NOAA Technical Memorandum OAR GSD-64
https://doi.org/10.25923/krbd-2a83

Ceiling and Visibility Analysis Products Assessment

November 2019

Kenneth R. Fenton Jr.
Dana M. Mueller
Tanya R. Peevey
Matthew S. Wandishin
Melissa A. Petty

Earth System Research Laboratory
Global Systems Division
Boulder, Colorado
November 2019

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

A Technical Memorandum OAR GSD-64

Ceiling and Visibility Analysis Products Assessment

Kenneth R. Fenton Jr.[1]
Dana M. Mueller[1]
Tanya R. Peevey[1]
Matthew S. Wandishin[2]
Melissa A. Petty[1,3]

[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
[3]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. The authors would like to thank the following organizations for providing the data needed for the evaluation: NOAA/NWS/MDL for providing the GLMP data and NOAA/NWS/EMC for providing the RTMA data.

UNITED STATES DEPARTMENT OF COMMERCE

Wilbur Ross
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Dr. Neil Jacobs
Acting 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) of the National Oceanic and Atmospheric Administration's (NOAA's) Global Systems Division (GSD) within the Earth System Research Laboratory (ESRL) was tasked by the Federal Aviation Administration's (FAA's) Aviation Weather Research Program (AWRP) to complete an assessment of ceiling and visibility (C&V) analysis products. The purpose of this assessment was to measure the performance of several C&V products in the context of their potential use in the Aviation Weather Center's (AWC) Helicopter Emergency Medical Services (HEMS) tool. The products evaluated in this assessment were the C&V analysis products from the National Ceiling and Visibility Analysis (NCVA), the Gridded Localized Aviation Model Output Statistics Program (GLMP), the Real-Time Mesoscale Analysis (RTMA), and the RTMA Rapid Updates (RTMA-RU). This is the second part of a two-part assessment; the first part analyzed the results from the cool season from January 1, 2018 through April 30, 2018, and the second part of the assessment covered the warm season from May 1, 2018 through August 30, 2018.

Using independent observation data sets (i.e., observations that were not used by the analyses and evaluated during the time period after the analyses are performed and before the next analysis is available) to assess product performance, the following main findings were observed:

- NCVA and GLMP generally outperformed RTMA and RTMA-RU in identifying low ceilings (Figure 1 left/Figure 29), but the highest scoring was not consistent through all flight categories.
- NCVA generally outperformed GLMP, RTMA, and RTMA-RU in identifying poor visibility and while the differences were statistically significant, they were not substantial (Figure 1 right/Figure 46).

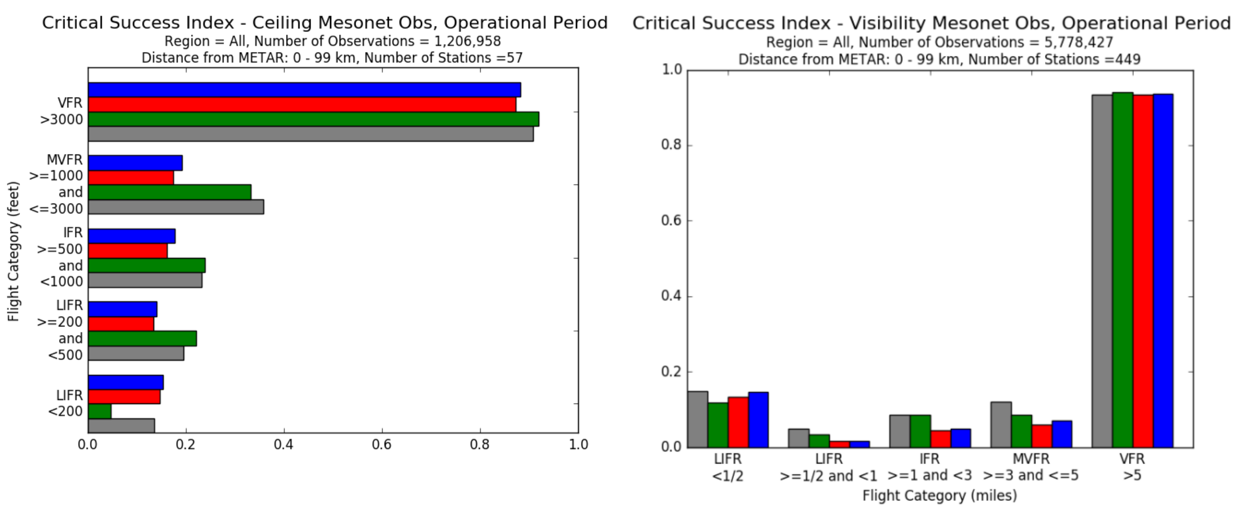


Figure 1. CSIs of the NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against ceiling (left) and visibility (right) mesonet observations during the operational verification period.

Using the much more prevalent dependent data (i.e., observations that were used by the analyses and/or were taken during the same period as observations that were used) to perform verification, it was discovered that:

- NCVA and GLMP outperformed RTMA and RTMA-RU by a wide margin in identifying both low ceilings (Figure 2 left/Figure 22) and poor visibilities (Figure 2 right/Figure 39).
- NCVA generally, though not uniformly, outperformed GLMP slightly in identifying low ceilings (Figure 2 left/Figure 22) and substantially outperformed GLMP in identifying poor visibility (Figure 2 right/Figure 39).

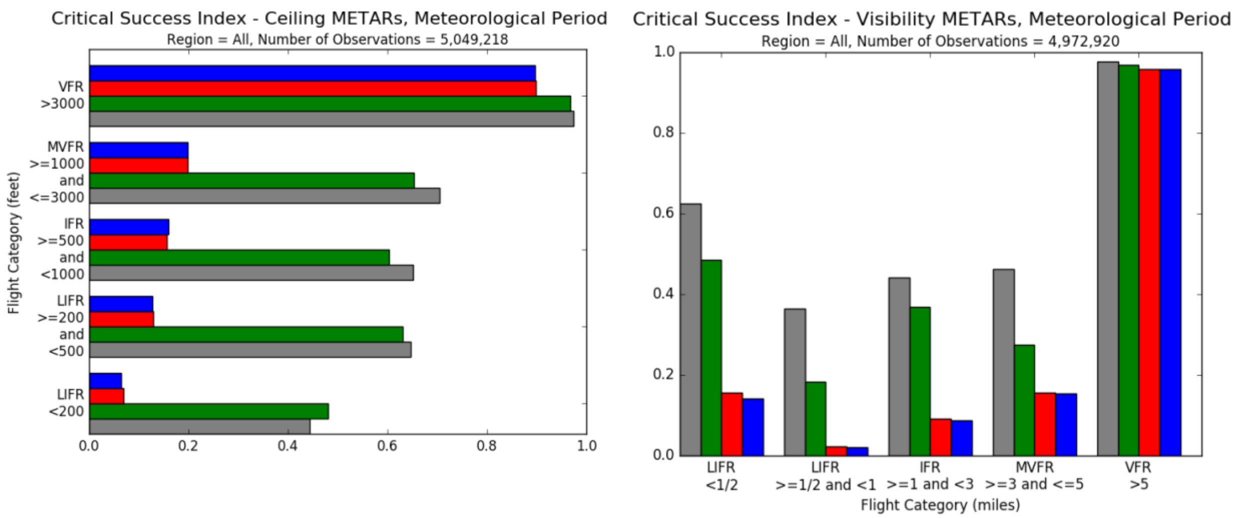


Figure 2. CSIs of NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against ceiling (left) and visibility (right) METARs during the meteorological verification period.

Factors contributing to the differences in product performance include:

- For ceilings, RTMA and RTMA-RU identify more events for all but the least restrictive flight categories compared to NCVA and GLMP (Figure 13) and the observations (Figure 14).
- For visibilities, NCVA and GLMP identify more visibility events for all but the most restrictive flight categories compare to RTMA and RTMA-RU (Figure 32); the observations match closely to the NCVA and GLMP event frequencies (Figure 33).
- RTMA and RTMA-RU event frequencies are typically more consistent with the observations in mountainous terrain, and less consistent over the eastern half of CONUS, with the largest errors along the eastern seaboard and the west coast (Figure 18, Figure 20, Figure 35, Figure 37).
- The lower skill for RTMA and RTMA-RU is the result of both fewer events captured and more false alarms (Figure 22, Figure 39, Figure 72 - Figure 81).

2 Introduction

This National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Global Systems Division (GSD) document outlines the results from the Quality Assessment (QA) Product Development Team's (PDT) assessment of ceiling and visibility (C&V) analysis products. The purpose of this assessment was to measure the performance of several C&V products in the context of their potential use in the Aviation Weather Center's (AWC) Helicopter Emergency Medical Services (HEMS) tool.

The products evaluated in this assessment included the ceiling and visibility analysis products from the Gridded Localized Aviation Model Output Statistics Program (GLMP), the Real Time Mesoscale Analysis (RTMA), and the RTMA Rapid Updates (RTMA-RU). The assessment included an evaluation of product skill, as well as several case studies. In addition to being compared between each other, the products were also compared against the current operational baseline, the National Ceiling and Visibility Analysis (NCVA).

The period for this assessment was May 1, 2018 through August 31, 2018. This is a follow-on to the cool season assessment period, which analyzed results from January 1, 2018 through April 30, 2018.

The assessment made use of mesonet and aviation routine weather reports (METAR) C&V data, mesonet dewpoint depression data, five-minute Automated Surface Observing System (ASOS) data, and satellite data from Geostationary Operational Environmental Satellite 16 (GOES-16). These observations were used to generate quantitative assessments of the quality of the products. The assessment included the following main areas of investigation:

1. Characteristics of the product fields
2. Statistical performance against truth observations
3. Case studies

3 Data

3.1 Products

3.1.1 National Ceiling and Visibility Analysis

The NCVA is an analysis product that uses METAR observations and GOES satellite information to display a real-time graphical output of flight categories. The outputs of the product are ceiling heights and visibility ranges at a 5-km resolution. In addition, the AWC creates a graphic that classifies each location as either the visual flight rules (VFR) or instrument flight rules (IFR) flight category based on the ceiling and visibility values. The product is updated every five minutes and covers the contiguous United States (CONUS). The NCVA maintains strict adherence to the observations at METAR locations and uses an interpolation scheme to estimate the conditions in between METAR stations. The product assimilates observations from the 80 minutes preceding the valid time, takes less than five minutes to run, and is updated every five minutes. A schematic of NCVA's operations timeline is shown in Figure 3. The product valid time is defined as the stated time that the analysis represents. The product available time is defined as the time that the product is published and becomes available to the user.

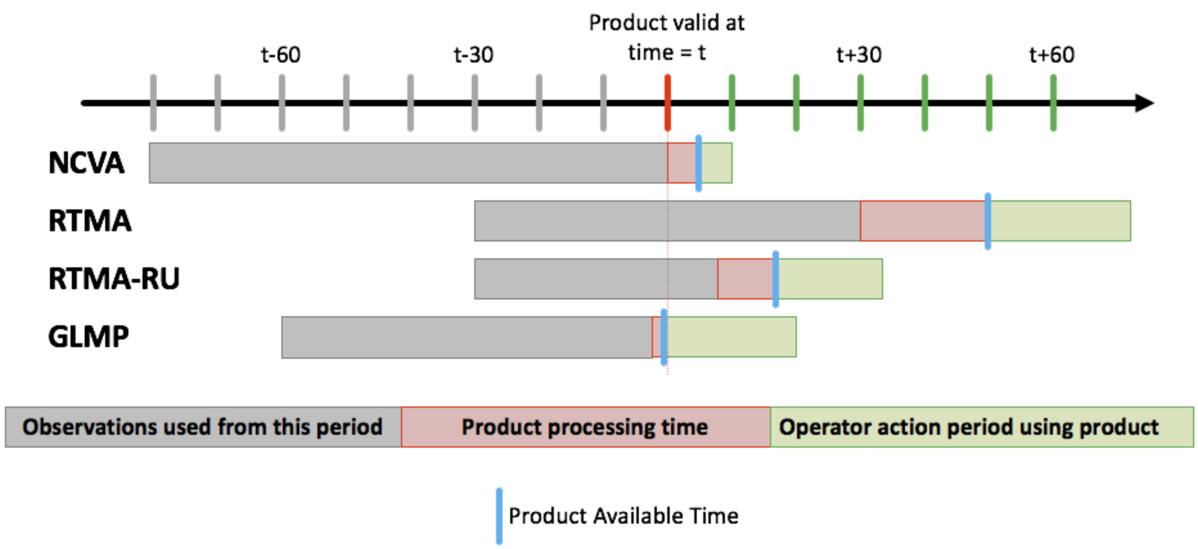


Figure 3. Timeline of the operations schedule for NCVA, RTMA, RTMA-RU, and GLMP.

NCVA separately processes ceiling and visibility fields before combining them into a flight category. First, the algorithm assigns the ceiling height of the nearest-neighboring METAR to each grid point to create an initial ceiling height field. Next, the height is corrected for changes in elevation between the METAR and the grid point. Then, a cloud mask from GOES satellite data is applied, which removes ceilings in locations where the satellite data indicates that clouds are not present. For visibility, the algorithm assigns the visibility of the nearest neighboring METAR to each grid point. The ceiling and visibility fields are then combined such that the flight category at each grid point represents the minimum flight category from the two fields.

3.1.2 Gridded Localized Aviation Model Output Statistics Program (GLMP)

The Gridded Localized Aviation Model Output Statistics (MOS) Program (GLMP) is an algorithm that provides a gridded analysis and forecasts of sensible weather elements. The GLMP produces an analysis every 15 minutes on a grid with a 2.5-km horizontal resolution. The output fields include near-surface variables such as temperature, dewpoint, ceiling height, ceiling probability at three categories, sky cover, visibility, visibility probability at three categories, wind gusts, wind speed, convection, and lightning. GLMP v2.1 was operationally implemented in January 2018.

GLMP uses the Bergthorssen, Cressman, Doos, and Glahn (BCDG) process to extend the Localized Aviation Model Output Statistics Program (LAMP) station-based ceiling and visibility analysis to the National Digital Forecast Database (NDFD) grid (Ghirardelli et al 2015). The LAMP station-based ceiling and visibility analysis is produced by integrating METAR stations and buoy reports that have been quality controlled and approved by the Meteorological Development Laboratory (MDL). The GLMP uses the most recent routine report or special report from each station during the last hour at the XX:30 and XX:45 analysis times. At the XX:00 and XX:15 analysis times, the window for observations is extended an additional 15 minutes to use observations that are up to one hour and 15 minutes prior to the analysis time. The GLMP is run three minutes before the analysis time (i.e., the analysis valid at 1200Z begins to run at 1157Z) and takes two minutes to complete. The analysis is typically available one minute prior to the valid time (i.e., the analysis valid at 1200Z is available at 1159Z). A schematic of GLMP's operations timeline is shown in Figure 3.

3.1.3 Real-Time Mesoscale Analysis

The Real-Time Mesoscale Analysis (RTMA) is an analysis and assimilation system for near-surface weather conditions. The RTMA produces an analysis at the top of every hour on a grid with 2.5-km horizontal resolution. The output fields include temperature, dewpoint, pressure, wind, humidity, visibility, cloud cover, and ceiling height. The analysis is typically available 50 minutes after the analysis time.

RTMA uses the Gridpoint Statistical Interpolation (GSI) system applied in two-dimensional variational mode to assimilate observations to a background model field. The background field is a blend of the HRRR and the NAM-nest downscaled to a 2.5-km grid (Pondeca et al 2015). Observations assimilated by RTMA include those from METARs and mesonets. From the METARs, the RTMA uses the temperature, dewpoint, wind speed, wind direction, wind gusts, pressure, sky cover, ceiling height, visibility, and wave heights, where applicable. RTMA uses a smaller list of data from mesonet sites, including temperature, dewpoint, wind speed, wind direction, wind gusts, and pressure. RTMA does not assimilate ceiling or visibility data from mesonets, making those a source of independent verification data. The algorithm also assimilates GOES sky cover data, but does not yet currently use that data to clear regions of ceilings. Observations that occur within plus or minus 30 minutes of each hour are assimilated (Yang et al. 2017). A schematic of the RTMA operations timeline is shown in Figure 3.

3.1.4 Real-Time Mesoscale Analysis Rapid Updates

The Real Time Mesoscale Analysis with Rapid Updates (RTMA-RU) is an analysis system similar to RTMA, but generates output every 15 minutes. In addition to the XX:15, XX:30, and XX:45 analysis times, the RTMA-RU is run concurrently, but separately from the RTMA at the top of every hour (Myrick 2017). Like the RTMA, the output is on a grid with 2.5-km horizontal resolution and the output variables include temperature, dewpoint, pressure, wind, humidity, visibility, cloud cover, and ceiling height. The analysis became operational in mid-December 2017.

The RTMA-RU assimilates observations from the same sources as the RTMA. If there are multiple observations available within the assimilation window, then the RTMA-RU does not use time interpolation and the observation nearest in time to the analysis time is used to modify the background field in the RTMA-RU. Observations are usually assimilated for a period from 30 minutes before to 8 minutes after the analysis time. The RTMA-RU uses the 15-minute forecast output from the most recent run of the HRRR as background fields for ceiling and visibility, and hourly output for the other surface variables (Yang et al. 2017). The analysis takes approximately 9 minutes to run and is available 17 minutes after the valid time. The operations timeline for RTMA-RU is shown in Figure 3.

3.2 Observations

3.2.1 METARs

METARs are automated and human-produced observations of the meteorological conditions occurring at a surface location. METARs are issued every hour as well as when the meteorological conditions cross predetermined categories. Temperature, dewpoint, pressure, wind speed, wind direction, visibility, and sky conditions are contained in METARs. Data from METARs, in particular sky condition and visibility, are assimilated by all the products in this study. Figure 4 shows the locations of the METARs used in this assessment.

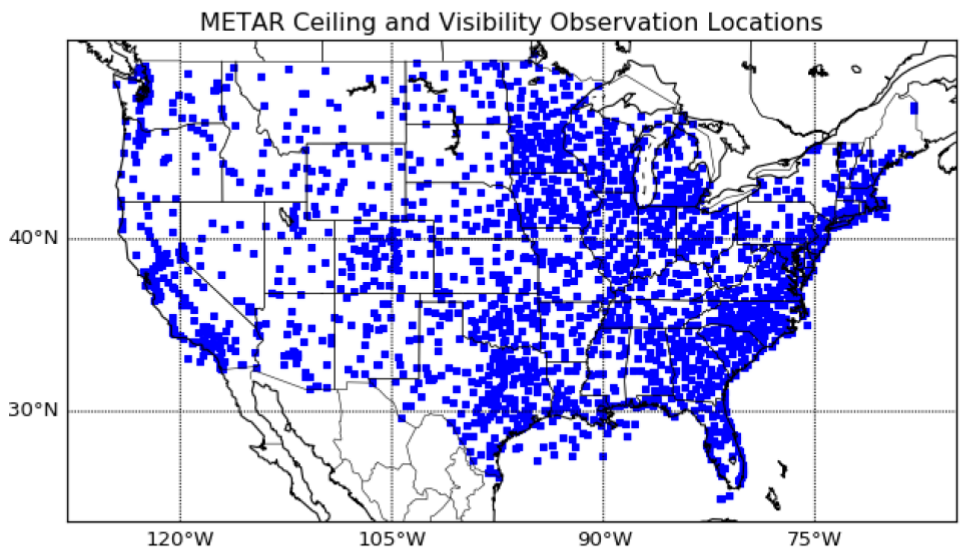


Figure 4. Map showing the locations of the 2,026 METAR stations used in this assessment.

3.2.2 Five-Minute ASOS Data

The Automated Surface Observing System (ASOS) is the set of meteorological sensors that record the atmospheric conditions that are transmitted in most METARs. ASOS stations are constantly updating with weather information and have a more temporally complete record of atmospheric conditions than METARs, which only report once per hour or whenever a predetermined category is crossed. Meteorological data, including visibility and cloud heights, are transmitted every five minutes from a subset of METAR locations (Figure 5) and was available to the QA PDT.

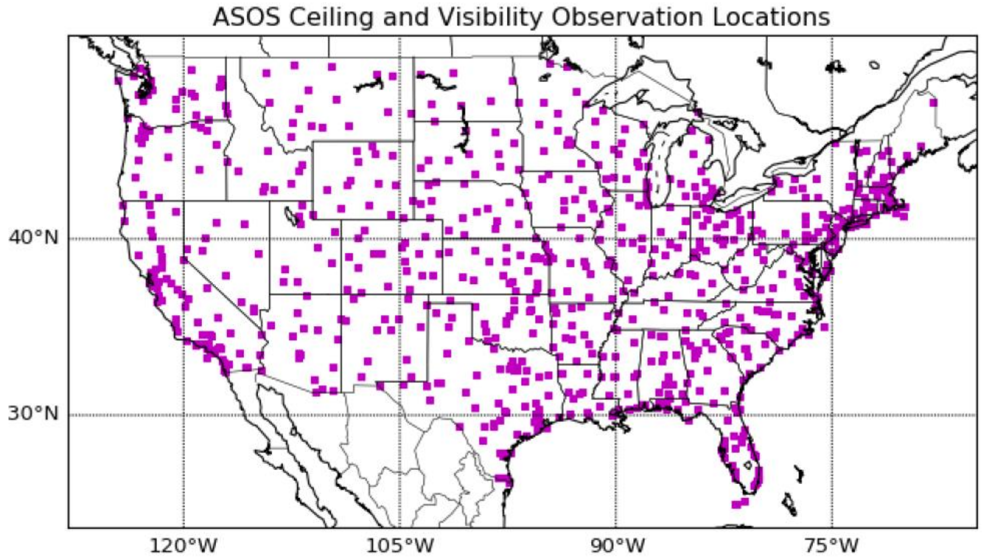


Figure 5. Map showing the locations of the 789 ASOS stations used in this assessment.

3.2.3 Meteorological Assimilation Data Ingest System (MADIS) Mesonet Data

Mesonets are regional networks of automated meteorological observing systems. They are usually operated on the state or local level, with the most common types belonging to a state's Department of Transportation. Mesonet sensors and reporting methods are not held to the same rigid standards as METARs and therefore, the data must undergo a quality control process by the user.

