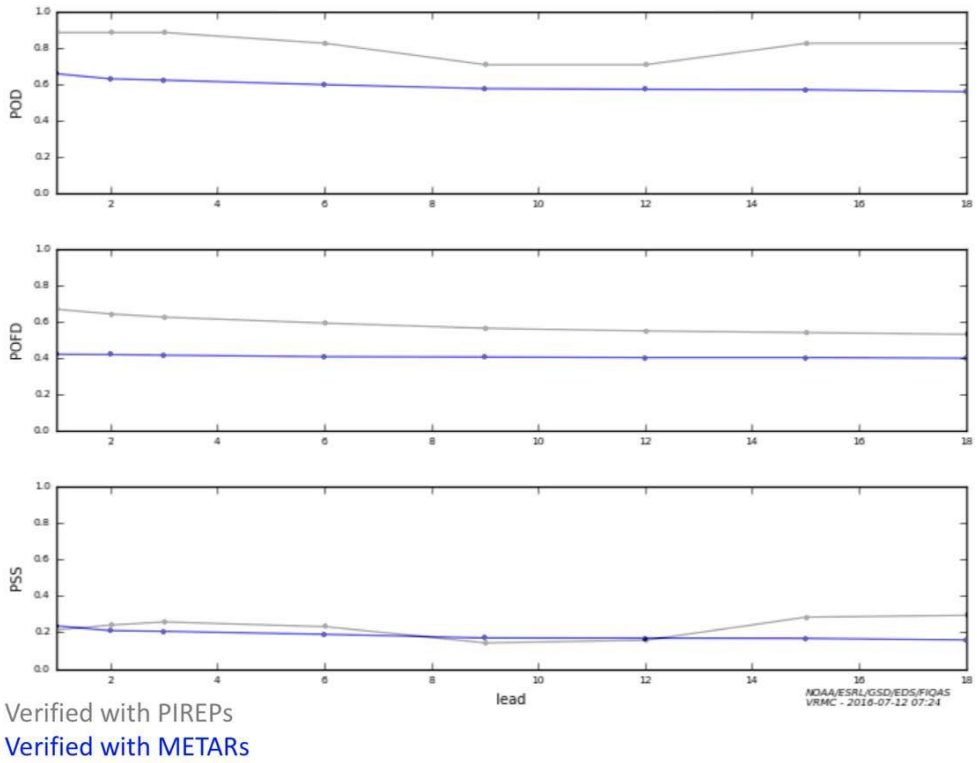


**Figure 4.27 IPA-F SLD POD (top), POFD (middle), and PSS (bottom) verified with PIREPs (gray) and METARs (blue) as a function of lead time (in hours). 'Unknown' SLD is treated as 'Yes' forecast.**

## 4.4 Case Studies

### 4.4.1 05 January 2016 0200Z

Multiple icing reports came in from aircraft in the vicinity of Cook Inlet between 2020 UTC 04 January and 2000 UTC 05 January 2016. Here, we investigate 0200 UTC January 05, when there were two PIREPs over the inlet. The PIREP severities and altitudes are shown in Figure 4.28.

Figure 4.29 shows the NASA Langley GOES cloud top and cloud base product (http://www-pm.larc.nasa.gov/cgi-bin/site/showdoc?docid=22&domain=alaska_nw-pacific&lkdomain=Y). During this icing event there was southeasterly flow over the Chugach and Kenai Mountains to the east of the Cook Inlet, leading to thick cloud layer over the inlet with bases at ~3000 ft and tops near ~35000 ft.