Many mesonets only report temperature and dewpoint, however, a subset of stations also report sky condition and visibility. Research conducted by the QA PDT has shown that some of the sky condition and visibility reports are of high quality and can be used in verification. The first part of the research focused on the ceiling reports from mesonet stations. During a sample period in November 2017, 63 mesonet stations reported ceiling observations. Each mesonet station from that set was then matched to the nearest METAR location. Five pairs of METAR and mesonet stations were located within 2-9 km of each other and had ceiling observations within a time difference of 5 minutes. In total, 1,405 observations were matched between mesonet and METAR ceiling observations. The correlation between the two observation sets was strong with R^2=0.97. Importantly, there were very few occurrences where one observation set had a ceiling below 1,000 ft while the other observation set had a ceiling above 1,000 ft. Overall, the mesonet ceiling observations align well with the METAR ceiling observations and were used to verify product ceilings in locations where METAR observations are not present. Figure 6 shows the locations of the mesonet ceiling stations used in this assessment.

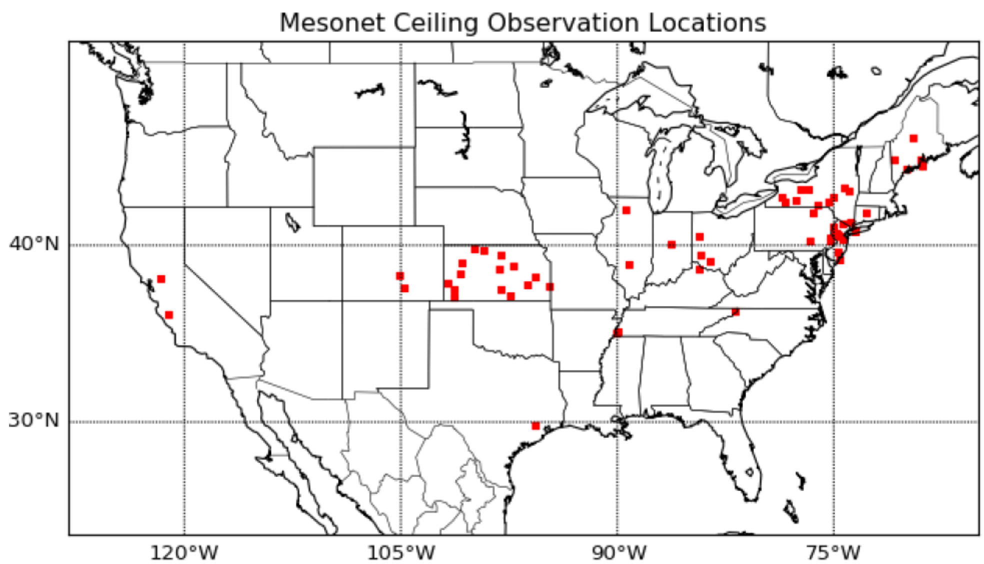


Figure 6. Map showing the locations of the 62 mesonet stations that were used for ceiling verification in this assessment.

The second part of the QA PDT's research focused on the visibility observations. Mesonet observations from a sample period were matched to METAR observations that occurred within 2-km and 5 minutes of one another. A filter was applied to remove mesonet stations that did not report a full range of visibility values, resulting in observation matches between the mesonet and METAR visibilities for 14 station pairs. There was a strong correlation (R^2=0.98) between the two data sets, indicating that the mesonet visibilities are consistent with neighboring METAR visibilities and can be used for verification.

A three-part filter was developed by the QA PDT to ensure that only stations with reasonable distributions of visibility observations were used for verification. The first part of the filter required that a mesonet station report an unrestricted visibility (> 11,000 m) in more than 60% of its reports. The next part required that at least one report occurred in the mid-range of visibility (between 4,000 and 8,100 m). The last part of the filter required the station report at least once per hour with an 80% reliability. The filter was applied to all available mesonet visibility stations, and 457 stations (Figure 7) passed the filter and were used in this assessment.

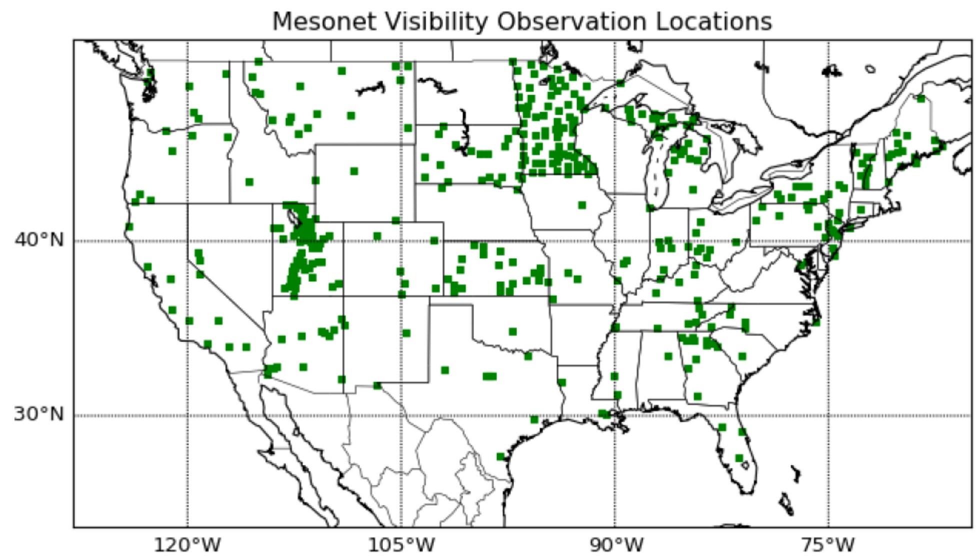


Figure 7. Map showing the locations of the 457 mesonet stations that were used for visibility verification in this assessment.

In addition to the C&V observations from mesonets, the temperature and dewpoint measurements were used to identify areas where reduced visibilities should not be present. A study by the QA PDT found that any location with a dewpoint depression (temperature minus dewpoint) greater than or equal to 3°C had a visibility greater than one mile 99% of the time (Fenton et al. 2017). Although the RTMA does assimilate temperature and dewpoint observations, it does so only for its temperature and humidity analysis fields and does not use that data for ceiling and visibility calculations. This assessment used the mesonet stations shown in Figure 8 to identify locations where visibilities should be greater than one mile because the measured dewpoint depression was greater than or equal to 3°C.

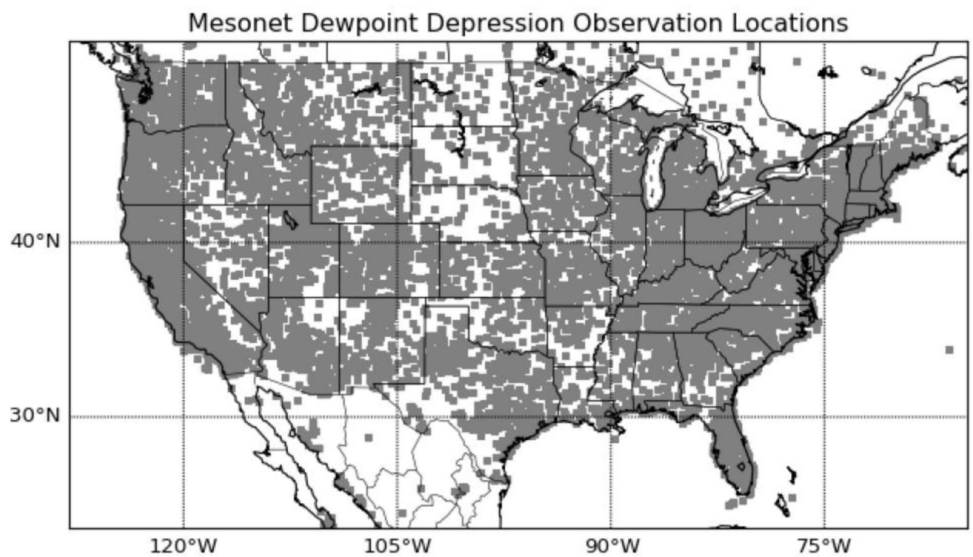


Figure 8. Map showing the locations of the 14,404 mesonet stations reporting dewpoint depression that were used in this assessment.

3.2.4 Satellite

GOES-16, also known as GOES-East, provides satellite imagery of clouds in the visible, near-infrared, and infrared wavelengths. With 16 channels (2 visible, 4 near-IR, and 10 IR) and a suite of derived products, the volume of data that is available is significant (Schmit et al. 2017). Unfortunately for the purposes of verifying C&V, most of the products can only act as bounds or have significant caveats that would not make them useable as direct observations. There is also a timing issue; for example, the Fog and Low Stratus Product could be useful, but because it is not fully implemented with GOES-16 data, it is not possible to evaluate its utility as an observation set yet. Therefore, this assessment only included the GOES-16 Clear Sky Mask product.

The Clear Sky Mask data was available as a provisionally mature product on February 16, 2018, but won't be fully operational until November 2018 (GOES-R 2018). The mask algorithm utilizes over half of the ABI channels to perform temporal, spatial, and spectral tests to create an intermediate 4-level mask for use in downstream products and for the official binary mask. It also utilizes ancillary data from the Global Forecast System (GFS) (snow cover, surface temperature, total precipitable water, tropopause level, etc.), the Community Radiative Transfer Model (clear sky transmittance profile, clear and cloudy radiances, clear and cloudy brightness temperatures, etc.), and dynamic auxiliary data (sun glint, solar zenith angle, etc.) (Harris Corporation 2017).

The Clear Sky Mask product is available both day and night at 2-km nadir resolution with a 1-km mapping accuracy (Harris Corporation 2017). Figure 9 shows the approximate pixel area for a 1-km resolution product, which can be multiplied by four for the pixel area of the 2-km Clear Sky Mask product. The southeast United States has a smaller pixel area and better resolution than the northwest United States, which is due to the satellite's position at 75.2 degrees West longitude. The Clear Sky Mask is available every 5 minutes with a latency of 271 seconds (Harris Corporation 2017). The product is required to achieve 87% correct detection (with good quality pixels that have a local zenith angle less than or equal to 70 degrees) and the provisional release statement indicates that the product reached accuracy specifications when evaluated against the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite data (GOES-R 2018). Finally, the mask is known to have artifacts near the terminator and coastal regions and there is less confidence in classifications over snow-covered areas at high angles and when warm, low clouds are present (Heidinger et al. 2016). Many of these issues are currently being investigated (GOES-R 2018).

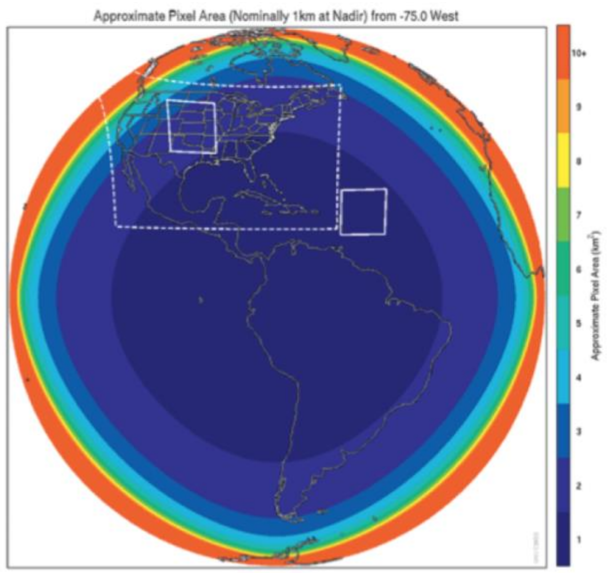


Figure 9. Approximate pixel area for a 1 km resolution product from GOES-16. The dashed box is the CONUS domain and the smaller, solid boxes are the movable mesoscale sectors. Image from Schmit et al. 2017.

4 Methodology

4.1 Stratifications

The assessment focused on well-established flight categories to stratify the results. Five flight categories were used including visual flight rules (VFR), marginal visual flight rules (MVFR), instrument flight rules (IFR), and two divisions of the lower instrument flight rules (LIFR). Table 1 defines each flight category according to the ceiling and visibility categories used in this assessment.

Table 1. Flight categories and their definitions used in the assessment.

Flight	Ceiling	Visibility
VFR	> 3,000 ft	> 5 miles
MVFR	>= 1,000 ft and <= 3,000 ft	>= 3 miles and <= 5 miles
IFR	>= 500 ft and < 1,000 ft	>= 1 mile and < 3 miles
Upper LIFR	>= 200 ft and < 500 ft	>= ½ mile and < 1 mile
Lower LIFR	< 200 ft	< ½ mile

The results were aggregated for all of the CONUS as well as the four geographic regions displayed in Figure 10. The regions were divided up as shown based on similar geography, climate, and to balance the number of observations available between the regions.

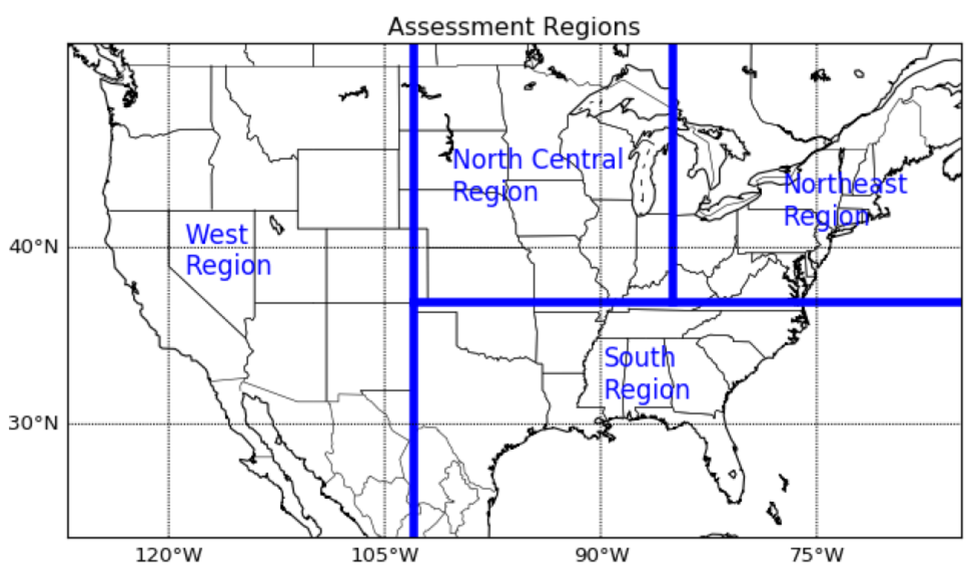


Figure 10. The four geographic regions used in the assessment.

4.2 Techniques to Pair Observations with Product Output

Observations, including METAR, ASOS, and mesonet observations, were matched to the products through the use of two methods:

1. Match observations to products near the product valid time
2. Match observations to products during the operator action period

METARs were assimilated by all of the products in this assessment, therefore, the first method is a check to ensure that the products are modifying their output to closely match the METARs. The first method matched each product to observations in the meteorological verification period, defined as the 15 minutes preceding the product valid time. A schematic of the matching is shown in Figure 11. Under this method, all products with the same valid time were compared against the same set of observations. The observations were matched spatially to the nearest grid point in the output of each algorithm. Only the observation closest in time to the product valid time was used in the event that there were multiple observations from one location in the 15-minute period.

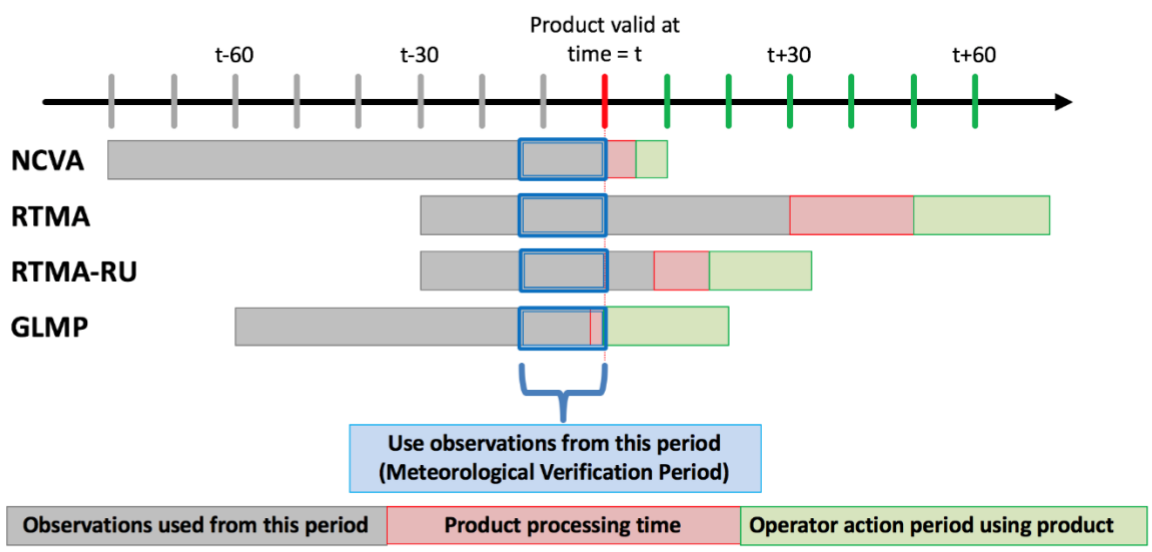


Figure 11. A sample operations timeline for the first method of matching observations to the C&V analysis products in the meteorological verification period. The observation from each location that occurred last in the 15 minutes preceding the product valid time was used in this method.

A second method was employed to better match how each product could be used in an operational setting by accounting for the latency between the valid time of each product and when it becomes available to the user. As mentioned in the product descriptions section, the time difference between the valid time and available time is typically five minutes for NCVA, one minute for GLMP, 50 minutes for RTMA, and 17 minutes for RTMA-RU. The second method compared products to observations that occurred during the operation verification period, defined as the interval of time from when one product becomes available to the operator until the next product becomes available. Contrary to the method that used a single observation during the meteorological verification period, the operational verification period compared each product to all of the observations that occurred within the period. For example, if the RTMA-RU valid at 1200Z became available at 1217Z, then that output will be compared to any observations with times between 1217Z and 1232Z. A sample timeline is shown in Figure 12.

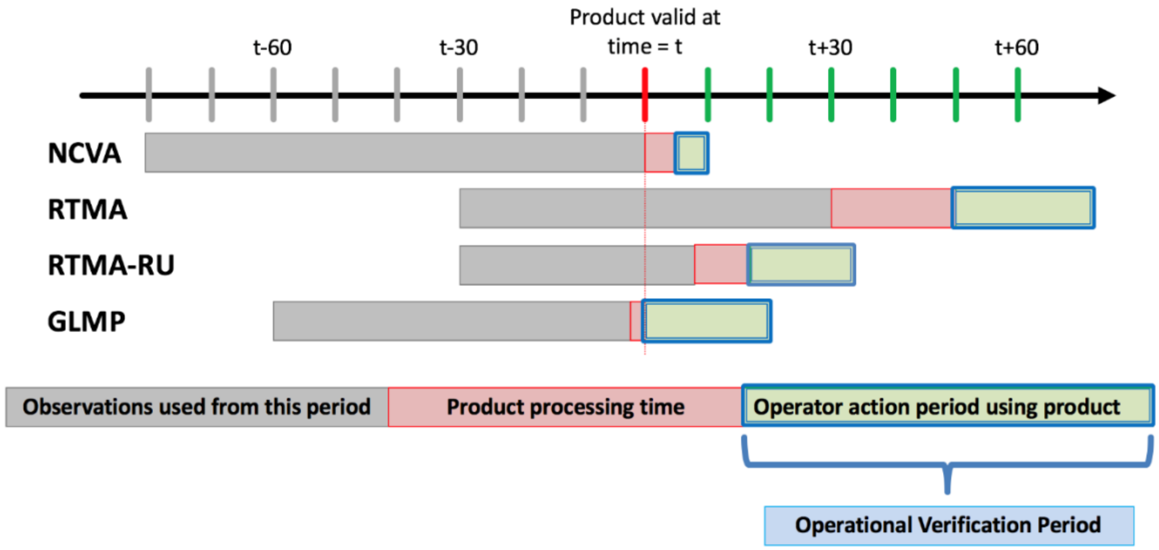


Figure 12. A sample operations timeline for the second method of matching observations to the C&V analysis products in the operational verification period. Observations that occurred after the product processing time and until the updated product became available were used in the operational comparison method.

4.3 Evaluations

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

Table 2. Contingency table matching product outputs to observations.

		Observation	
		Yes	No
Product Output	Yes	Hit	False Alarm
	No	Miss	Correct No

Table 3. Definitions and descriptions of each skill score statistic.

Statistic	Definition	Description
POD:	$\frac{Hits}{Hits + Misses}$	Probability of Detection, proportion of all observed events that are correctly forecast to occur, in this case, of detecting ceilings or visibility at a specific flight category
POFD:	$\frac{False\ Alarms}{False\ Alarms + Correct\ No's}$	Probability of False Detection, proportion of all observed non-events that are mistakenly forecast to be events
FAR:	$\frac{False\ Alarms}{False\ Alarms + Hits}$	False Alarm Ratio, proportion of all forecast events that are incorrectly identified
CSI:	$\frac{Hits}{Hits + False\ Alarms + Misses}$	Critical Success Index, ratio of correctly forecast events to the number of events plus the number of false alarms; also known as the Threat Score (TS)
PSS:	$POD - POFD$	Pierce Skill Score; also known as the True Skill Score

4.3.1 Field Distributions

Distributions were created for each ceiling and visibility product binned by the flight categories presented in Table 1. The distributions were calculated as normalized frequencies by summing the number of occurrences within a flight category over the assessment period and then dividing by the total number of occurrences in any flight category. Only grid points that were within the intersection of the geographic coverage area of all four products were counted.

In addition to this first type of distribution, a second set of distributions was created that was event equalized to METAR observations. Product output was only counted at the grid point nearest in space and time (from the meteorological verification period) to a METAR.

4.3.2 Distribution Maps

Spatial distributions were created that show the frequency of each product by flight category. Again, two types of distributions were generated; the first type used product counts from the intersection of the geographic coverage over the full assessment period, while the second type was event equalized to the METARs as described in the Field Distributions Section. In addition to the distribution maps of the second type, differences maps were constructed by subtracting the frequency of each METAR from the frequency of each product at that location, and was performed for each flight category.

4.3.3 Skill Score Calculations

4.3.3.1 METARs

METAR observations were assimilated by all of the products in this assessment, therefore, there should be a very high correlation between METAR observations and the product outputs. The meteorological verification period was used to check this correlation by measuring the POD, FAR, CSI, and PSS (Table 3) for each product compared against METARs that are nearest and prior to the product valid time. The operational verification period was used to measure the POD, FAR, CSI, and PSS (Table 3) for each product compared against METARs available during the operator action period. The resulting statistics here should be lower than in the first method and more representative of the skill of the products when used in operations.

Lower and upper bounds to the CSI were calculated at the 95% confidence level using a bootstrap method. The bootstrap method employed for this assessment sampled the total number of observations over 1,000 iterations with replacement. The 95% confidence level was calculated by taking the values at the 2.5% and 97.5% categories from the distribution of the iterations.

4.3.3.2 ASOS

The five-minute ASOS data is not assimilated by the products, but is not fully independent either because of temporal auto-correlations in the ceiling and visibility fields. Therefore, the ASOS data is used for verification in the same manner as described above for METARs.

4.3.3.3 Mesonet Ceiling and Visibility Data

The ceiling and visibility observations from mesonet sites are independent data that was not assimilated by any of the products. The skill scores from Table 3 were calculated against the mesonet observations for the both the meteorological and operational verifications periods. Lower and upper bounds to the CSI were calculated at the 95% confidence level as described above in the METAR section.

4.3.3.4 Mesonet Dewpoint Depression Data

The mesonet dewpoint depression data was used as a bound for locations where low visibilities should not exist. Only the negative condition in the observation field was used to calculate a false positive rate, which is defined the same as the false alarm ratio in Table 3.

4.3.3.5 Satellite

The GOES-16 clear sky mask was used as a bound for areas where clouds do not exist and consequently, locations where low ceilings should not be found. Similar to the mesonet dewpoint depression data, only the negative condition in the observation field was used, resulting in the calculation of a false positive rate for ceilings. The domain used for this assessment covered most of the CONUS except for parts of

California, Arizona, Nevada, Oregon, and Washington which were removed due to parallax issues near the edge of the image.

5 Results – Ceiling

5.1 Distributions

The frequency of lower ceilings decreased in the warm season as compared to the cool season, as expected. Figure 13 shows the warm season frequencies of each product at all of the restricted flight categories. The frequencies of each product in the MVFR, IFR, and upper LIFR categories were roughly half the rate in the warm season as they were during the cool season. NCVA and GLMP had higher frequencies than the RTMA products at the MVFR category, but the RTMA products had higher frequencies at all of the lower categories. The relative differences seen between the products were smaller than those observed during the cool season. Similar to the cool season, GLMP again had a much lower frequency of LIFR ceilings than the other products.

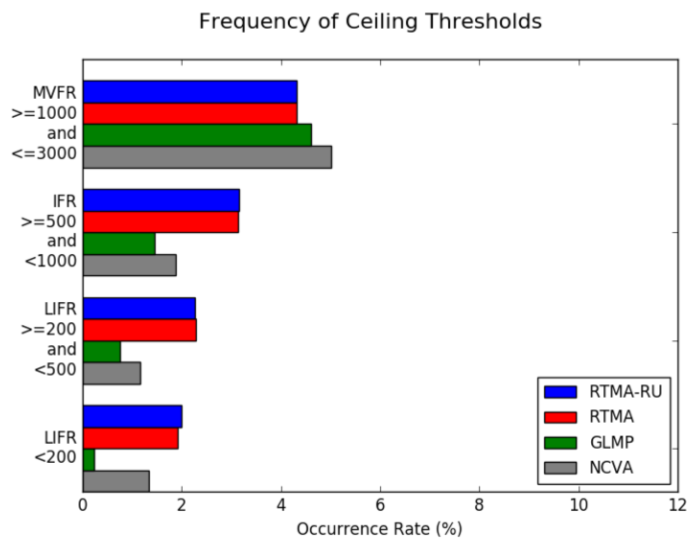


Figure 13. Distribution of warm season ceiling heights by aviation category. Products were equalized to the same geographic boundaries and frequencies were normalized to the number of grid points for each product.

The distributions were then recalculated for each product as event-equalized with METARs. Figure 14 shows the resulting distribution for each product plotted with the distribution from the METARs. Similar to the full distributions above, the warm season rates for the MVFR, IFR, and upper LIFR categories were about half the value as those from the cool season. Again, the NCVA was the closest match to the METAR frequencies at the MVFR, IFR, and upper LIFR categories, while the GLMP was the closest match at the lower LIFR category. The relative differences between the products was similar to the full distributions, except at the lower LIFR threshold where the NCVA had a much lower frequency when event equalized to the METARs. The higher rate seen in the full distribution was likely due to the addition of possible terrain obscurations that were identified by the NCVA in areas away from the METAR locations.

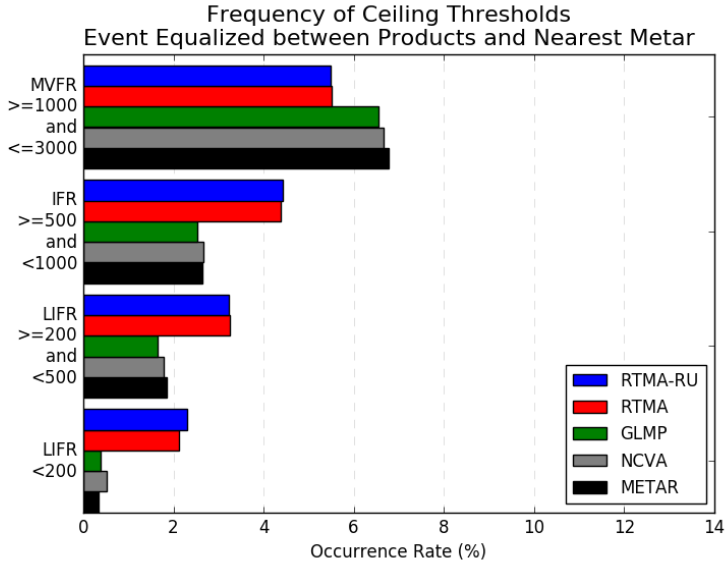


Figure 14. Distribution of ceiling heights by aviation category for the four products and METARs. The distribution was taken from the product point nearest to each METAR site and event equalized.

5.2 Geographic Distributions

Frequencies of each product were calculated at every grid point and plotted for their common domains. The geographic distributions of the RTMA and RMTA-RU were similar so only the RTMA-RU will be presented in the following figures. Figure 15 shows the geographic distributions of the products and the METARs at the MVFR flight category. As was seen in the distributions above, the NCVA and GLMP have similar frequencies that are higher than those of RTMA-RU. The differences are most noticeable in the northeast, southern Texas, and along the west coast where the NCVA and GLMP products have darker shading than the RMTA-RU.

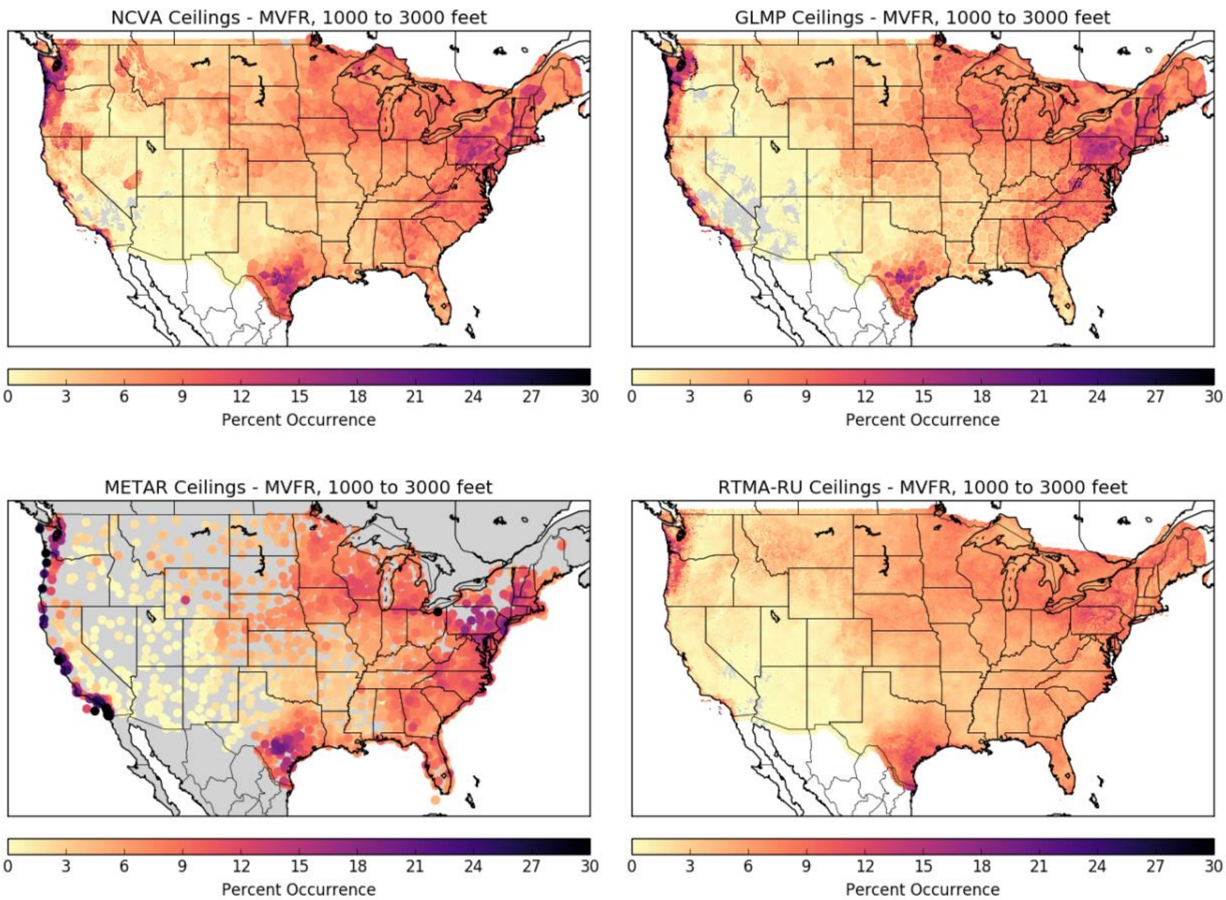


Figure 15. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the MVFR ceiling flight category.

In addition to the general geographic distributions, the distribution of each product at every METAR location was calculated. The METAR frequency was then subtracted from each product frequency at every METAR location to create the difference maps shown in Figure 16. The difference maps provide a quantification of the difference between each product frequency and the METAR frequency at every METAR location. The difference maps show that NCVA and GLMP were both close matches to the METARs at most locations. Both products had a mix of frequencies that were slightly above or below those at the METAR locations and a general tendency for slightly higher frequencies in the southeast and along coastal California. The RTMA-RU frequencies were overall too low as compared to the METARs, but it also had a mix of higher and lower frequencies. The strongest differences were seen through the Appalachians and the west coast, where the RTMA-RU frequencies were lower than those of the METARs. Additionally, the RTMA-RU frequencies were too high over Florida and coastal Texas.

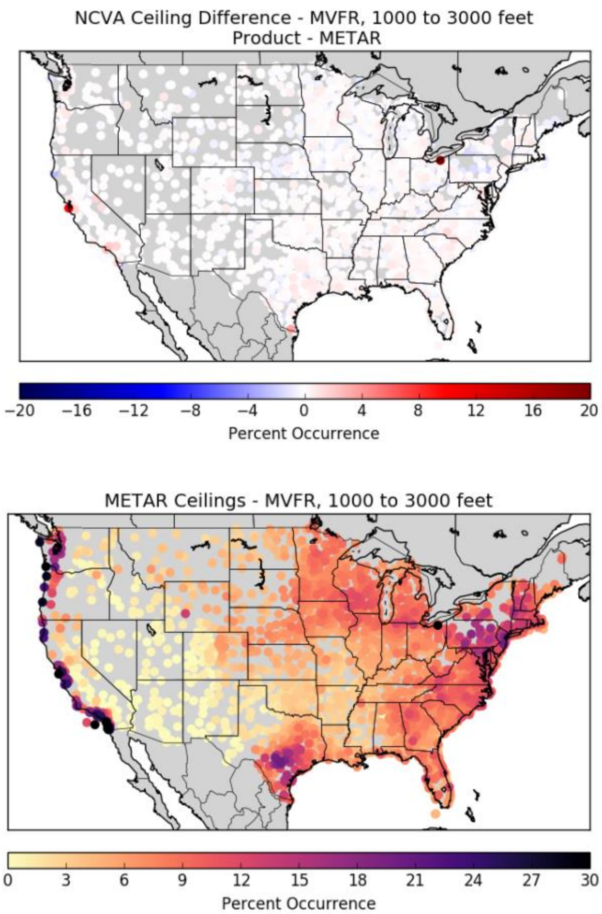


Figure 16. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the MVFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

The relative differences among the products and between the products and the METARs were similar at the IFR and upper LIFR categories, so only the geographic distributions for the upper LIFR category are shown (Figure 17). The IFR results can be found in the Appendix. The NCVA and GLMP had similar frequencies, although the NCVA had higher occurrence rates at this category over the eastern seaboard than the GLMP. The RTMA-RU was generally higher than both of the other products, especially over the eastern half of the country.

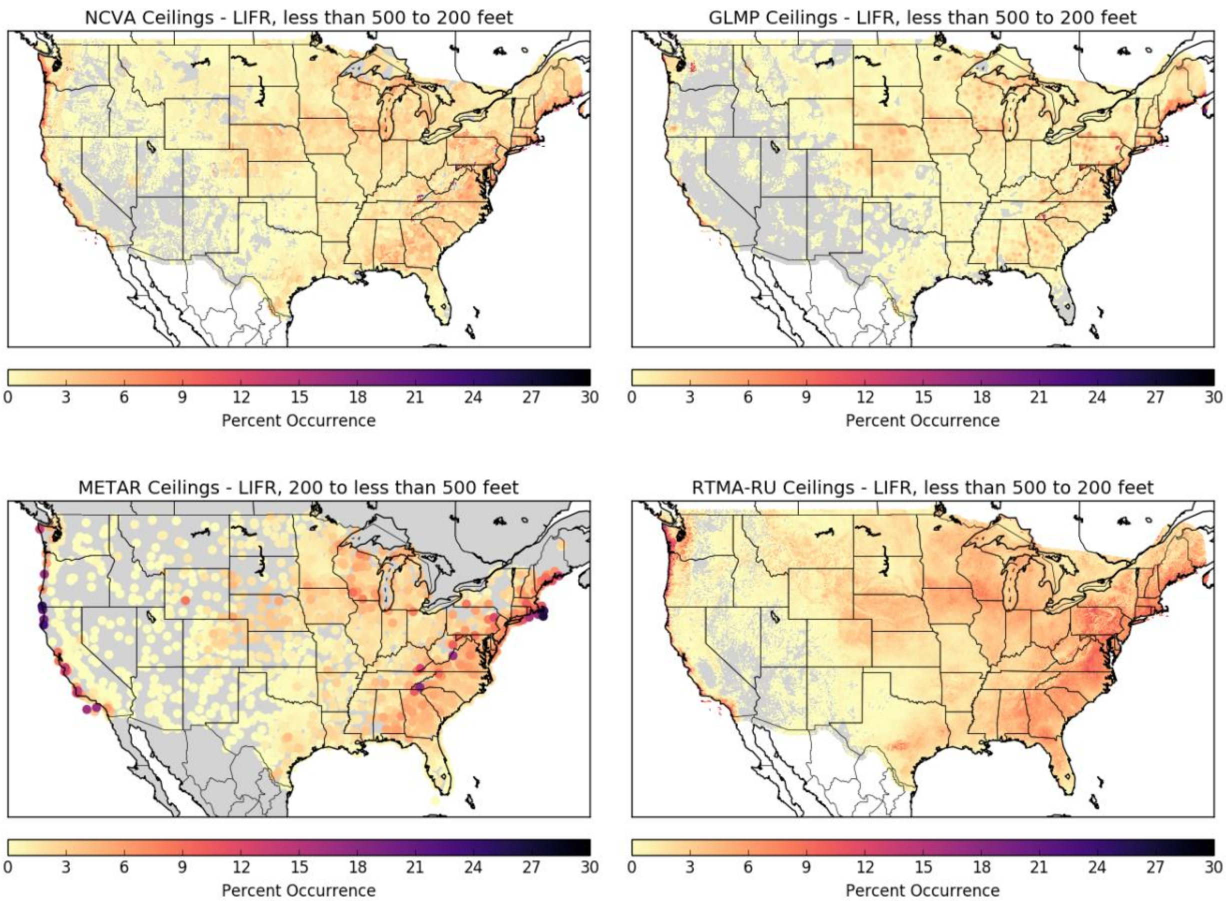


Figure 17. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the upper LIFR ceiling flight category.

The difference maps for the upper LIFR category are show in Figure 18. The NCVA was the closest match to the METARs and the GLMP matched well with the exception of having lower frequencies than the METARs through the Appalachians. The RTMA-RU had frequencies that were higher than the METARs over the west coast and across most of the eastern US, with a few exceptions along coastal New England and in the Appalachians.

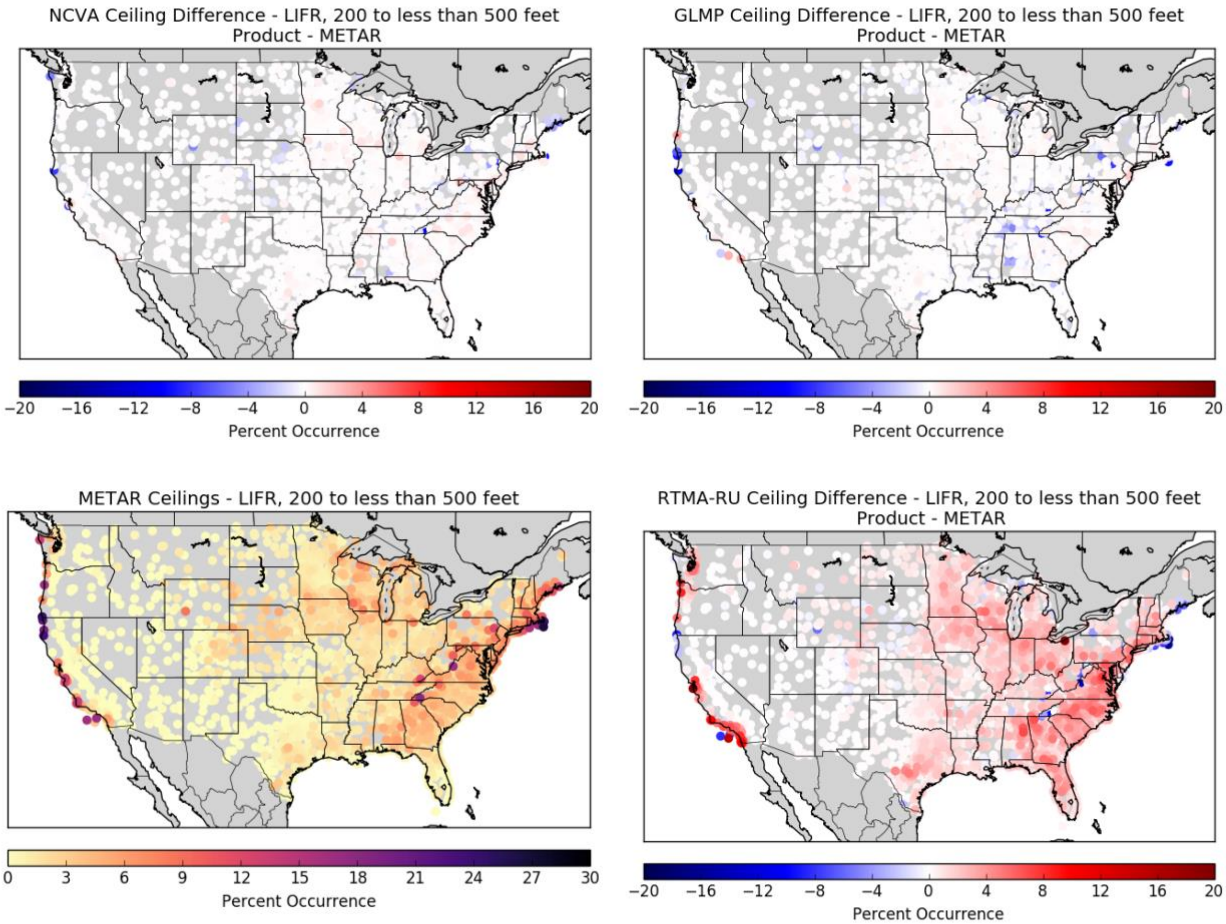


Figure 18. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the upper LIFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

Figure 19 shows the geographic distributions for each product along with the METARs at the lower LIFR flight category. As was seen in the bar graph distribution, the GLMP has overall lower frequencies at this category than NCVA and RTMA-RU. In addition to numerical output of ceiling height, NCVA identified some grid points as possible terrain obscurations when a nearby ceiling intersected with higher terrain. For this assessment, the possible terrain obscurations were treated as a ceiling height of zero feet because a pilot would have to exercise extreme caution when flying into an area analyzed in this manner. Many of the higher frequencies of low ceiling on the NCVA map are areas of high terrain that were regularly identified as possible terrain obscurations in the NCVA output. RTMA-RU had higher frequencies over much of the eastern half of the country and in particular, higher frequencies over the northern Great Lakes.