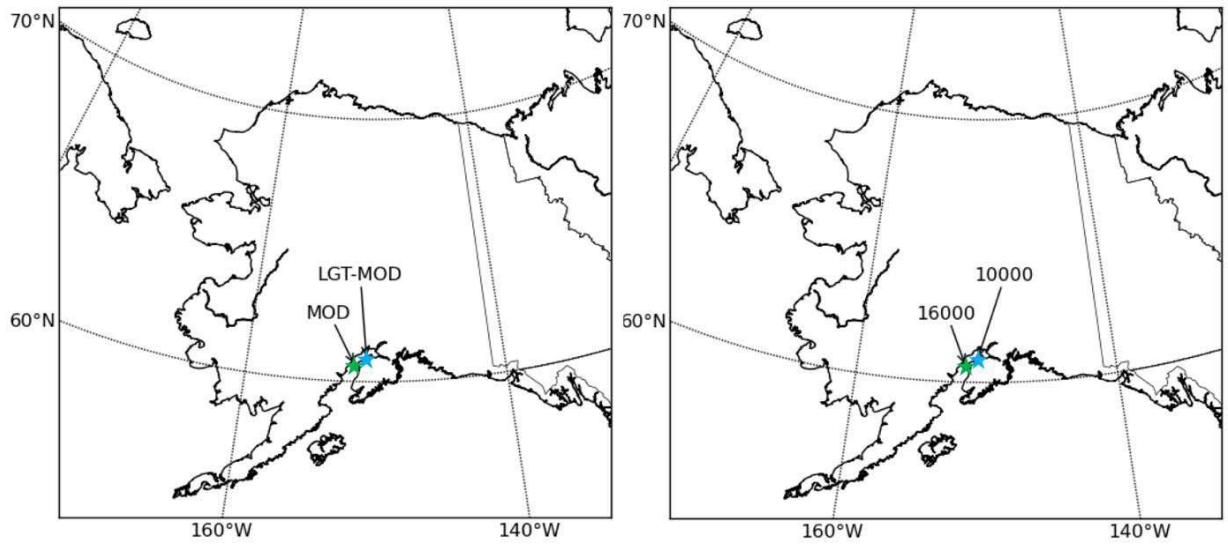


**Figure 4.28 Severity and altitude (ft) of two PIREPs over the Cook Inlet between 0200–0300 UTC on 05 January 2016.**

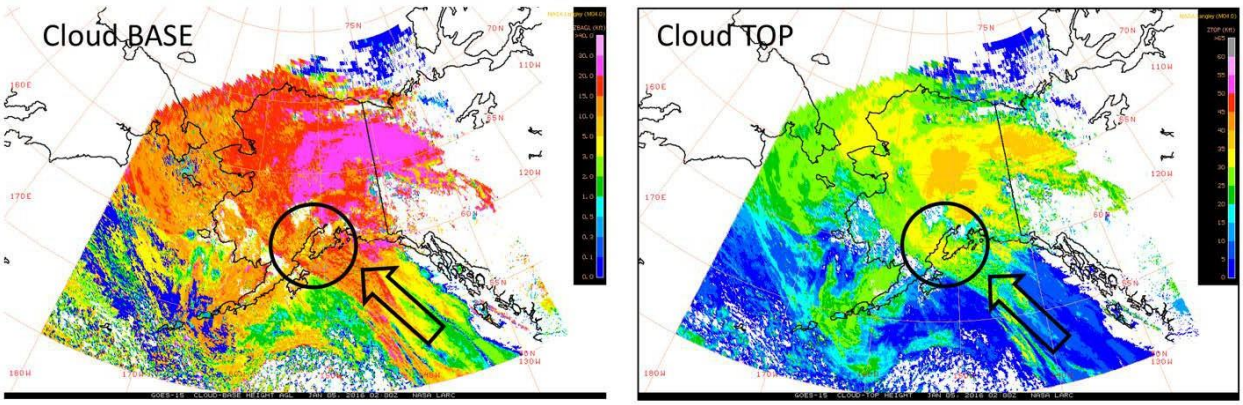


**Figure 4.29 NASA Langley GOES cloud base (left) and top (right).**

Figure 4.30 shows the FIP and IPA-F forecasts for the 16000-ft layer, the flight level of the MOD PIREP (green star). One notable disadvantage of FIP is that because of issuance times, a forecast for the exact valid time (0200 UTC) is unavailable; thus, forecasts valid at 0300 UTC and the 0000 UTC analysis are investigated. When looking at IPA-F severity forecasts, there is a moderate signal over the Cook Inlet consistent with the PIREP. For the FIS PDT severity forecasts, there is consistent light severity over Cook Inlet, which is weaker than what the PIREP and IPA-F suggests. For this case, the probability is relatively low and this signal is consistent between IPA-F and FIP (not shown).

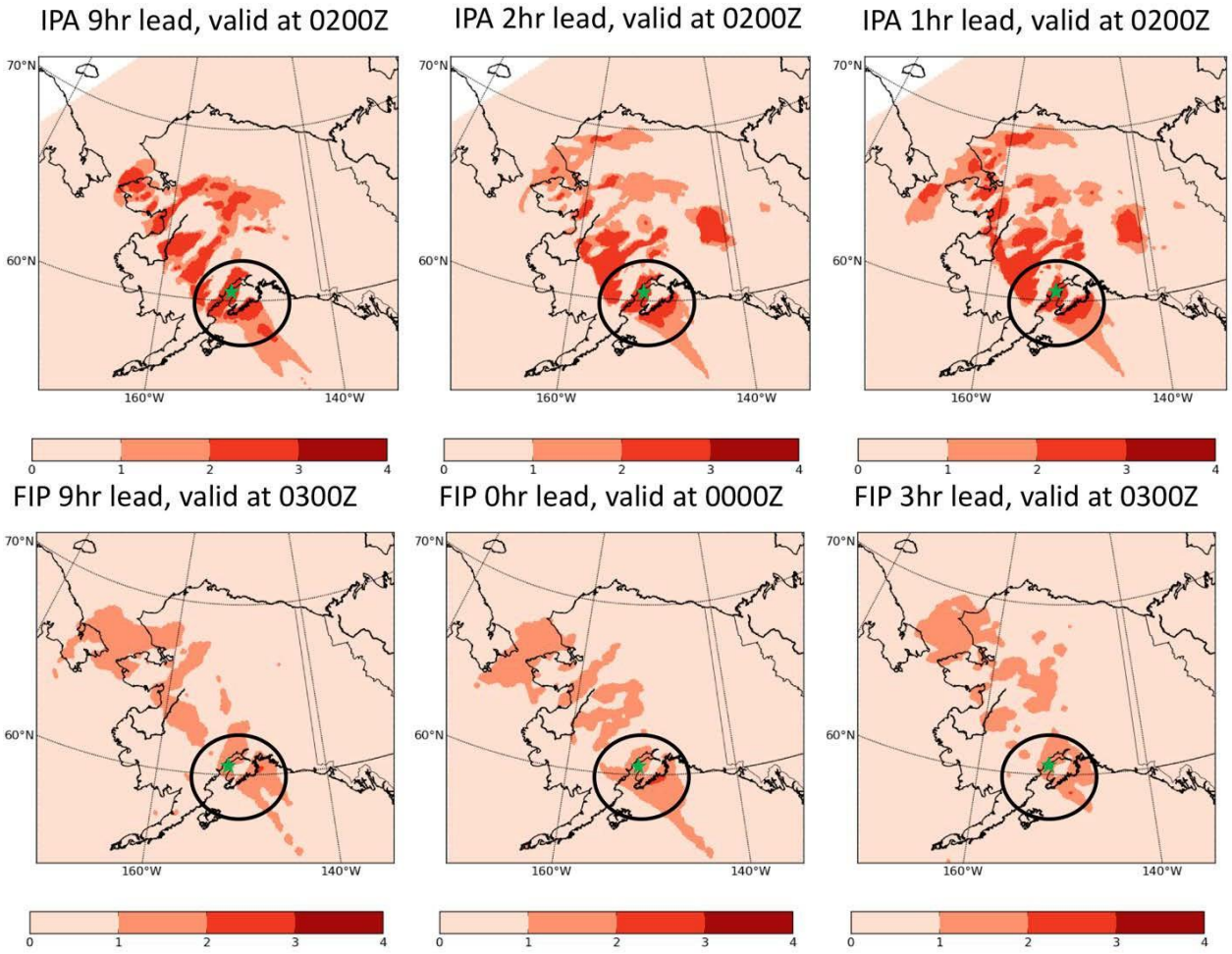


**Figure 4.30 IPA (top row) and FIS PDT (bottom row) severity forecasts at 16000 ft valid at or near 0200 UTC on 05 January 2016. Forecast severity categories are identified as 0=none, 1=light, 2=moderate, and 3= severe.**

Figure 4.31 shows a similar plot, but now for the 10000-ft layer, corresponding to a LGT-MOD PIREP. There is a consistent moderate to severe signal in IPA-F forecasts over Cook Inlet, with the PIREP falling along a strong gradient. The FIS forecast using PDT thresholds has a consistent moderate signal over Cook Inlet, which is weaker than what IPA-F suggests, but consistent with the PIREP. IPA-F and FIP both show high probability for this case (not shown).