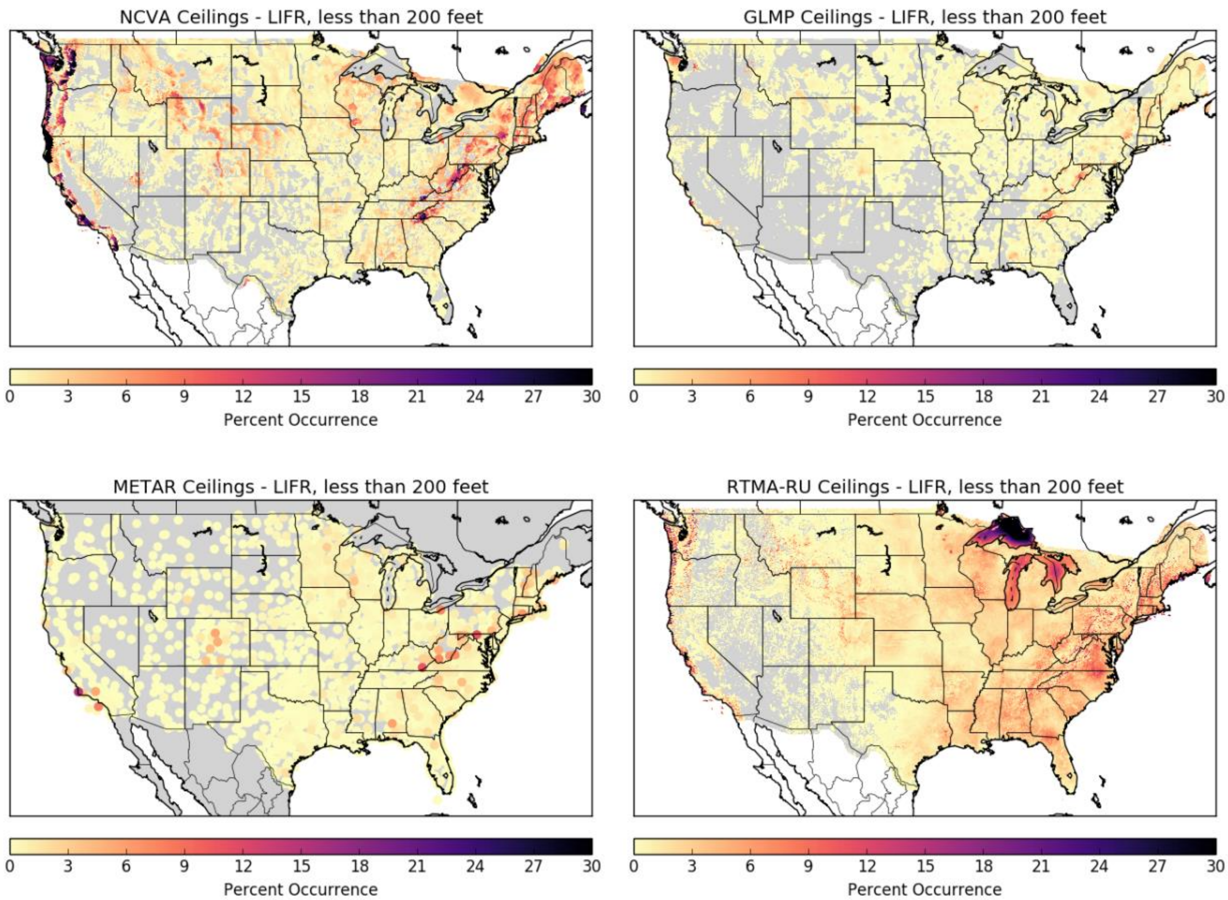


Figure 19. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the lower LIFR ceiling flight category.

The difference maps for the lower LIFR category are shown in Figure 20. As was seen in the bar graph distributions, the GLMP most closely matched the METAR observations at this category, followed by NCVA. NCVA had frequencies that were mostly too high along the eastern seaboard. RTMA-RU had higher frequencies than the METARs across much of the country with the exception of the mountain west. In particular, the RTMA-RU frequencies were much higher than observations in the mid-Atlantic region and along the west coast.

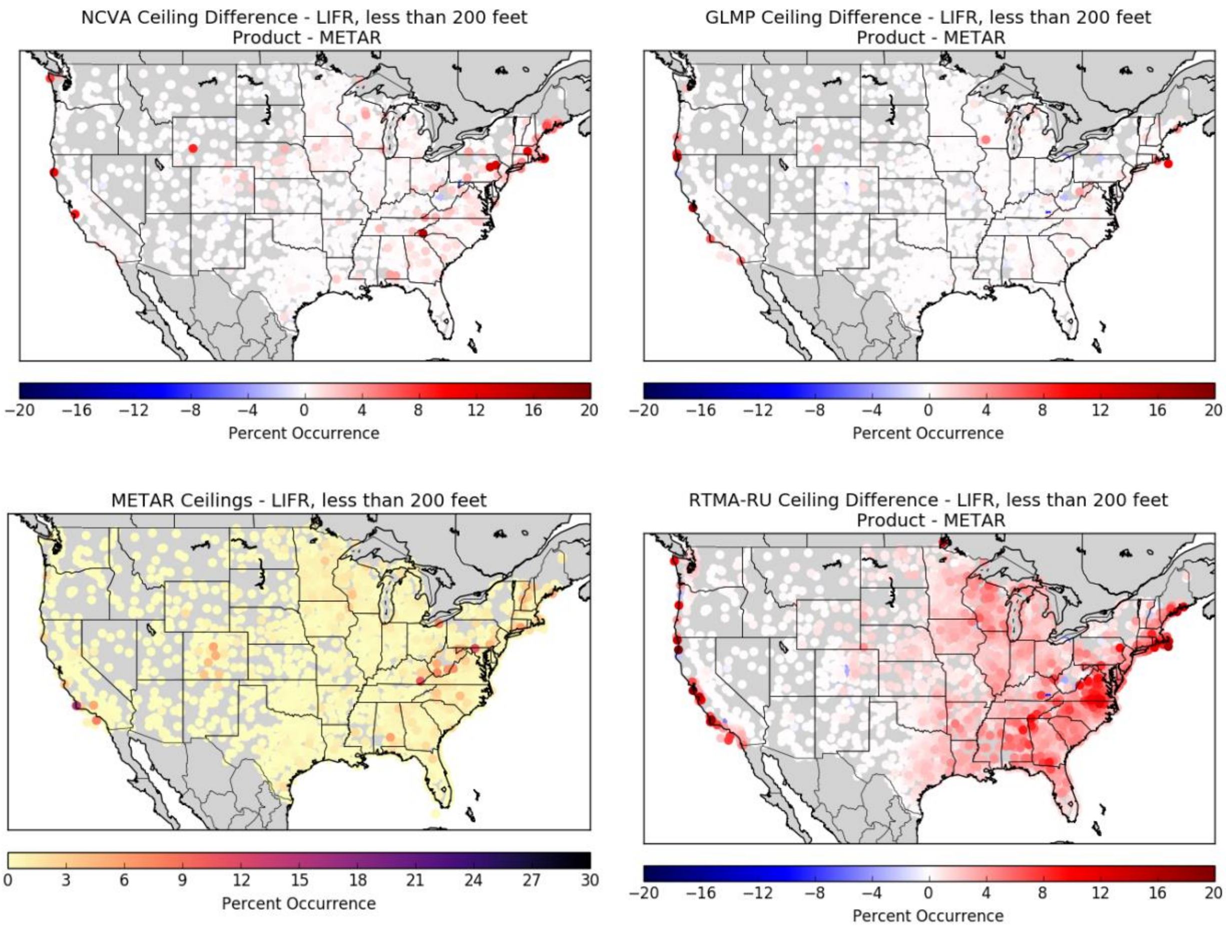


Figure 20. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the lower LIFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

5.3 Conditional Distributions

The distribution of each product was also calculated contingent on the category of the METAR observations, which will be referred to as conditional distributions. Conditional distributions show the percentage of time that the product output resided in each flight category when the METAR observations were in a given category. The results were normalized to each METAR flight category making the percent occurrence of a particular product sum to 100% when aggregated over all five flight categories (each vertical column in a graph sums to 100%). The results of the conditional distributions for the NCVA, GLMP, and RTMA-RU are presented in Figure 21 along with an example of a perfect product's conditional distribution. As an example, a perfect product would be in the VFR category 100% of the time when the METAR observations were VFR, in the MVFR category 100% of the time when the METAR observations were in the MVFR category, and so on along the diagonal line from the lower left to the upper right of the graph, as seen in the lower left distribution. NCVA was in the correct category the highest percentage of the time, followed by GLMP. Both NCVA and GLMP were most often within one category of the observations (boxes along or adjacent to the diagonal). RTMA-RU was more likely to be in the correct category when the observations were VFR and lower LIFR, but spread its analyses across four or five flight categories (relatively high percentages in the incorrect flight categories) when the observations were in the MVFR, IFR, and upper LIFR categories. All three products had slightly elevated percentages of being in the VFR category when the observations were lower LIFR (lower right corner of each graph).

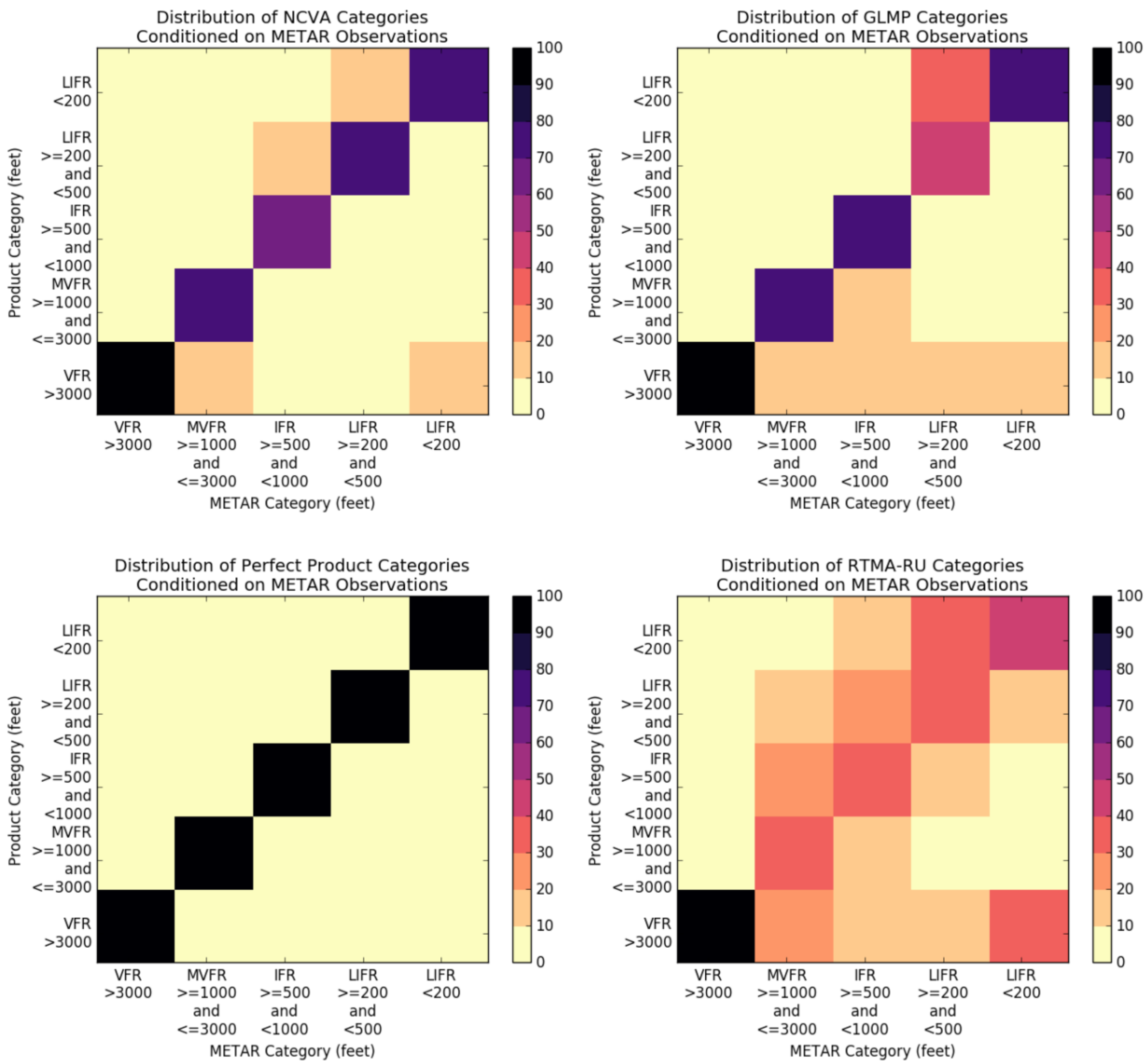


Figure 21. Conditional distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right), conditioned on the flight category of the METAR observation. An example of a perfect conditional distribution (lower left) is shown for reference.

5.4 Meteorological Verification Period

5.4.1 Skill Scores Measured Against METAR

Figure 22 shows the POD, FAR, PSS, and CSI of each analysis product measured against METARs during the meteorological verification period. NCVA and GLMP had higher PODs, PSSs, and CSIs than RTMA and RTMA-RU at every flight category. NCVA and GLMP were able to achieve higher PODs than RTMA and RTMA-RU despite identifying lower occurrence rates at the IFR, upper LIFR, and lower LIFR categories. The PODs for NCVA and GLMP generally decreased with lower ceiling categories except for NCVA at the lower LIFR category. RTMA and RTMA-RU did not decrease evenly like the other two products and had their lowest PODs at the MVFR and upper LIFR flight categories. In addition, the FARs for NCVA and GLMP were lower than RTMA and RTMA-RU in every flight category and generally increased at lower ceiling categories. The PSSs were very similar to the PODs because of the large sample size of over 5,000,000 observations with a high percentage of correct no's in the score calculations for the flight restricted categories. NCVA and GLMP had the two highest CSI scores at every flight category. NCVA had the highest score at the VFR, MVFR, IFR, and upper LIFR categories, while GLMP had the highest score at the lower LIFR category. The lower and upper bounds to the CSI were calculated for the 95% confidence interval. All of the difference between the NCVA, GLMP, and the RTMA products were statistically significant at the 95% confidence level. CSI was selected as the most representative statistic because it does not include the large group of correct no's and will be the primary statistic discussed for the remaining data sets.

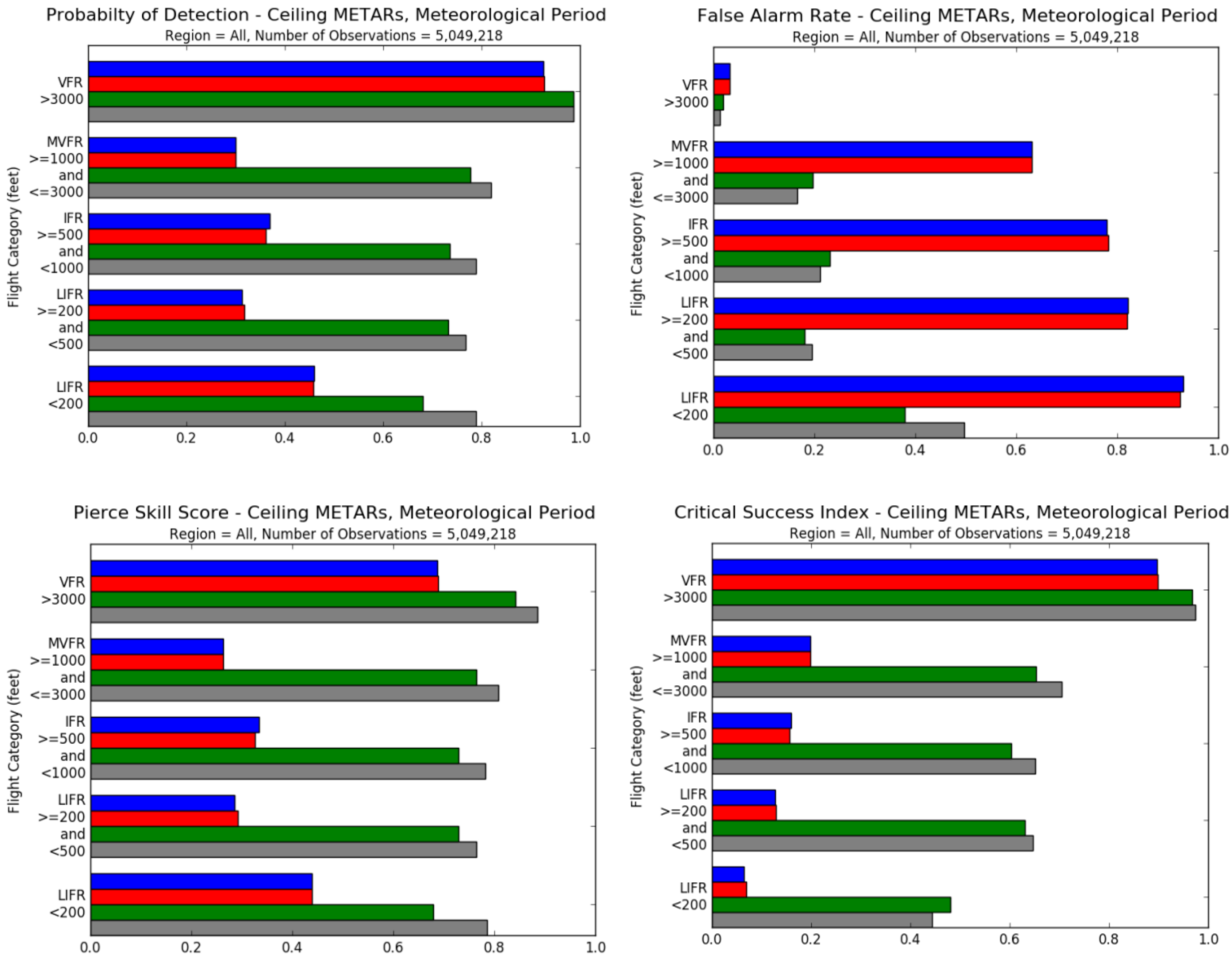


Figure 22. The POD (upper left), FAR (upper right), PSS (lower left), and CSI (lower right) of the ceiling products measured against METARs during the meteorological verification period. NCVA is gray, GLMP is green, RTMA is red, and RTMA-RU is blue.

		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.9565*	**0.9567**	*0.9569*
	RTMA	*0.9575*	**0.9577**	*0.9578*
	GLMP	*0.9683*	**0.9685**	*0.9686*
	NCVA	*0.9756*	**0.9757**	*0.9759*
MVFR	RTMA-RU	*0.1523*	**0.1539**	*0.1554*
	RTMA	*0.1538*	**0.1554**	*0.1570*
	GLMP	*0.2718*	**0.2739**	*0.2758*
	NCVA	*0.4602*	**0.4622**	*0.4643*
IFR	RTMA-RU	*0.0841*	**0.0860**	*0.0878*
	RTMA	*0.0893*	**0.0911**	*0.0930*
	GLMP	*0.3650*	**0.3681**	*0.3714*
	NCVA	*0.4371*	**0.4404**	*0.4437*
Upper LIFR	RTMA-RU	*0.0191*	**0.0207**	*0.0223*
	RTMA	*0.0204*	**0.0219**	*0.0234*
	GLMP	*0.1772*	**0.1820**	*0.1867*
	NCVA	*0.3588*	**0.3645**	*0.3693*
Lower LIFR	RTMA-RU	*0.1393*	**0.1418**	*0.1443*
	RTMA	*0.1528*	**0.1555**	*0.1582*
	GLMP	*0.4797*	**0.4850**	*0.4903*
	NCVA	*0.6178*	**0.6235**	*0.6287*

Figure 23. The lower and upper bounds for the 95% confidence interval of the CSI for ceiling products measured against METARs during the meteorological verification period.

5.4.2 Skill Scores Measured Against ASOS

The ASOS data came from a subset of the METAR locations (789 of the 2,026 sites) that reported at a five-minute frequency. The XX:00, XX:15, XX:30, and XX:45 ASOS reports were matched to the NCVA, GLMP, and RMTA-RU products with the same valid time. The hourly updating RMTA product was not used in this calculation in order to take advantage of the sub-hourly reports that make the ASOS data valuable. Figure 24 shows the CSIs measured against the ASOS data for NCVA, GLMP, and RTMA-RU. The results are similar to those measured against the METARs, with NCVA and GLMP scoring significantly higher than RTMA-RU at all flight categories. NCVA was the highest scoring product at the VFR, MVFR, and IFR categories, and GLMP was the highest scoring at the upper and lower LIFR categories. POD, FAR, and PSS results can be found in the Appendix.

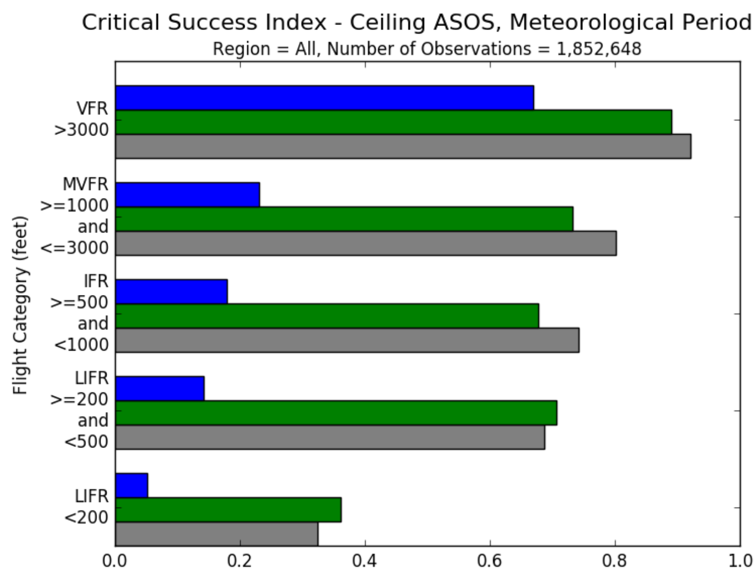
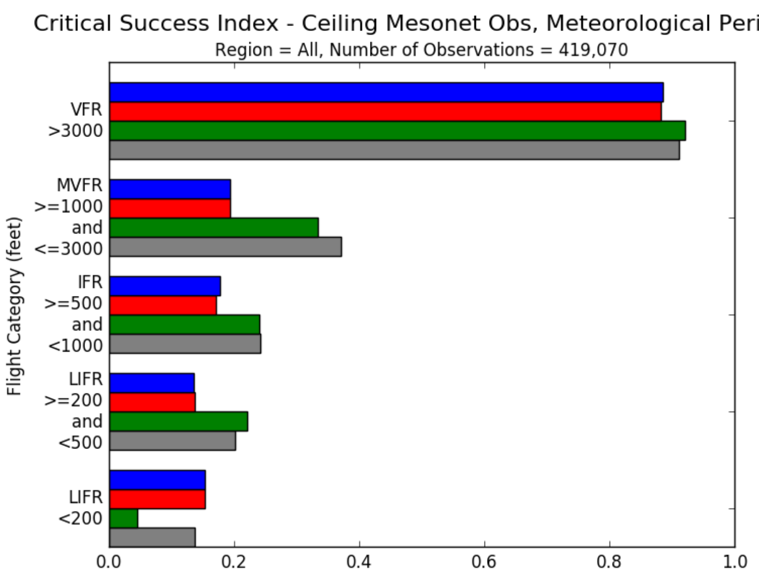


Figure 24. CSIs for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS ceiling observations during the meteorological verification period.