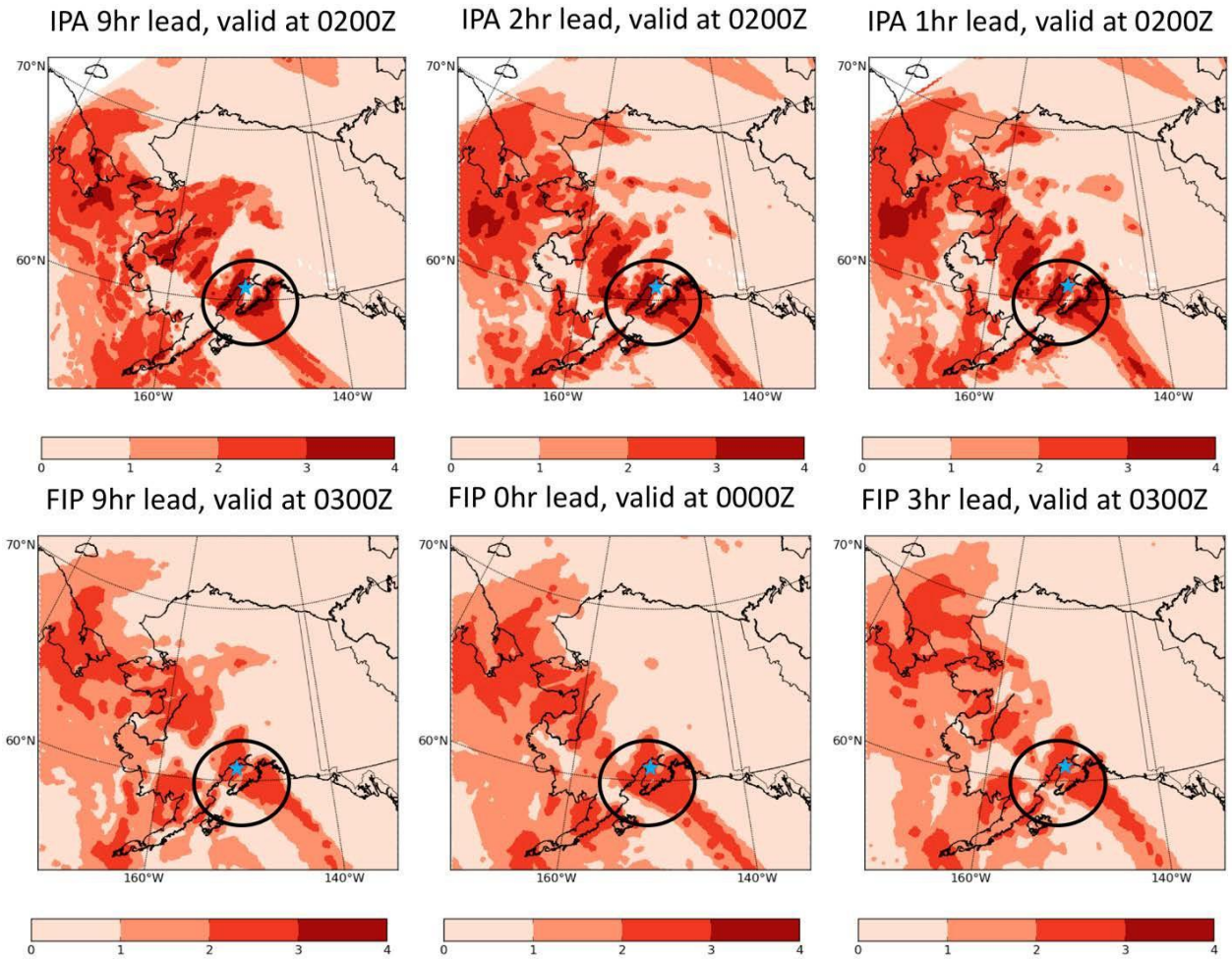


**Figure 4.31 Similar to Figure 4.30, but for 10000 ft**

### 4.4.2 20 SEPTEMBER 2015 1800 UTC

On 19–20 September 2015, a stable stratus icing layer set up along the northern coast of Alaska. The levels of the reports ranged from 1400–5000 ft, with the majority of the reports between 3700 and 4500 ft, coinciding with the top of the status layer. Here we investigate 1800 UTC September 20, when there are three PIREPs over the Barrow. The PIREP severities and altitudes are shown in Figure 4.32.

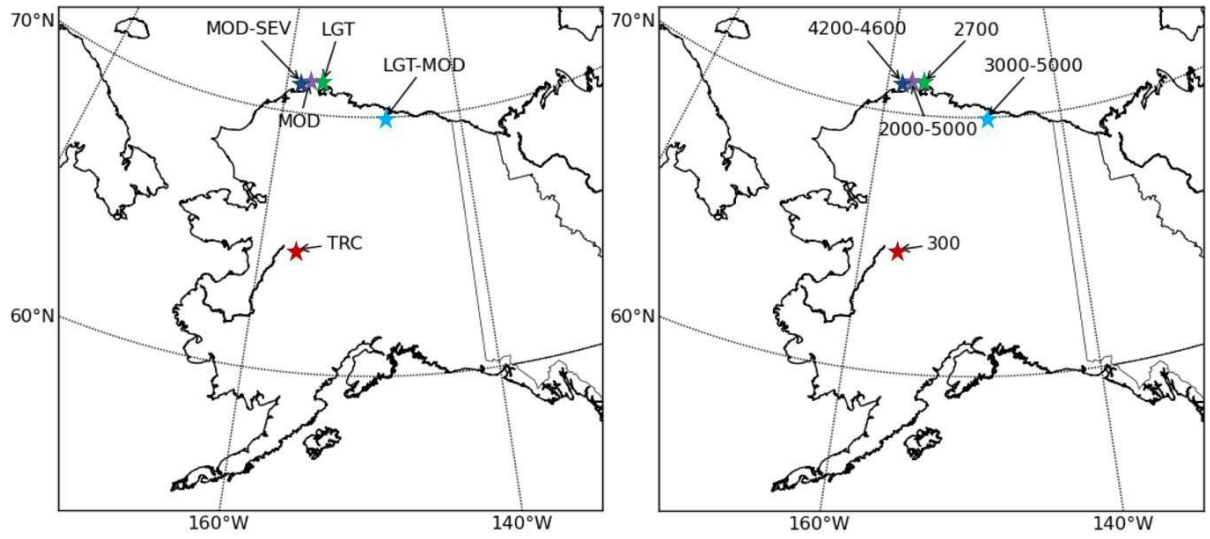


**Figure 4.32 Severity and altitude (ft) of five PIREPs between 1800–1900 UTC on 20 September 2015, three of which are over Point Barrow.**

Figure 4.33 shows the FIP and IPA-F forecasts for the 4500-ft layer, where three of the five PIREPs are valid. For the 6-hr forecast, IPA-F shows a ribbon of higher probabilities along the northern coast of Alaska collocated with the MOD and MOD-SEV PIREPs (purple/navy stars), but there is very low probability collocated with the LGT-MOD PIREP (light blue star). FIP has a splotchy pattern along the northern coast, with lower probabilities compared to IPA-F, although FIP forecasts slightly higher probabilities associated with the LGT-MOD PIREP.

Looking at the 1-hr forecast, the IPA-F shows a consistent bullseye of high probability associated with the PIREPs along the northern coast (purple/navy stars), and correctly increases the probability forecast for the LGT-MOD PIREP (light blue star). The FIP analysis at 1800 UTC is relatively consistent with the 6-hr lead, missing the PIREPs along the northern coast.

Both FIS PDT and IPA-F are consistent in showing Light severity in the vicinity of the PIREPs, thus underpredicting the more severe PIREPs (not shown).

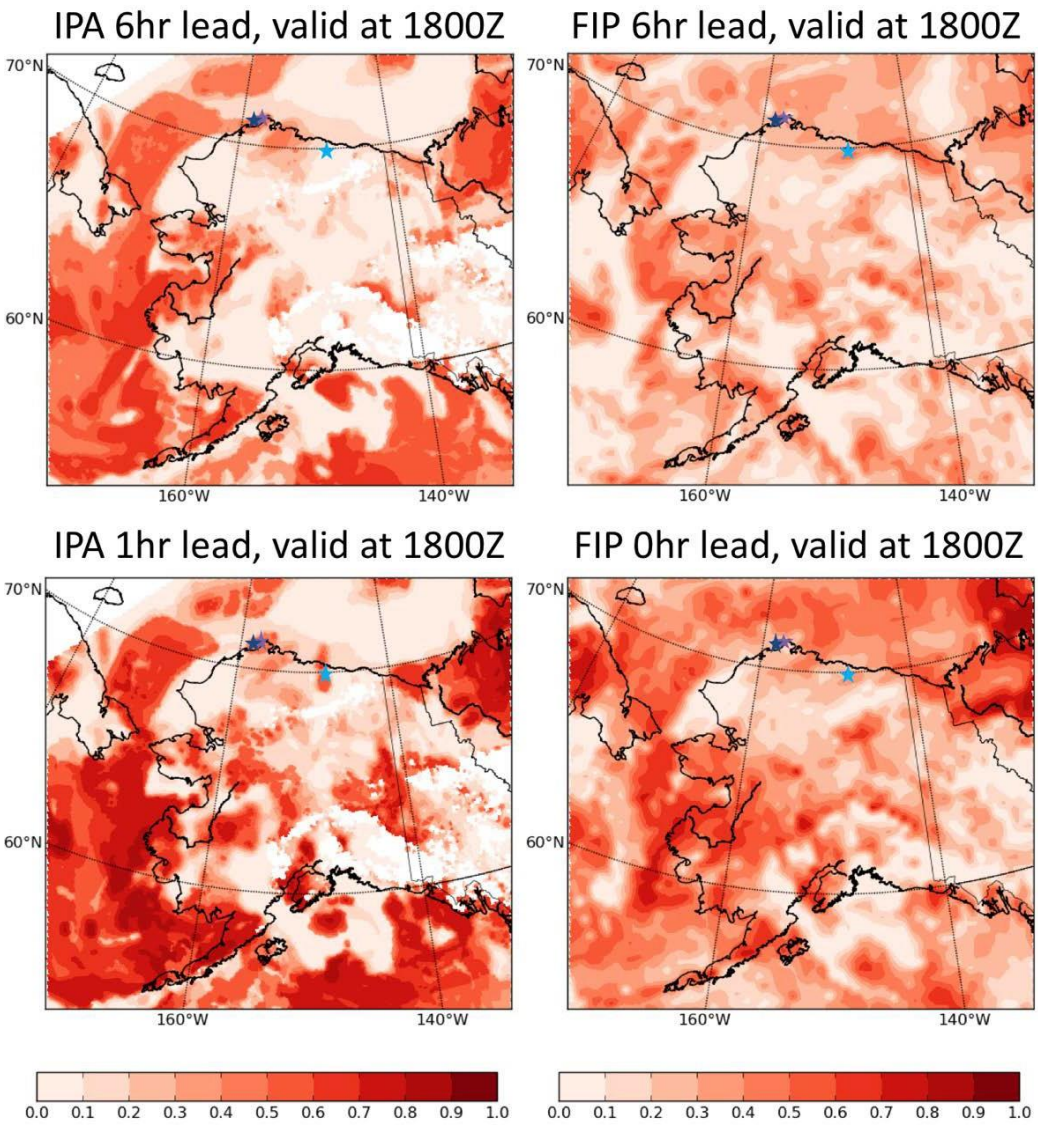


**Figure 4.33 IPA-F (left column) and FIS PDT (right column) severity forecasts at 4500 ft valid at 1800 UTC on 20 September 2015.**