5.4.3 Skill Scores Measured Against Mesonet

Statistical scores were also computed for the mesonet observations, with the full set of scores available in the Appendix. Figure 25 shows the CSIs of each product measured against the independent data from the mesonet ceiling locations. The mesonet data was drawn from 62 reporting locations that had a higher station density in the Northeast, Ohio River Valley, and Kansas. The Northwest, Southwest, and Southern regions of the CONUS were underreported in this data set. The mesonet results were similar in pattern to the METAR results at the VFR, MVFR, IFR, and upper LIFR flight categories. NCVA and GLMP again had CSIs higher than RTMA and RTMA-RU at these categories, with NCVA performing best in the MVFR and IFR categories, while GLMP scored highest in the VFR and upper LIFR categories. NCVA and GLMP had large reductions in skill away from the METAR sites, but their skill still remained above that of the RTMA products at the upper four flight categories. NCVA and GLMP outperformed RTMA and RTMA-RU at the IFR and upper LIFR categories despite analyzing ceilings at these categories roughly 1% less frequently than the RTMA products (Figure 13). RTMA and RTMA-RU had slightly higher scores than NCVA at the lower LIFR category, with GLMP performing worst in this category. GLMP's low score at this category was likely due to its much lower frequency than the other products (Figure 13). The differences between NCVA and GLMP were statistically significant at VFR, MVFR, and upper LIFR flight categories. The differences between NCVA and GLMP, and between RTMA and RTMA-RU, were not significant at the IFR and lower LIFR categories, respectively. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.8842*	**0.8852**	*0.8861*
	RTMA	*0.8816*	**0.8827**	*0.8838*
	GLMP	*0.9209*	**0.9218**	*0.9226*
	NCVA	*0.9113*	**0.9123**	*0.9132*
MVFR	RTMA-RU	*0.1892*	**0.1930**	*0.1968*
	RTMA	*0.1895*	**0.1930**	*0.1965*
	GLMP	*0.3291*	**0.3335**	*0.3380*
	NCVA	*0.3659*	**0.3704**	*0.3747*
IFR	RTMA-RU	*0.1727*	**0.1773**	*0.1813*
	RTMA	*0.1664*	**0.1708**	*0.1748*
	GLMP	*0.2348*	**0.2407**	*0.2461*
	NCVA	*0.2361*	**0.2419**	*0.2477*
Upper LIFR	RTMA-RU	*0.1315*	**0.1361**	*0.1408*
	RTMA	*0.1321*	**0.1365**	*0.1408*
	GLMP	*0.2154*	**0.2220**	*0.2284*
	NCVA	*0.1956*	**0.2013**	*0.2074*
Lower LIFR	RTMA-RU	*0.1478*	**0.1539**	*0.1595*
	RTMA	*0.1479*	**0.1541**	*0.1595*
	GLMP	*0.0413*	**0.0461**	*0.0507*
	NCVA	*0.1307*	**0.1371**	*0.1436*

Figure 25. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet ceiling observations during the meteorological verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

5.4.4 Skill Scores Measured Against the GOES-16 Clear Sky Mask

The GOES-16 clear sky mask data was used to calculate a false positive rate for each algorithm. The false positive rate is the frequency that each product analyzed a ceiling of 3,000 ft or lower when the clear sky mask identified no clouds to be present, or put another way, the rate at which the product does not correctly identify VFR flight conditions in areas with no clouds present. The GLMP had the lowest false positive rate (lower scores indicate better performance), followed by the NCVA and then the RTMA-RU and RTMA.

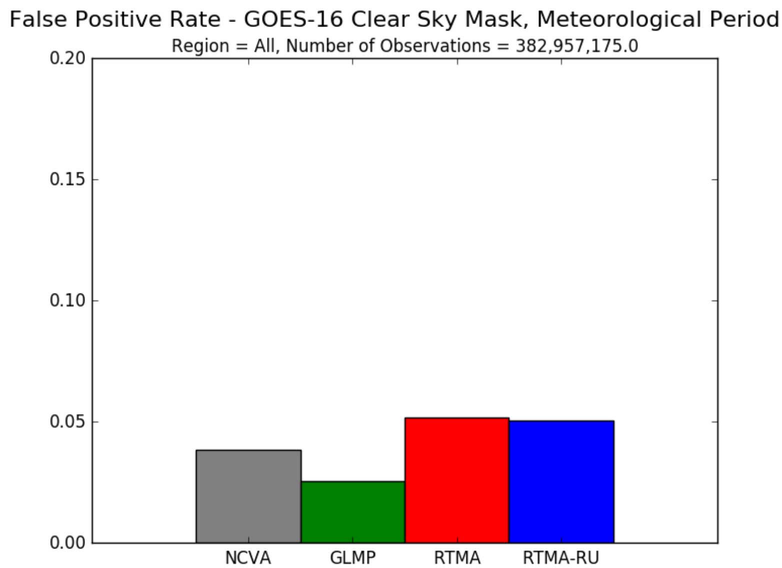
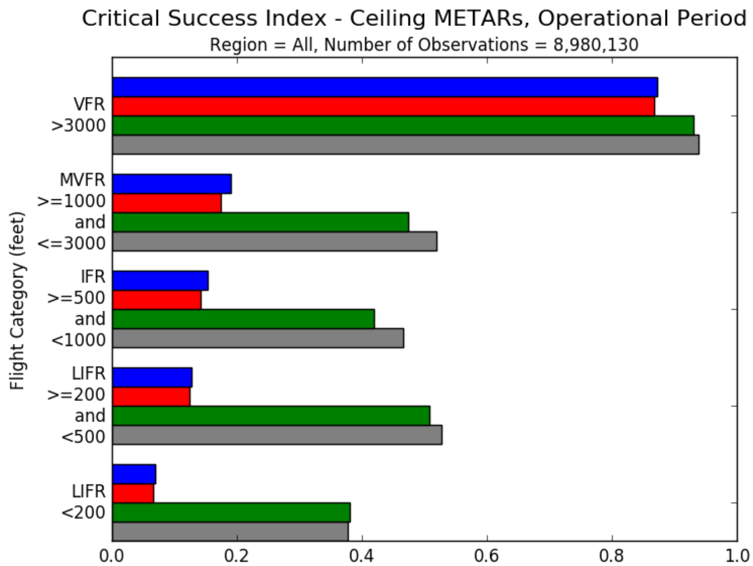


Figure 26. False positive rate for each product measured against the GOES-16 clear sky mask. The false positive rate is the frequency that each product analyzed a ceiling of 3,000’ or lower when the clear sky mask identified no clouds to be present.

5.5 Operational Verification Period

5.5.1 Skill Scores Measured Against METAR

The results against the METAR ceiling data from the operational verification period are shown in Figure 27. Overall, the scores of all the products were slightly lower during the operational verification period than during the meteorological period. The relative scores were similar to what was observed during the meteorological verification period (Figure 22). Similar to the meteorological period, NCVA and GLMP scored significantly higher than the RTMA products at all categories. Again, it is noteworthy that NCVA and GLMP were able to achieve higher CSIs while analyzing a lower frequency of upper LIFR and IFR ceilings than the RTMA products. GLMP also had the highest score of the group at the lower LIFR category while analyzing a lower frequency of ceilings at that category (Figure 14). All of the differences between NCVA, GLMP, and RTMA-RU were statistically significant at the 95% confidence level, with the exception of GLMP and NCVA at the lower LIFR category. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.8729*	**0.8731**	*0.8733*
	RTMA	*0.8670*	**0.8673**	*0.8675*
	GLMP	*0.9311*	**0.9313**	*0.9315*
	NCVA	*0.9383*	**0.9385**	*0.9387*
MVFR	RTMA-RU	*0.1891*	**0.1898**	*0.1905*
	RTMA	*0.1738*	**0.1745**	*0.1752*
	GLMP	*0.4739*	**0.4749**	*0.4758*
	NCVA	*0.5193*	**0.5203**	*0.5213*
IFR	RTMA-RU	*0.1530*	**0.1540**	*0.1548*
	RTMA	*0.1415*	**0.1424**	*0.1433*
	GLMP	*0.4185*	**0.4201**	*0.4216*
	NCVA	*0.4650*	**0.4664**	*0.4679*
Upper LIFR	RTMA-RU	*0.1272*	**0.1282**	*0.1292*
	RTMA	*0.1240*	**0.1249**	*0.1258*
	GLMP	*0.5060*	**0.5079**	*0.5098*
	NCVA	*0.5261*	**0.5278**	*0.5295*
Lower LIFR	RTMA-RU	*0.0683*	**0.0692**	*0.0702*
	RTMA	*0.0653*	**0.0663**	*0.0672*
	GLMP	*0.3770*	**0.3806**	*0.3843*
	NCVA	*0.3743*	**0.3775**	*0.3812*

Figure 27. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against METAR ceiling observations during the operational verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

5.5.2 Skill Scores Measured Against ASOS

Figure 28 shows the CSIs measured against ASOS ceiling observations during the operational verification period. In general, the scores were similar to the meteorological verification period (Figure 24), except that the NCVA scores decreased more than the GLMP scores. This enables GLMP to have the highest scores at the IFR, upper LIFR, and lower LIFR categories. RTMA-RU had lower scores than NCVA and GLMP at every flight category. POD, FAR, and PSS results can be found in the Appendix.

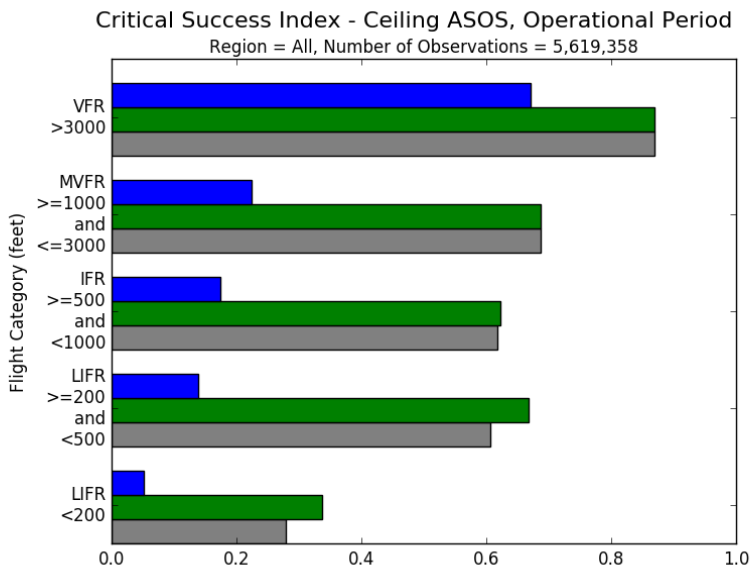
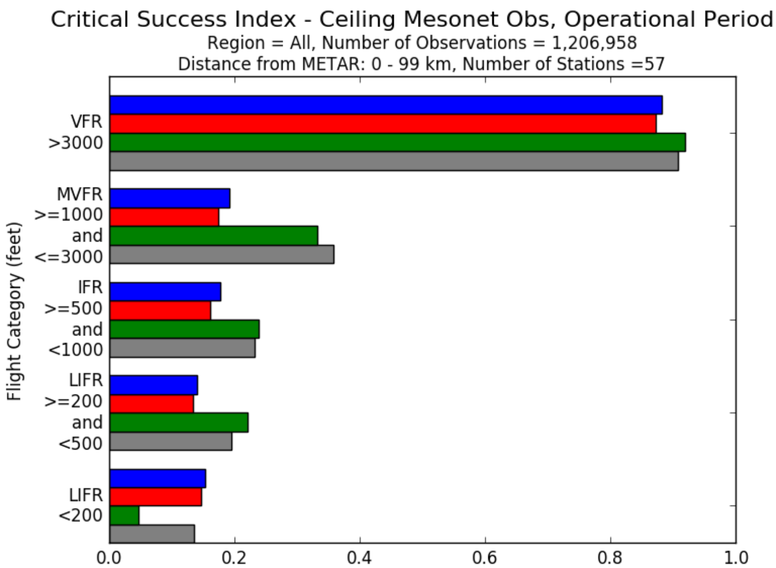


Figure 28. CSIs for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS ceiling observations during the operational verification period.

5.5.3 Skill Scores Measured Against Mesonet

The CSIs from the operational verification period measured against mesonet ceilings are presented in Figure 29. The results are similar to the meteorological period (Figure 25). NCVA and GLMP had higher CSIs than the RTMA products at the VFR, MVFR, IFR, and upper LIFR flight categories. RTMA-RU and RTMA had the highest CSIs at the lower LIFR category, as was the case in the meteorological verification period. All of the differences between NCVA, GLMP, and RTMA-RU were statistically significant at the 95% confidence level. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.8818*	**0.8825**	*0.8831*
	RTMA	*0.8717*	**0.8723**	*0.8729*
	GLMP	*0.9196*	**0.9201**	*0.9206*
	NCVA	*0.9074*	**0.9079**	*0.9084*
MVFR	RTMA-RU	*0.1906*	**0.1925**	*0.1947*
	RTMA	*0.1721*	**0.1741**	*0.1761*
	GLMP	*0.3305*	**0.3330**	*0.3356*
	NCVA	*0.3565*	**0.3591**	*0.3616*
IFR	RTMA-RU	*0.1759*	**0.1783**	*0.1811*
	RTMA	*0.1587*	**0.1611**	*0.1634*
	GLMP	*0.2363*	**0.2394**	*0.2427*
	NCVA	*0.2298*	**0.2331**	*0.2363*
Upper LIFR	RTMA-RU	*0.1378*	**0.1404**	*0.1433*
	RTMA	*0.1320*	**0.1346**	*0.1373*
	GLMP	*0.2170*	**0.2209**	*0.2250*
	NCVA	*0.1921*	**0.1955**	*0.1990*
Lower LIFR	RTMA-RU	*0.1495*	**0.1528**	*0.1564*
	RTMA	*0.1437*	**0.1468**	*0.1497*
	GLMP	*0.0435*	**0.0464**	*0.0493*
	NCVA	*0.1318*	**0.1351**	*0.1388*

Figure 29. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet ceiling observations during the operational verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

5.5.4 Skill Scores Measured Against GOES-16 Clear Sky Mask

Figure 30 shows the false positive rate for each product measured against the GOES-16 Clear Sky Mask during the operational verification period. The relative performance of each product was similar to the meteorological verification period (Figure 26), except for the larger increase in false positive rate in the RTMA. The longer latency of the RTMA product caused it to incorrectly place MVFR and lower ceilings more often in regions that the Clear Sky Mask identified as having no clouds present.

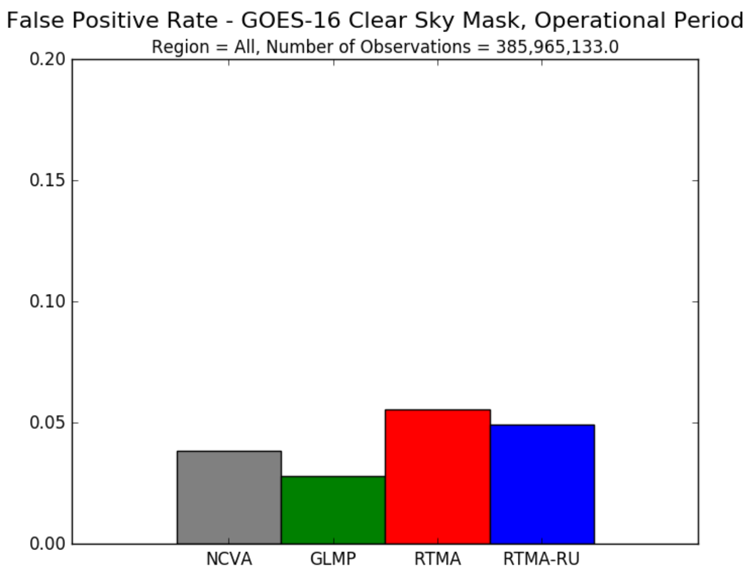


Figure 30. False positive rate for each product measured against the GOES-16 clear sky mask. The false positive rate is the frequency that each product analyzed a ceiling of 3,000' or lower when the clear sky mask identified no clouds to be present.

5.6 Regional Analysis

In addition to the CONUS analysis, a regional analysis was also performed. Figure 31 shows the CSI measured against ceiling METARs for the North Central, Northeast, West, and South regions, as well as the full CONUS statistics for reference. The CSIs in all of the regions generally followed the same pattern that was observed in the full CONUS statistics. NCVA and GLMP were higher than RTMA and RTMA-RU in every flight category. NCVA had the highest and GLMP the second highest CSI for the VFR, MVFR, and IFR categories in all of the regions. GLMP had the highest CSI at the upper and lower LIFR categories in the North Central and Northeast, while NCVA had the highest scores in the West and South regions. The most notable difference from the full CONUS statistics occurred at the lower LIFR category, where NCVA was able to obtain the highest CSI in the West and South regions, which differed from GLMP's higher performance in the full CONUS statistics. The regional graphics created against the mesonet ceiling data are included in the Appendix.

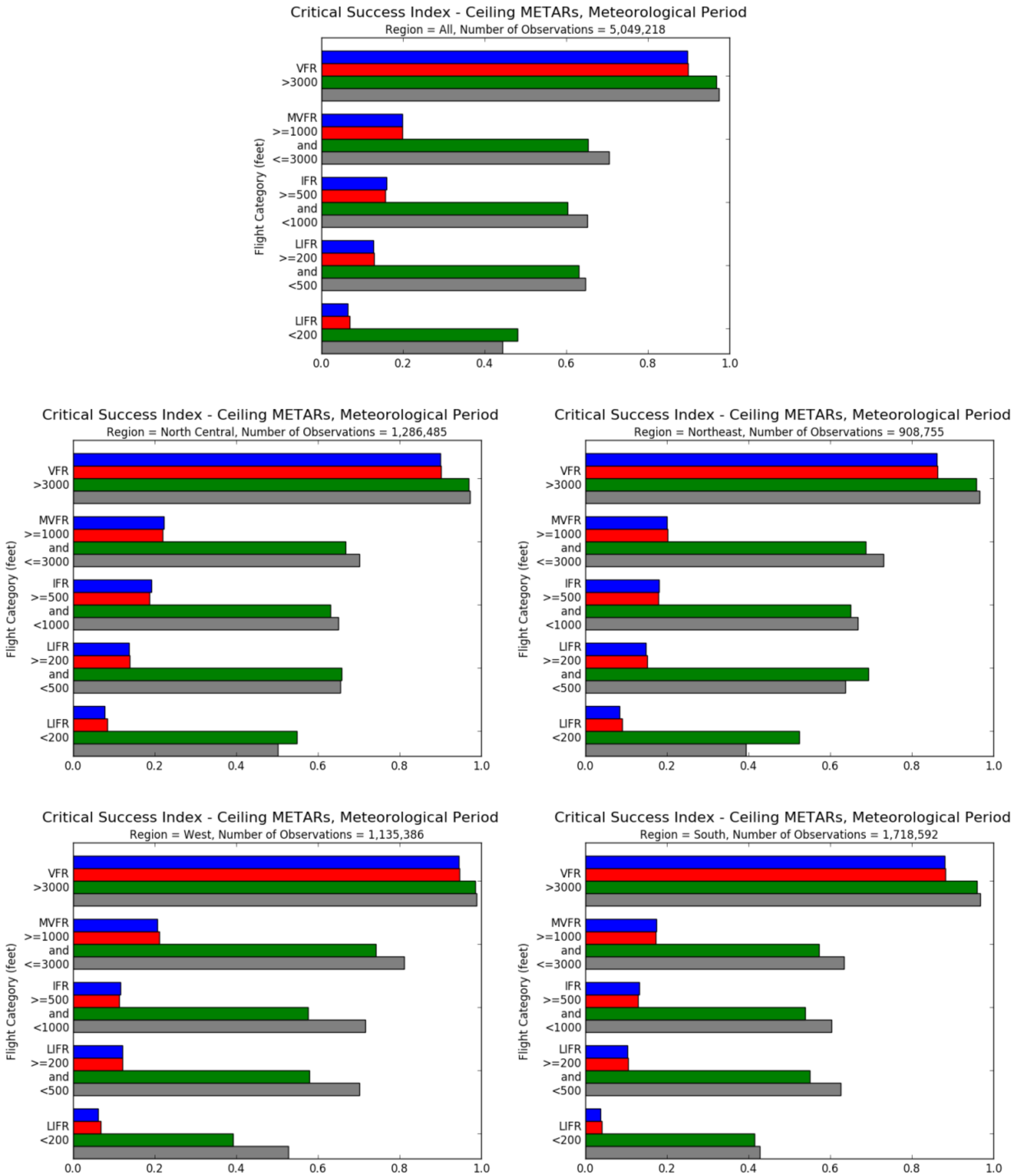


Figure 31. CSI in the North Central (center left), Northeast (center right), West (bottom left), and South (bottom right) regions of NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) during the meteorological verification period measured against ceiling METARs. The top panel showing the CSI for all regions was previously presented in Figure 22 and is presented for comparison.

6 Results – Visibility

6.1 Distributions

The four products also showed differences in their relative outputs of visibility for the different flight categories. Figure 32 shows the distribution of the visibility frequencies by flight category for each product. Similar to the ceiling results, each flight restricted visibility category during the warm season occurred at a frequency roughly half that of the cool season. RTMA and RMTA-RU had higher rates of lower LIFR visibility compared to NCVA and GLMP, but had rates in between NCVA and GLMP at the upper LIFR category. GLMP had the lowest occurrence rate at both of the LIFR categories. Both GLMP and NCVA were higher than RTMA and RTMA-RU at the IFR category. NCVA had a 1% higher frequency of MVFR visibility than any of the other products.

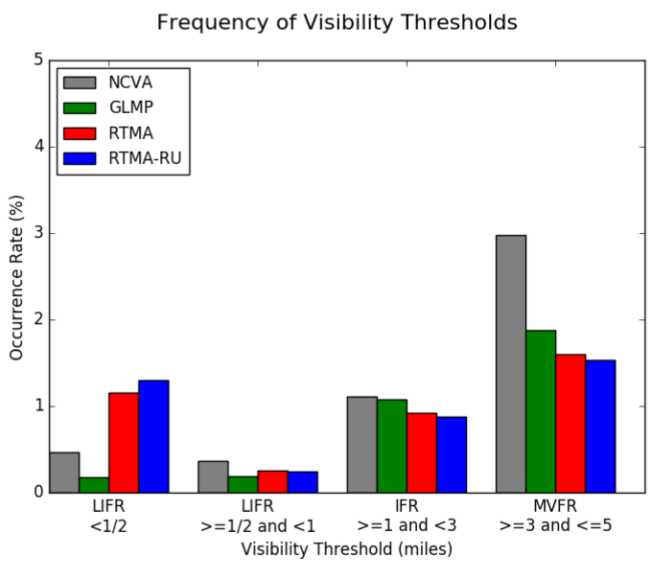


Figure 32. Distribution of visibility frequencies by aviation category. Products were equalized to the same geographic boundaries and frequencies were normalized to the number of grid points for each product.

The relative distributions changed at the lower LIFR and IFR categories when compared only at the METAR locations. At this subset of locations, shown in Figure 33, GLMP had higher occurrence rates of lower LIFR and IFR visibilities than NCVA. NCVA was the closest match to the METAR frequency in each flight category. GLMP was the second-best match at the lower LIFR, IFR, and MVFR categories, while RTMA was the second-best match at the upper LIFR category. Similar to the ceiling results, RTMA and RTMA-RU frequencies were too high in the lower LIFR category and too low in the MVFR category when compared against the METARs. GLMP had similar occurrence rates to NCVA at all categories except the MVFR category, which contrasts with the ceiling results where GLMP was similar to NCVA at all categories including the MVFR category.

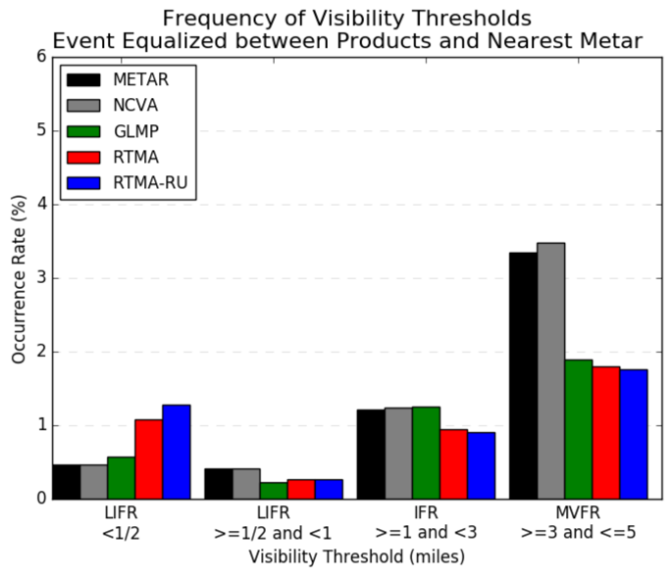


Figure 33. Distribution of visibilities by aviation category for the four products and METARs. The distribution was taken from the product point nearest to each METAR site and event equalized.

6.2 Geographic Distributions

The most significant differences in the relative distributions of the products were seen at the lower LIFR and MVFR flight categories and are presented in the following figures. Geographic distribution figures for the upper LIFR and IFR categories can be found in the Appendix. The geographic distributions at the lower LIFR flight category are displayed in Figure 34. GLMP had the lowest occurrence rates of this category across most of the CONUS, which agrees with the distribution shown in Figure 32. RTMA-RU had higher occurrence rates than the other products in most areas, most notably along the eastern seaboard and over the Great Lakes.

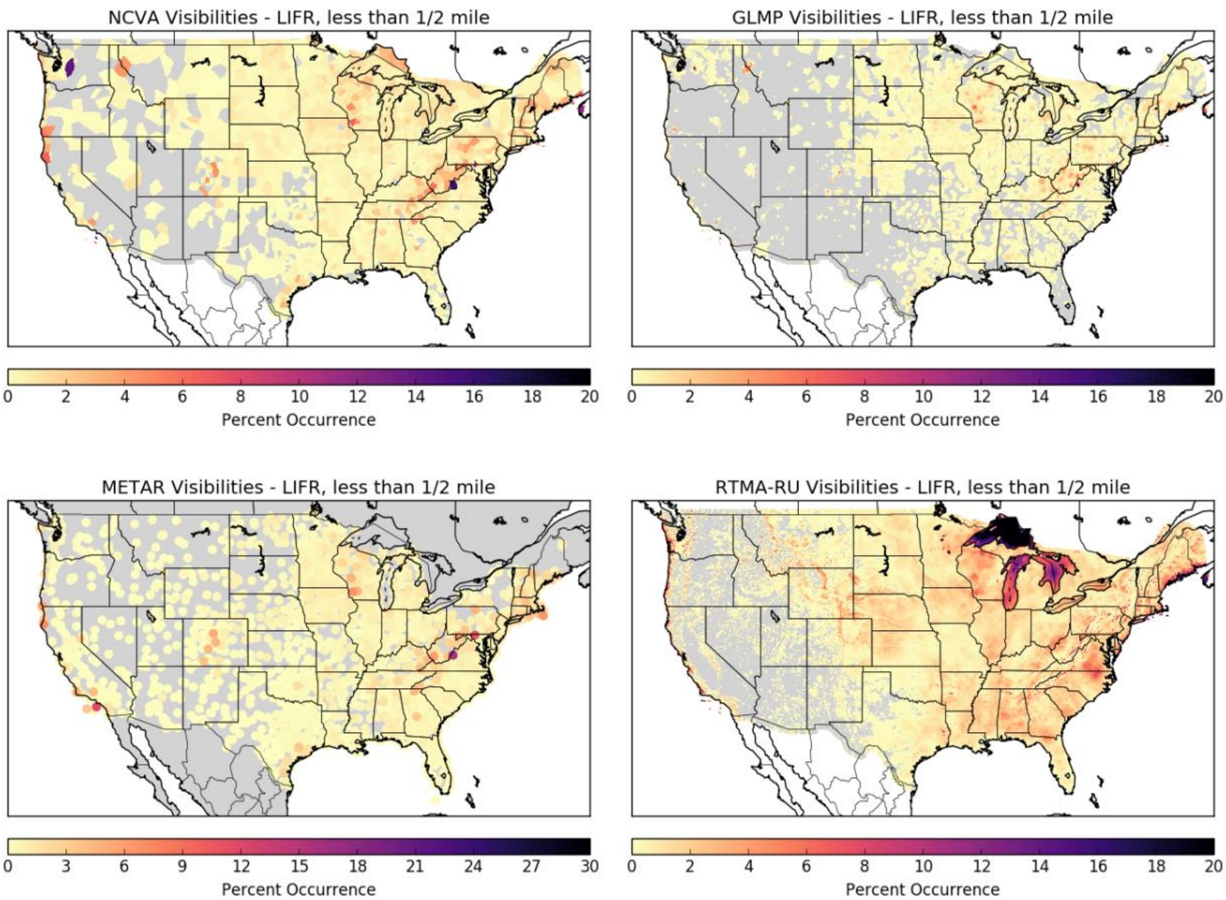


Figure 34. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the lower LIFR visibility flight category.

The difference maps for the lower LIFR category are shown in Figure 35. NCVA was the closest match at most METAR locations, followed by the GLMP. The GLMP had occurrence rates that were generally too high when compared to METARs in the eastern half of the US, with some exceptions through the Appalachians. RTMA-RU also had occurrence rates that were too high in most places east of the Rockies, except a few locations in the Appalachians. RTMA-RU better matched the METARs throughout the mountain west than in other regions at this category.

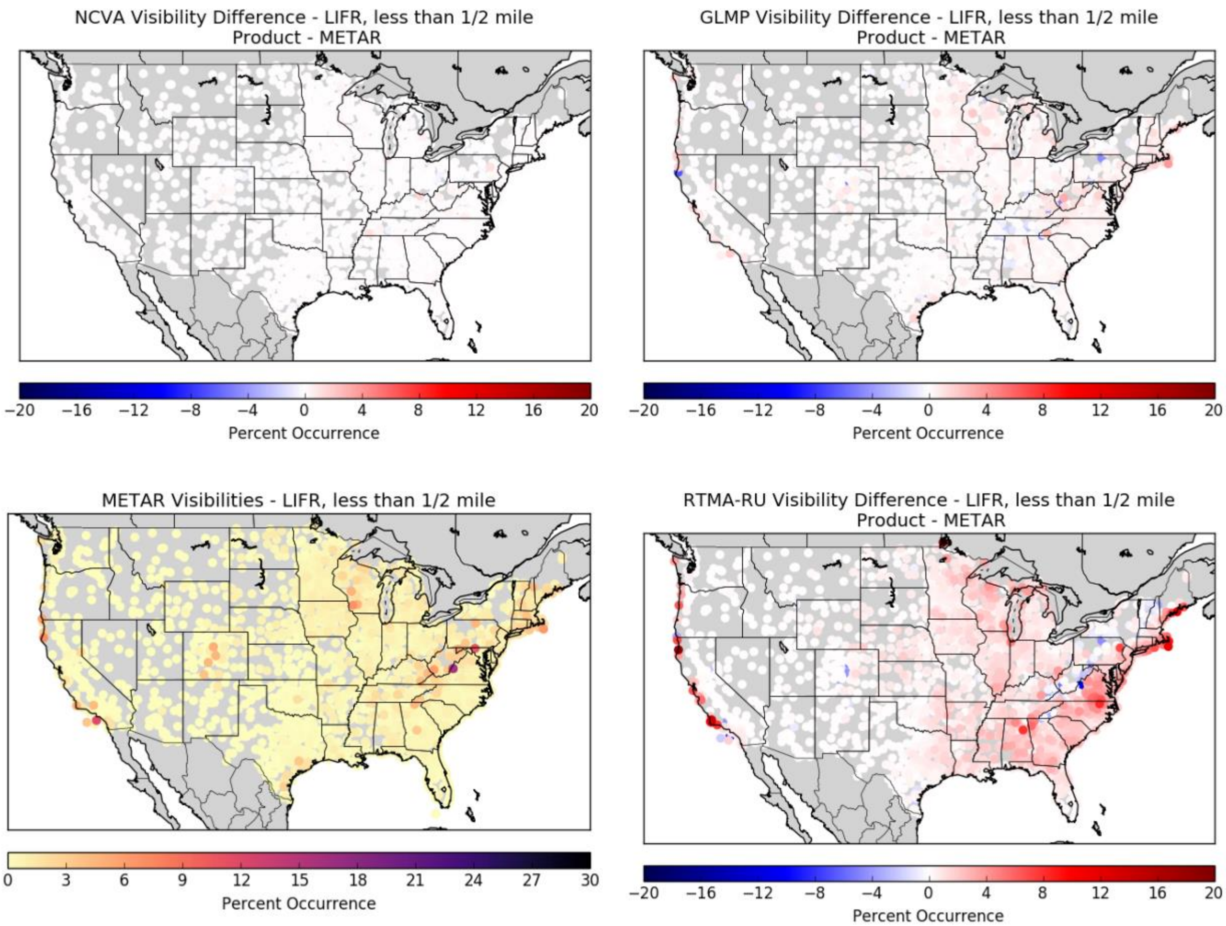


Figure 35. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the lower LIFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

NCVA had a higher frequency of visibilities at the MVFR category than the other products, as was shown in Figure 32. This is also demonstrated by the geographic distribution for each product at this category as presented in Figure 36. While GLMP and RTMA-RU had similar geographic frequencies overall, NCVA had higher frequencies than both products in most locations, particularly over the eastern half of the US. Additionally, NCVA, and to a lesser extent GLMP, captured the isolated METAR locations that reported higher frequencies of restricted visibilities, in the Carolinas and Arkansas. In each case, NCVA placed a larger geographic coverage around the METAR location than GLMP. This is in contrast to the ceiling fields where the application of terrain information created higher spatial resolution.

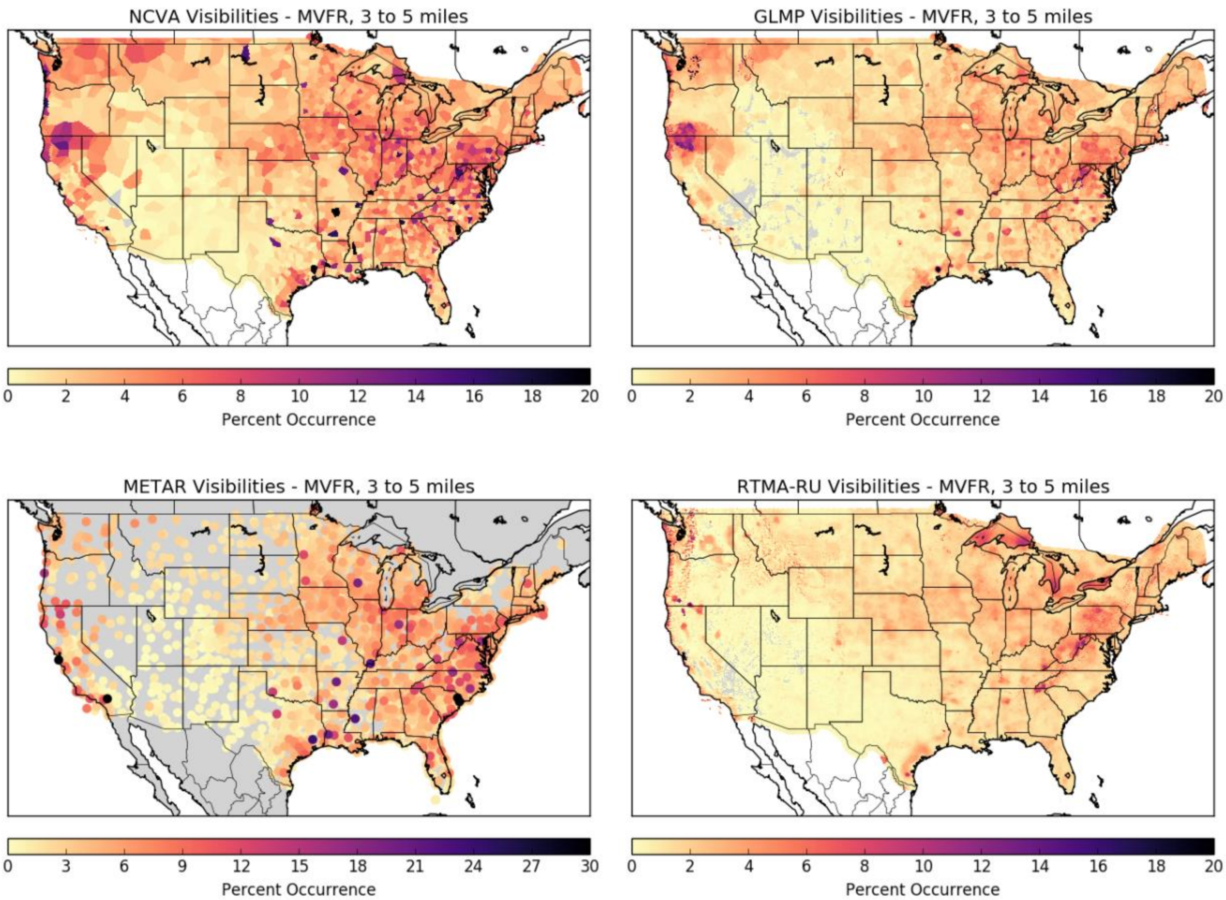


Figure 36. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the MVFR visibility flight category.

Figure 37 shows the difference maps for each product at the MVFR category. NCVA had frequencies closest to, albeit slightly higher than, those of the METARs, compared with the other products. GLMP and RTMA-RU both generally had frequencies that were less than the METAR frequencies. RTMA-RU did have some regions that were higher than the METAR occurrence rates, mainly in the mountain west, southern California, and at isolated stations in the Appalachians.

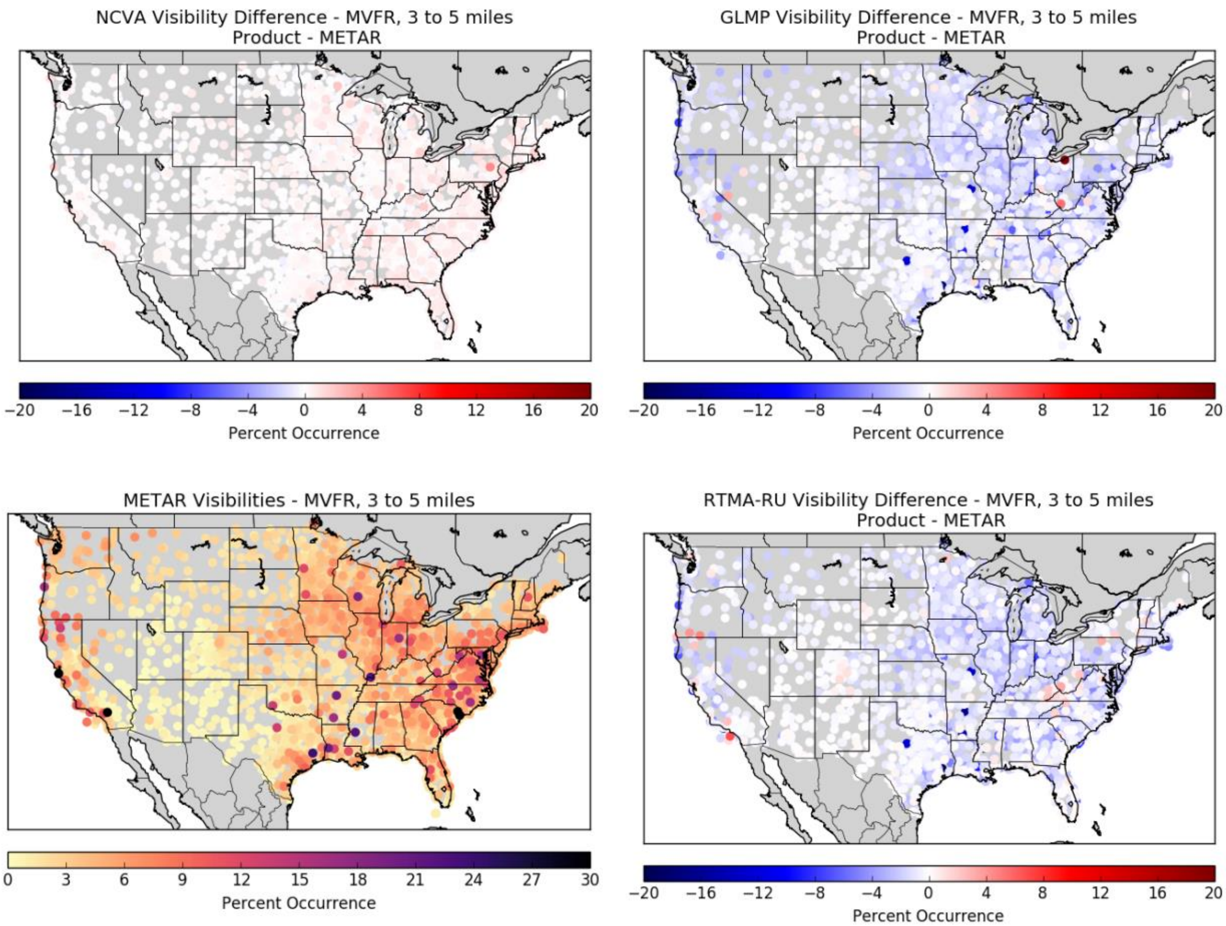


Figure 37. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the MVFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

6.3 Conditional Distributions

The conditional distributions for each product contingent on the METAR observations are shown in Figure 38. A sample perfect product is also shown for reference, where the sample product perfectly matched the METAR observations along the diagonal. NCVA was the best match to the METARs; it was in the correct category more often than GLMP and RTMA-RU. Additionally, when NCVA was incorrect, it was often only one category off from the observation. GLMP tended to analyze the visibilities greater than what was recorded in the observations, seen in the higher percentage of outputs along the horizontal VFR product category. RTMA-RU also tended to analyze the visibility greater (i.e., less restrictive conditions) than the METAR observations. In particular, RMTA-RU struggled at the upper LIFR category and in fact, placed analyzed visibilities more frequently in any other category than the correct upper LIFR category according to the observations.