# 5 CONCLUSIONS

The QA PDT was tasked to complete a follow-on quality assessment of the Icing Product Alaska Forecast (IPA-F) algorithm developed by the In-Flight Icing (IFI) PDT at the National Center for Atmospheric Research. The purpose of this second assessment was to compare the performance of the IPA-F to the current suite of Alaska icing forecast products, including the Forecast of Icing Probability (FIP), the Forecast of Icing Severity (FIS), and the gridded-Alaska Aviation Weather Unit (AAWU) icing product. The results of this study are intended to assist the Technical Review Panel in determining the readiness of IPA-F to be transitioned into operations. The study was conducted on data from the autumn of 2016, with particular emphasis on the area of responsibility for the AAWU. The forecasts were compared to each other as well as to observations, including pilot reports (PIREPs), aviation routine weather reports (METARs), and upper-air soundings.

In general, the assessment findings include:

- IPA-F outperformed FIP/FIS and the gridded -AAWU forecast
- IPA-F improvement was the greatest for higher probability thresholds and higher severities of icing (moderate or greater)
- IPA-F captured more moderate or greater (MOG) icing events than the gridded AAWU forecasts while forecasting that severity over a smaller area
- IPA-F forecasted larger areas of high probability and zero probability of icing than FIP; FIP forecasted a larger area of low probability of icing than IPA-F
- IPA-F forecasted a larger area of MOG icing than FIS

Overall, IPA-F outperformed FIP/FIS and the gridded AAWU forecasts when compared against observations. In particular, IPA-F shows an increase in skill when forecasting MOG icing, compared to FIP/FIS and gridded AAWU forecasts. First, IPA-F had a higher probability of detection (POD) than FIP on three of the five probability thresholds for light icing and on all five of the thresholds for moderate icing. This resulted in higher Pierce Skill Scores (PSS) for IPA-F than the FIP on four of the five probability thresholds for light icing and all five of the thresholds for moderate icing. Next, IPA-F was compared against the AAWU forecasts, where IPA-F had a higher POD at all thresholds for both light and moderate icing, leading to higher PSS for IPA-F on four of the five probability thresholds for light icing and all five thresholds for moderate icing.

IPA-F performed particularly well against the gridded AAWU forecasts. IPA-F captured more MOG icing events than the gridded AAWU forecast while forecasting the MOG intensity threshold over a smaller area. Compared to the FIP, IPA-F forecasted a greater area of high probabilities, but also a larger area of zero probabilities of icing. FIP forecasted a larger area of low probabilities for icing. Given the stronger performance of IPA-F at higher probability thresholds, IPA-F's increased forecast certainty yielded more accurate forecasts as evaluated by the PSS metric.

# 6 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.

# 7 REFERENCES


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

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)

Brown, and G.A. Grell, 2004: A 13-km RUC and beyond Recent developments and future plans. 11th Conference on Aviation, Range, and Aerospace Meteorology, Hyannis, MA, October 2004, American Meteorological Society (Boston).

Brown, B.G., and G.S. Young, 2000: Verification of icing and icing forecasts: Why some verification statistics can't be computed using PIREPs. Preprints, 9th conference on Aviation, Range, and Aerospace Meteorology, Orlando, FL, Sep. 11-15, American Meteorological Society (Boston), 393-398.

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

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

Kay, M.P., C. Lu, S. Madine, J. L. Mahoney, and P. Li, 2009: Detecting cloud icing conditions using CloudSat datasets. Preprints, 23rd Conference on Weather Analysis and Forecasting, 1-5 June, Omaha, NE, Amer. Met. Soc.

Kucera, P. A., B. G. Brown, and A. Holmes, 2007: Calibration of the Forecast Icing Product (FIP) – Icing Potential. Report to the FAA. (DETAILS?!?!)

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.

McDonough, F., 2010: Regional differences in icing clouds across Alaska. *3rd Annual Alaska Weather Symposium, Fairbanks, AK.*, (Conference Proceeding).

Murphy, M. P., 2010: Product Description Document, Graphical Airman's Meteorological Advisory (AAWU AIRMET). Available at: http://aviationweather.gov/static/docs/gairmet/AAWU AIRMETPDD_2010.pdf

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 situ 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.

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).