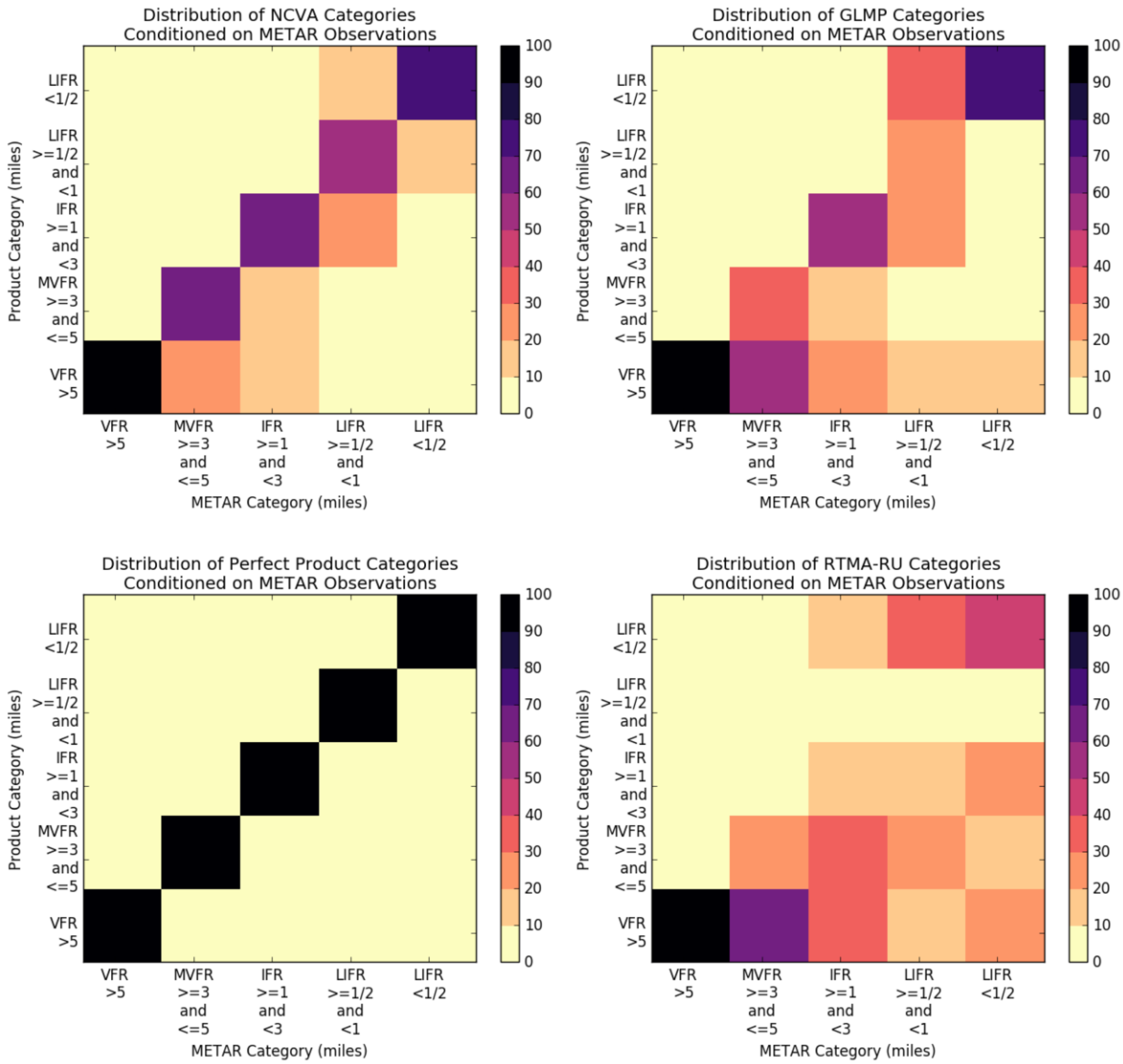


Figure 38. Conditional distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right), conditioned on the flight category of the METAR observation. An example of a perfect conditional distribution (lower left) is shown for reference.

6.4 Meteorological Period

6.4.1 Skill Scores Measured Against METAR

Figure 39 shows the POD, FAR, PSS, and CSI of each analysis product measured against METARs during the meteorological verification period. NCVA and GLMP had higher PODs, PSSs, and CSIs than RTMA and RTMA-RU at every flight category. The PODs for all of the algorithms were better in the lower LIFR category than in the upper LIFR, IFR, and MVFR categories, which indicates that it was easier to analyze severely restricted visibility than to exactly match a less degraded visibility. As with ceilings, RTMA/RTMA-RU captured fewer lower LIFR events despite identifying that category more often than NCVA and GLMP. GLMP, on the other hand, identified relatively fewer upper LIFR and subsequently captured fewer events in that category. NCVA had the highest and GLMP had the second highest CSIs in every flight category. In addition, the FARs for NCVA and GLMP were lower than RTMA and RTMA-RU in every flight category. The PSSs were very similar to the PODs because of the large sample size of nearly 5,000,000 observations, which contained a large group of correct no's for restricted visibility in the score calculation for the restricted flight categories. Similar to the ceiling METARs, CSI was selected as the most representative visibility statistic because it does not include this large group of correct no's and will be the primary statistic discussed for the remaining data sets. The differences between NCVA, GLMP, and RTMA-RU were statistically significant at every flight category at the 95% confidence level (Figure 40). POD, FAR, and PSS plots can be found in the Appendix.

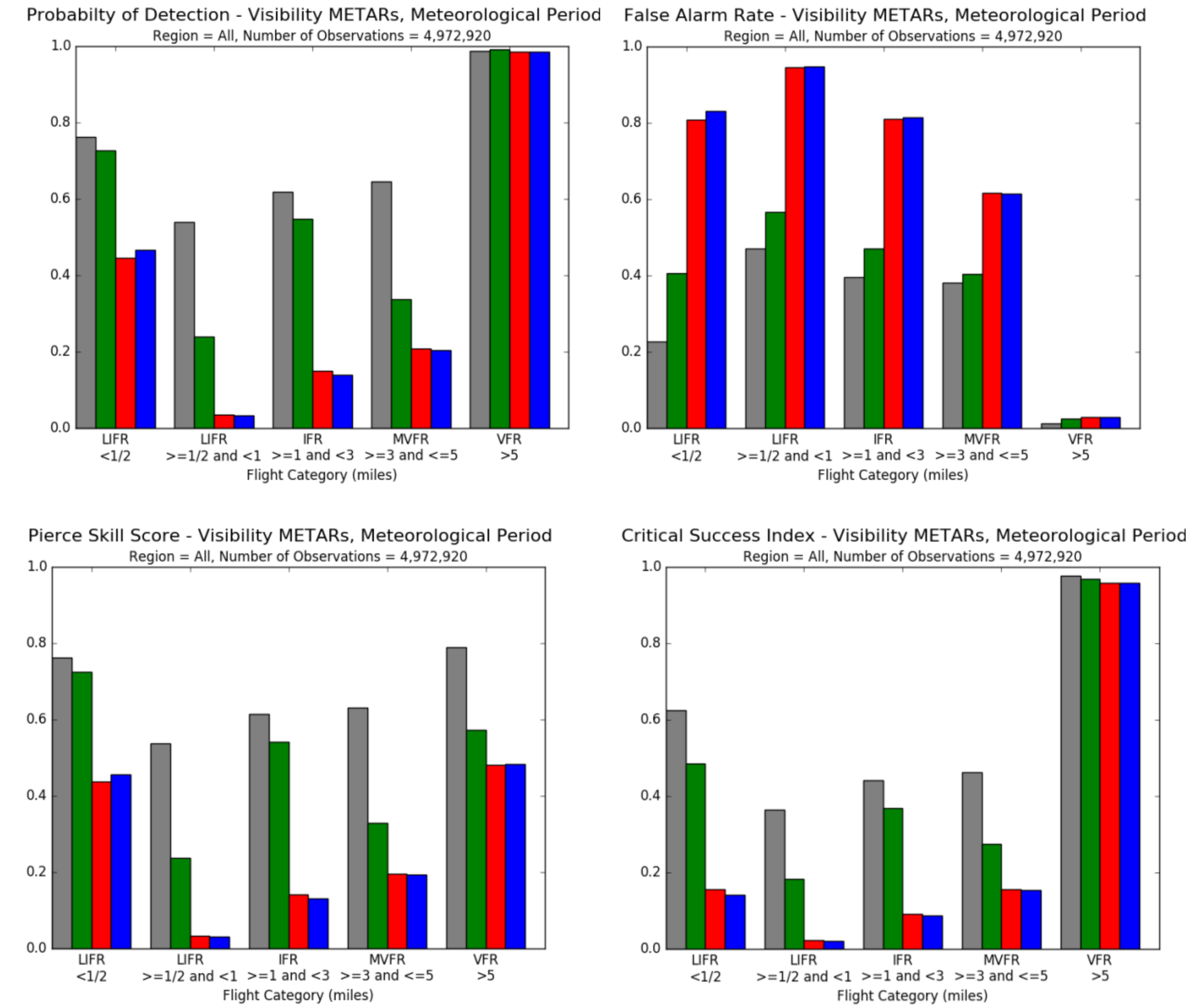


Figure 39. The POD (upper left), FAR (upper right), PSS (lower left), and CSI (lower right) of the visibility products measured against METARs during the meteorological verification period.

		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.9565*	**0.9567**	*0.9569*
	RTMA	*0.9575*	**0.9577**	*0.9578*
	GLMP	*0.9683*	**0.9685**	*0.9686*
	NCVA	*0.9756*	**0.9757**	*0.9759*
MVFR	RTMA-RU	*0.1523*	**0.1539**	*0.1554*
	RTMA	*0.1538*	**0.1554**	*0.1570*
	GLMP	*0.2718*	**0.2739**	*0.2758*
	NCVA	*0.4602*	**0.4622**	*0.4643*
IFR	RTMA-RU	*0.0841*	**0.0860**	*0.0878*
	RTMA	*0.0893*	**0.0911**	*0.0930*
	GLMP	*0.3650*	**0.3681**	*0.3714*
	NCVA	*0.4371*	**0.4404**	*0.4437*
Upper LIFR	RTMA-RU	*0.0191*	**0.0207**	*0.0223*
	RTMA	*0.0204*	**0.0219**	*0.0234*
	GLMP	*0.1772*	**0.1820**	*0.1867*
	NCVA	*0.3588*	**0.3645**	*0.3693*
Lower LIFR	RTMA-RU	*0.1393*	**0.1418**	*0.1443*
	RTMA	*0.1528*	**0.1555**	*0.1582*
	GLMP	*0.4797*	**0.4850**	*0.4903*
	NCVA	*0.6178*	**0.6235**	*0.6287*

Figure 40. The lower and upper bounds for the 95% confidence interval of the CSI for visibility products measured against METARs during the meteorological verification period.

6.4.2 Skill Scores Measured Against ASOS

Similar to the ceiling ASOS data, the visibility ASOS data came from a subset of the METAR locations (789 of the 2,026 sites) that reported at a five-minute frequency. The XX:00, XX:15, XX:30, and XX:45 ASOS reports were matched to the NCVA, GLMP, and RTMA-RU products with the same valid time. The hourly updating RTMA product was not used in this calculation in order to take advantage of the sub-hourly reports that make the ASOS data valuable. Figure 41 shows the CSIs measured against the ASOS data for NCVA, GLMP, and RTMA-RU. The results are similar to those measured against the METARs, with NCVA scoring highest and GLMP scoring second highest at all flight categories. Again, GLMP had its poorest performance relative to NCVA at the upper LIFR category. POD, FAR, and PSS results can be found in the Appendix.

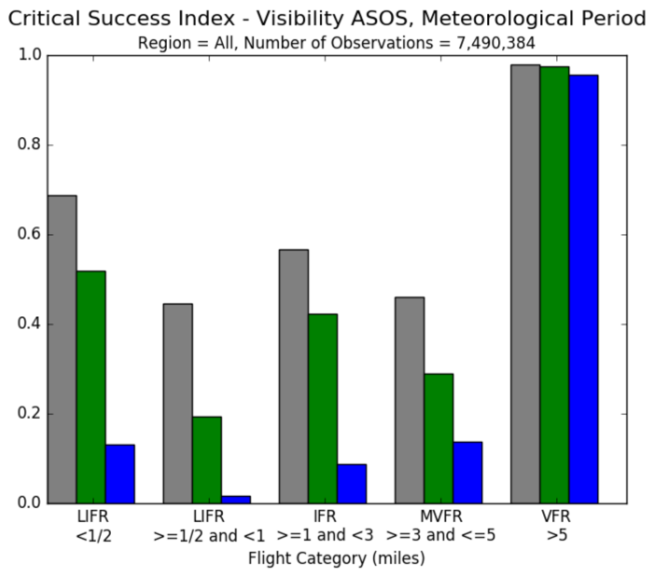
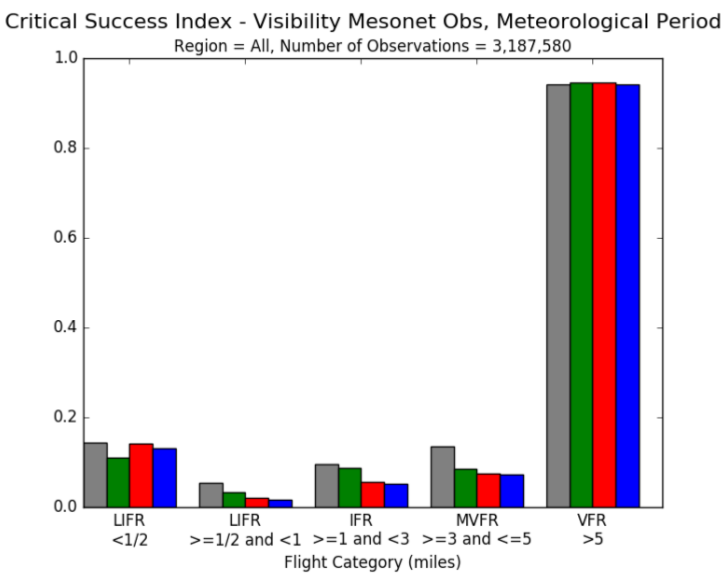


Figure 41. CSIs for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS visibility observations during the meteorological verification period.

6.4.3 Skill Scores Measured Against Mesonet

Statistical score calculations were also computed for the mesonet observations, with the full set of scores available in the Appendix. Figure 42 shows the CSIs of each product measured against the independent data from the mesonet visibility locations. The mesonet data was drawn from 457 reporting locations. Minnesota and Utah had the highest number of stations reporting during the period, however, all regions of the CONUS were represented. The CSIs for all the products were much closer to each other when compared against the mesonet data versus the METAR data. NCVA had the highest CSI score at the MVFR, IFR, upper LIFR, and lower LIFR categories. GLMP had the highest score in the VFR category and the second highest score in the MVFR, IFR, and upper LIFR categories. RTMA had the second highest score in the lower LIFR and VFR categories. All of the differences were statistically significant at the 95% confidence level except between NCVA and RTMA in the lower LIFR category. Although RTMA was able to perform near NCVA at the lower LIFR category, it needed over double the occurrence rate of that flight category to do so (Figure 32). Each product had lower CSI scores measured against the mesonet observations versus the METAR observations, but the NCVA and GLMP saw their scores reduced to a greater extent than the RTMA products. The differences seen in the CSI scores among the products were primarily the result of differences in POD. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.9416*	**0.9419**	*0.9421*
	RTMA	*0.9444*	**0.9446**	*0.9449*
	GLMP	*0.9459*	**0.9462**	*0.9464*
	NCVA	*0.9406*	**0.9408**	*0.9411*
MVFR	RTMA-RU	*0.0699*	**0.0714**	*0.0727*
	RTMA	*0.0729*	**0.0743**	*0.0757*
	GLMP	*0.0832*	**0.0847**	*0.0862*
	NCVA	*0.1338*	**0.1354**	*0.1372*
IFR	RTMA-RU	*0.0496*	**0.0512**	*0.0528*
	RTMA	*0.0548*	**0.0564**	*0.0581*
	GLMP	*0.0845*	**0.0864**	*0.0884*
	NCVA	*0.0942*	**0.0962**	*0.0982*
Upper LIFR	RTMA-RU	*0.0138*	**0.0155**	*0.0172*
	RTMA	*0.0181*	**0.0200**	*0.0217*
	GLMP	*0.0310*	**0.0337**	*0.0361*
	NCVA	*0.0502*	**0.0527**	*0.0554*
Lower LIFR	RTMA-RU	*0.1278*	**0.1309**	*0.1339*
	RTMA	*0.1380*	**0.1416**	*0.1448*
	GLMP	*0.1063*	**0.1106**	*0.1146*
	NCVA	*0.1396*	**0.1440**	*0.1476*

Figure 42. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet visibility observations during the meteorological verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

Mesonet dewpoint depressions from 14,404 stations were used to calculate a false positive rate for visibilities less than one mile. The false positive rate is the frequency that each product analyzed the visibility to be less than one mile when the mesonet dewpoint depression was 3°C or greater. Figure 43 shows the false positive rates for each product. RTMA-RU had the highest false positive rate while GLMP had the lowest false positive rate. The differences in false positive rate between each product were statistically significant at the 95% confidence level. These findings reinforce the trend from the METAR, ASOS, and mesonet results that the NCVA and GLMP products had lower false alarm rates than the RTMA products at the VFR threshold.

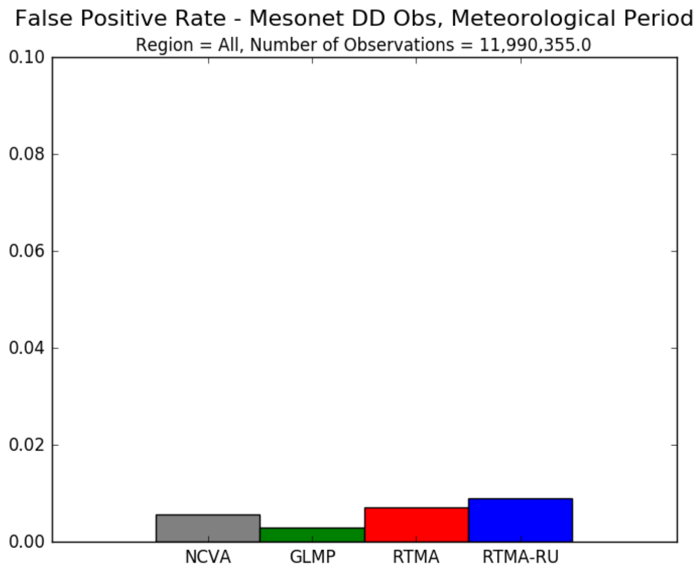
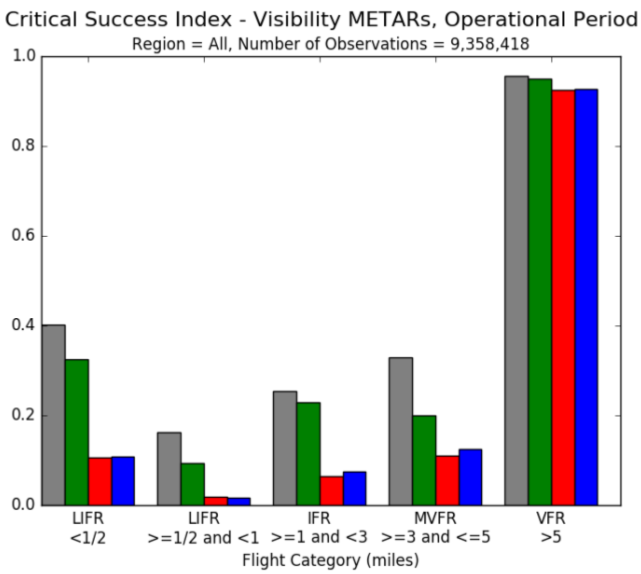


Figure 43. False positive rate for each product measured against the mesonet dewpoint depression observations. The false positive rate is the frequency that each product analyzed a visibility of less than one mile when the dewpoint depression was 3°C or greater.

6.5 Operational Period

6.5.1 Skill Scores Measured Against METAR

Figure 44 shows the CSIs measured against visibility METARs during the operational verification period. The relative differences between the products are similar to what was seen during the meteorological period (Figure 39), but the overall scores are lower. Similar to the meteorological period, NCVA had the highest and GLMP the second highest CSIs for every flight category. Again, NCVA and GLMP had higher scores than the RTMA products at the lower LIFR category while identifying visibilities at lower frequencies (Figure 33). All of the differences between NCVA, GLMP, and RTMA-RU were statistically significant at the 95% confidence level. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.9264*	**0.9266**	*0.9268*
	RTMA	*0.9249*	**0.9250**	*0.9252*
	GLMP	*0.9491*	**0.9492**	*0.9494*
	NCVA	*0.9563*	**0.9564**	*0.9565*
MVFR	RTMA-RU	*0.1242*	**0.1251**	*0.1259*
	RTMA	*0.1079*	**0.1088**	*0.1096*
	GLMP	*0.1980*	**0.1991**	*0.2001*
	NCVA	*0.3268*	**0.3280**	*0.3293*
IFR	RTMA-RU	*0.0729*	**0.0739**	*0.0748*
	RTMA	*0.0641*	**0.0649**	*0.0657*
	GLMP	*0.2278*	**0.2293**	*0.2307*
	NCVA	*0.2527*	**0.2541**	*0.2557*
Upper LIFR	RTMA-RU	*0.0149*	**0.0157**	*0.0165*
	RTMA	*0.0176*	**0.0185**	*0.0194*
	GLMP	*0.0923*	**0.0940**	*0.0958*
	NCVA	*0.1588*	**0.1609**	*0.1630*
Lower LIFR	RTMA-RU	*0.1068*	**0.1082**	*0.1096*
	RTMA	*0.1038*	**0.1052**	*0.1067*
	GLMP	*0.3211*	**0.3242**	*0.3272*
	NCVA	*0.3972*	**0.4006**	*0.4039*

Figure 44. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against METAR visibility observations during the operational verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

6.5.2 Skill Scores Measured Against ASOS

The CSIs for ASOS visibility during the operational verification period are shown in Figure 45. Overall, the scores are lower than they were during the meteorological period (Figure 41). NCVA had the highest scores for the lower LIFR, upper LIFR, IFR, and MVFR flight categories, while GLMP had the highest score in the VFR category. Although GLMP again had the second highest scores in the lower LIFR, upper LIFR, IFR, and MVFR categories, it was closer to NCVA than it was during the meteorological period. POD, FAR, and PSS results can be found in the Appendix.

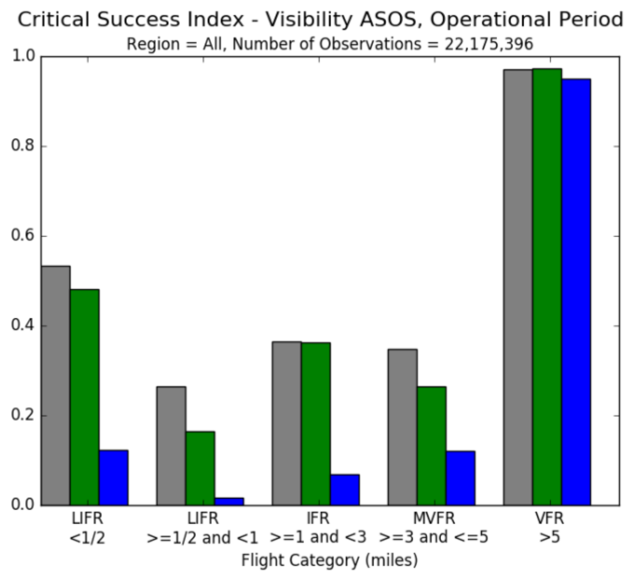
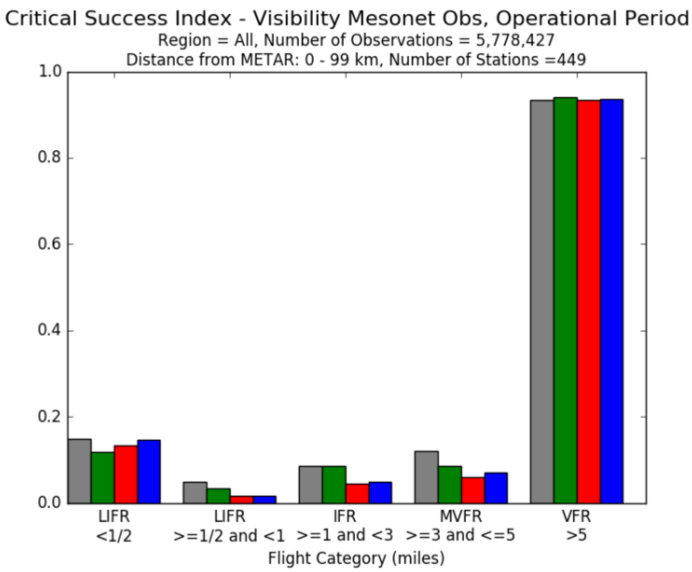


Figure 45. CSIs for NCVA (gray), GLMP (green) and RTMA-RU (blue) measured against ASOS visibility observations during the operational verification period.

6.5.3 Skill Scores Measured Against Mesonet

Figure 46 shows the CSIs during the operational verification period measured against the mesonet visibility observations. The relative differences between the products are similar to those seen during the meteorological period. The scores were closer between the products when compared against the mesonet data than the METAR data, indicating closer performance between the visibility products in areas where METARs are not available. Again, NCVA and GLMP were able to achieve similar CSIs to RTMA and RTMA-RU while having lower rates of occurrence at the lower LIFR category (Figure 32). The scores of each product were similar, but slightly lower than those during the meteorological verification period (Figure 42). The differences between NCVA, GLMP, and RTMA-RU were statistically significant at the 95% confidence level for the upper LIFR, MVFR, and VFR flight categories. POD, FAR, and PSS results can be found in the Appendix.



		Lower Bound	CSI	Upper Bound
VFR	RTMA-RU	*0.9354*	**0.9356**	*0.9358*
	RTMA	*0.9345*	**0.9347**	*0.9349*
	GLMP	*0.9409*	**0.9411**	*0.9412*
	NCVA	*0.9332*	**0.9334**	*0.9336*
MVFR	RTMA-RU	*0.0702*	**0.0712**	*0.0722*
	RTMA	*0.0597*	**0.0607**	*0.0617*
	GLMP	*0.0843*	**0.0853**	*0.0864*
	NCVA	*0.1192*	**0.1203**	*0.1214*
IFR	RTMA-RU	*0.0477*	**0.0487**	*0.0499*
	RTMA	*0.0430*	**0.0441**	*0.0451*
	GLMP	*0.0842*	**0.0856**	*0.0870*
	NCVA	*0.0850*	**0.0864**	*0.0878*
Upper LIFR	RTMA-RU	*0.0146*	**0.0158**	*0.0170*
	RTMA	*0.0161*	**0.0174**	*0.0186*
	GLMP	*0.0330*	**0.0347**	*0.0366*
	NCVA	*0.0467*	**0.0485**	*0.0503*
Lower LIFR	RTMA-RU	*0.1449*	**0.1472**	*0.1495*
	RTMA	*0.1311*	**0.1334**	*0.1355*
	GLMP	*0.1162*	**0.1190**	*0.1219*
	NCVA	*0.1457*	**0.1485**	*0.1513*

Figure 46. CSIs (left) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet visibility observations during the operational verification period. The lower and upper bounds (right) for the 95% confidence interval of the CSIs.

Figure 47 shows the false positive rates for each product measured against the mesonet dewpoint depression observations during the operational verification period. The scores are very similar to those from the meteorological verification period (Figure 43) which was caused by the large number of cases when the products correctly identified VFR visibilities. Again, the GLMP had the lowest false positive rate and the RTMA-RU the highest rate, and all differences between the products were statistically significant at the 95% confidence level.

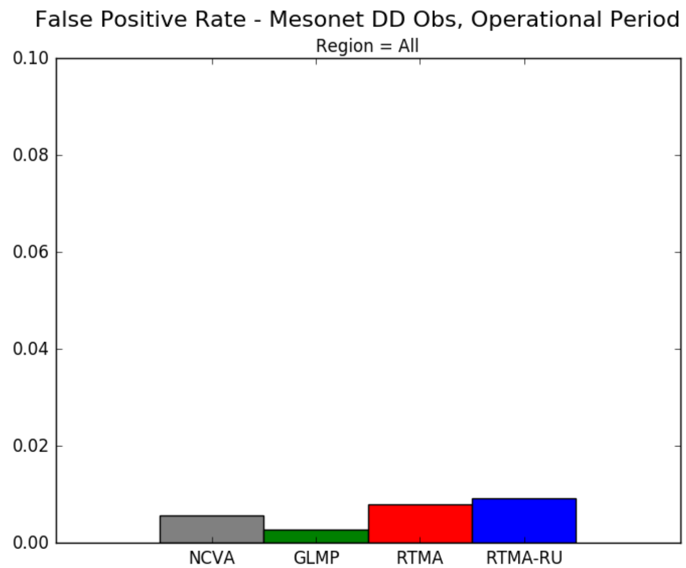


Figure 47. False positive rate for each product measured against the mesonet dewpoint depression observations during the operational verification period. The false positive rate is the frequency that each product analyzed a visibility of less than one mile when the dewpoint depression was 3°C or greater.

6.6 Regional Analysis

Figure 48 shows the CSI measured against the visibility mesonet observations for the four regions and the full CONUS statistics for reference. The largest deviations from the full CONUS statistics occurred in the West and South regions. All of the algorithms performed worse in the West region at categories less than VFR, but NCVA and GLMP experienced greater decreases in skill than RTMA and RTMA-RU at the lower LIFR flight category. NCVA stood out in the South region with the highest scores in all categories. GLMP, RTMA and RTMA-RU had comparatively worse scores in the IFR, upper LIFR, and lower LIFR categories in the South region compared to the full CONUS statistics.

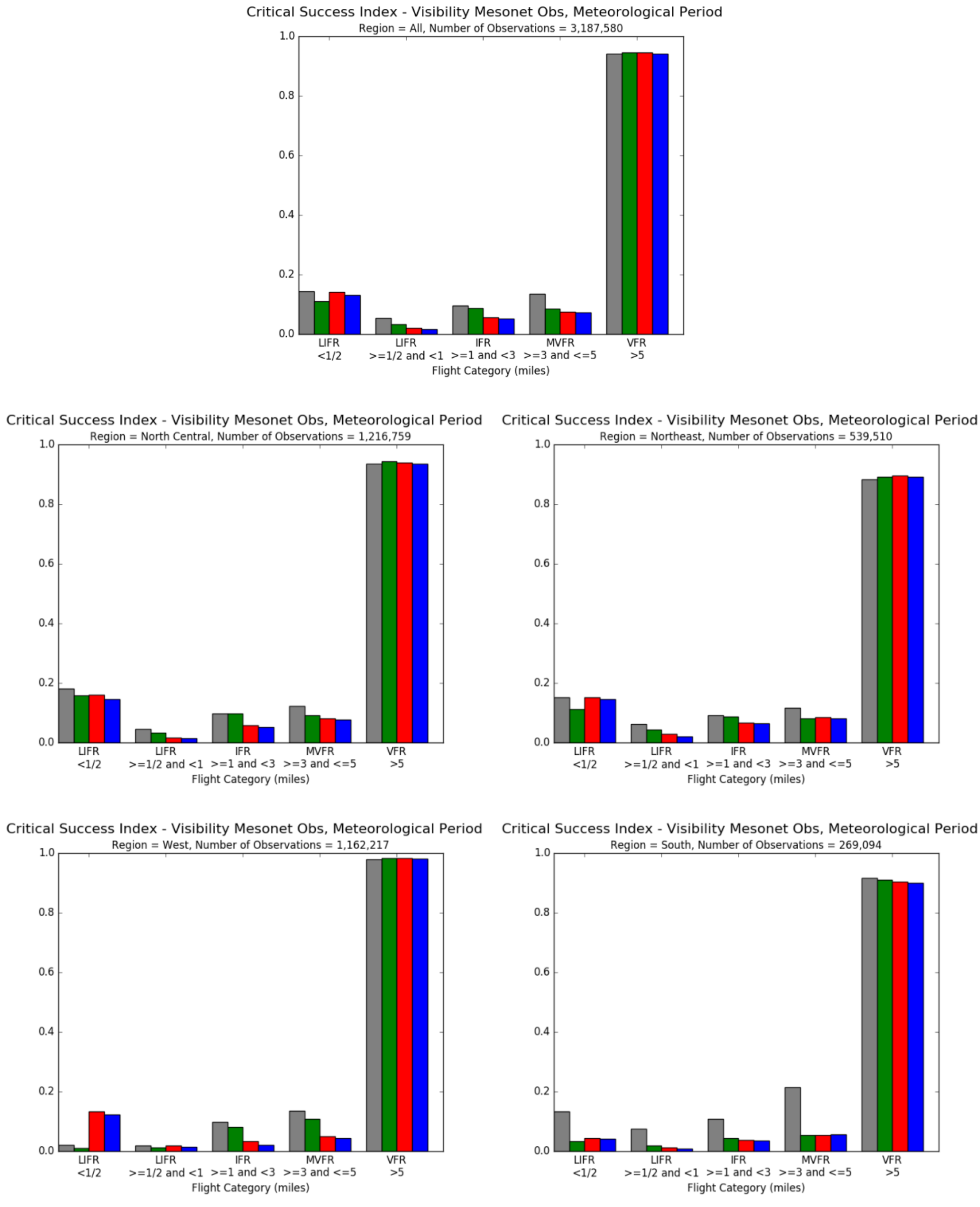


Figure 48. CSI in the North Central (center left), Northeast (center right), West (bottom left), and South (bottom right) regions of NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) during the meteorological verification period measured against mesonet visibility observations. The top panel showing the CSI for all regions was previously presented in Figure 42 and is presented for comparison.

7 Case Studies

7.1 Low Ceilings Over San Francisco on August 14, 2018

The first case study occurred when low ceilings were present over the San Francisco Bay area on August 14, 2018 that led to flight delays greater than two hours. Figure 49 shows the ceiling height output from each analysis product at 1600Z overlaid with the METAR and mesonet ceiling observations from the preceding 15-minute time period. There are significant differences between the product outputs throughout the region. First, the difference in resolution between the coarser NCVA and the finer GLMP and RTMA products was apparent and impacts the ability to resolve the small-scale terrain differences on the San Francisco Peninsula. The NCVA matched the category of all of the METAR observations, including the IFR conditions present at San Francisco Airport (SFO). NCVA does not extend over the water and slight differences between the NCVA terrain mask and the actual coastline of the bay were apparent. GLMP matched most of the flight categories observed at the METAR sites, including the IFR conditions at SFO. The RTMA products did not fare as well, recording few matches between the product output category and the observed METAR flight category. At SFO, the RTMA products both analyzed VFR ceilings when IFR ceilings were observed. The swath of VFR conditions extending through the bay did not match with any of the observations along the coastline. In addition, there were large difference between RTMA and RTMA-RU output over the Pacific Ocean and southeast of SFO.

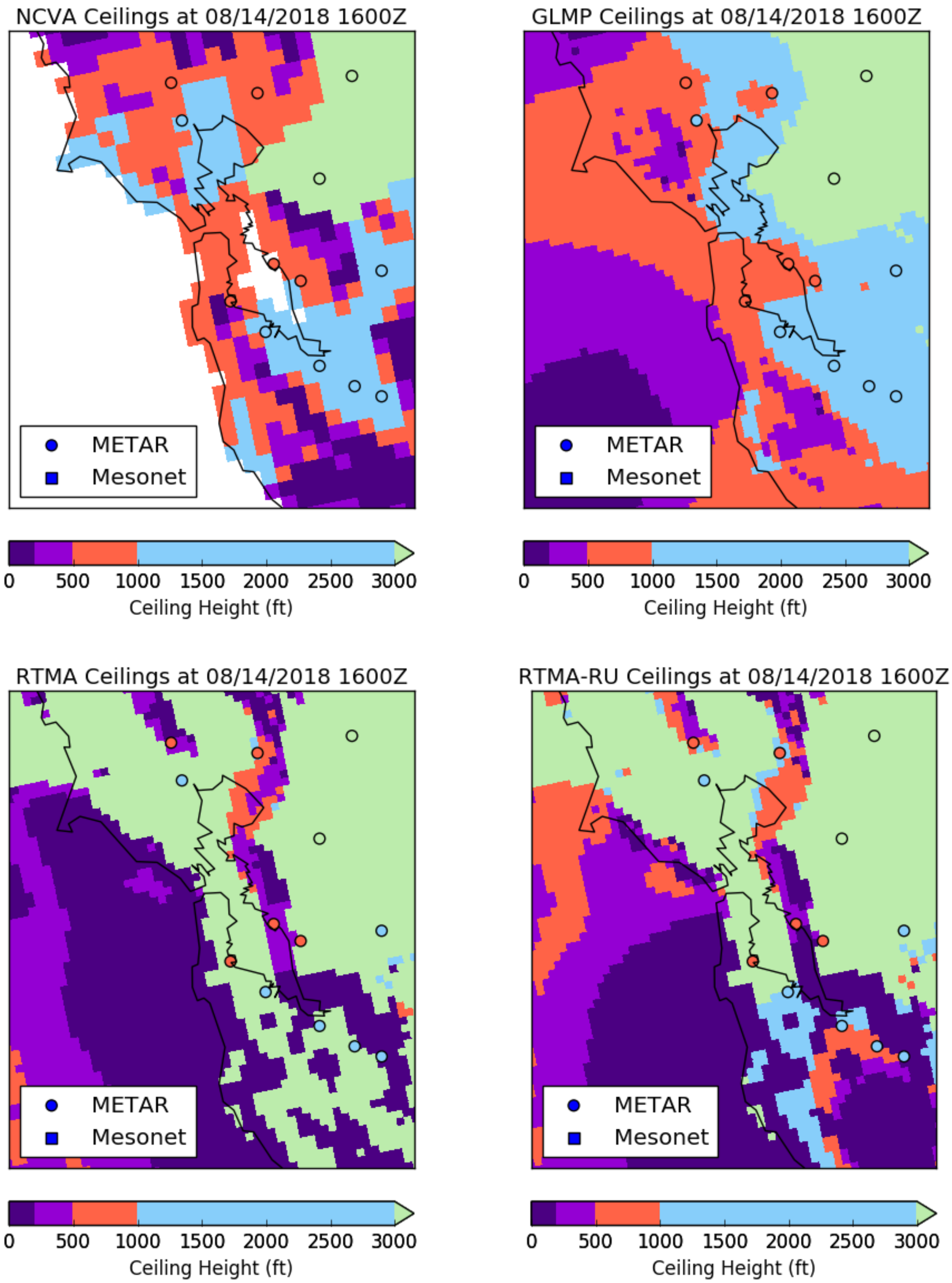


Figure 49. Ceiling output from the NCVA (upper left), GLMP (upper right), RTMA (lower left), and RTMA-RU (lower right) in the San Francisco Bay area on August 14, 2018 at 1600Z. METAR and mesonet observations from the previous 15-minute period are overlaid using the same color bar for comparison to the products. The colors represent the flight categories of VFR (green), MVFR (blue), IFR (red), upper LIFR (medium purple), lower LIFR (dark purple).

7.2 Reduced Visibility in Tennessee on May 23, 2018

Figure 50 shows the product outputs and METAR and mesonet visibility reports from a fog case in Tennessee. All four of the products identified areas of reduced visibility along the Tennessee-North Carolina border, but the placement of each flight category varied among the products. This case focuses on the reports from two mesonet stations, one east and one south of Knoxville. The mesonet station to the east reported IFR conditions, which were correctly analyzed by NCVA and GLMP, while the RTMA products had lower LIFR conditions present. The mesonet station to the south reported upper LIFR visibility. NCVA and the RTMA products had lower LIFR conditions surrounding this location and GLMP had MVFR conditions nearby. This case illustrates the difficulty in accurately analyzing visibility in regions of complex terrain.

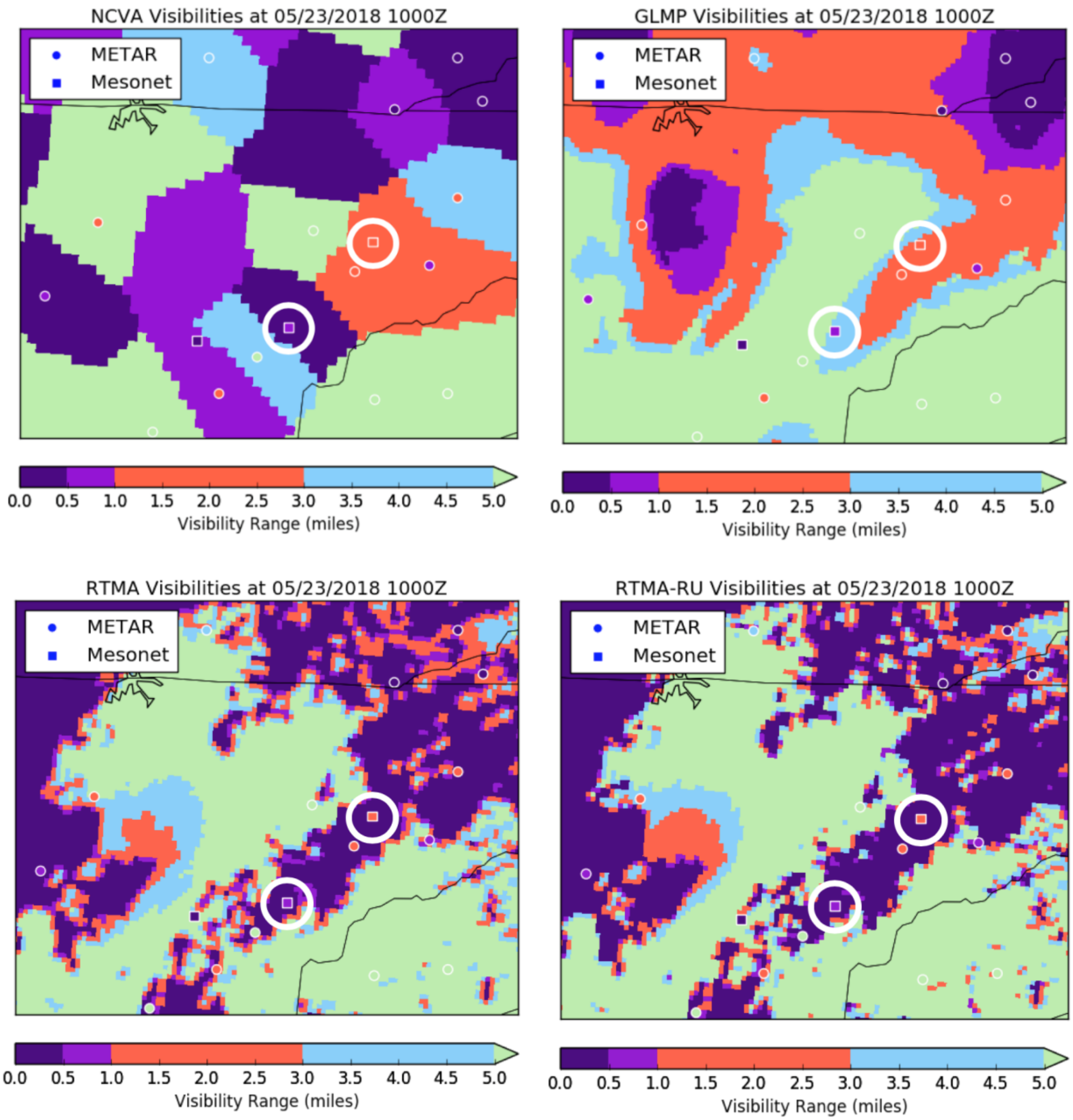


Figure 50. Visibility output from the NCVA (upper left), GLMP (upper right), RTMA (lower left), and RTMA-RU (lower right) around Salt Lake City, Utah during a snow event on May 23, 2018 at 1000Z. METAR and mesonet observations from the previous 15-minute period are overlaid using the same color bar for comparison to the products.

7.3 Reduced Visibility in Maine

The last case illustrates an example of product performance in a region of sparse METAR coverage. Figure 51 shows the visibility output for each of the four products over Maine on August 5, 2018. There are two mesonet observations highlighted in western Maine that were a considerable distance from the nearest METAR station. At 10Z, the western of the two mesonet locations recorded lower LIFR visibility and the eastern station VFR visibility. The nearest METAR and mesonet stations to the south also reported visibilities in the lower LIFR category. The time series of visibility observations from the eastern station was checked and the station had recently reported restricted visibilities, which increased faith that the VFR report was correct even with the surrounding lower visibilities. This case also illustrates the ephemeral nature of low visibilities because these two stations had reliable reports at either end of the flight category spectrum. In the region near these mesonet locations, NCVA had lower to upper LIFR visibilities, GLMP had IFR to lower LIFR visibilities, and the RTMA products had mostly VFR conditions. NCVA best matched the western mesonet station while RTMA best matched the eastern station. NCVA and GLMP were better matches to both the METAR and mesonet observations in southern Maine that reported mostly lower LIFR conditions.

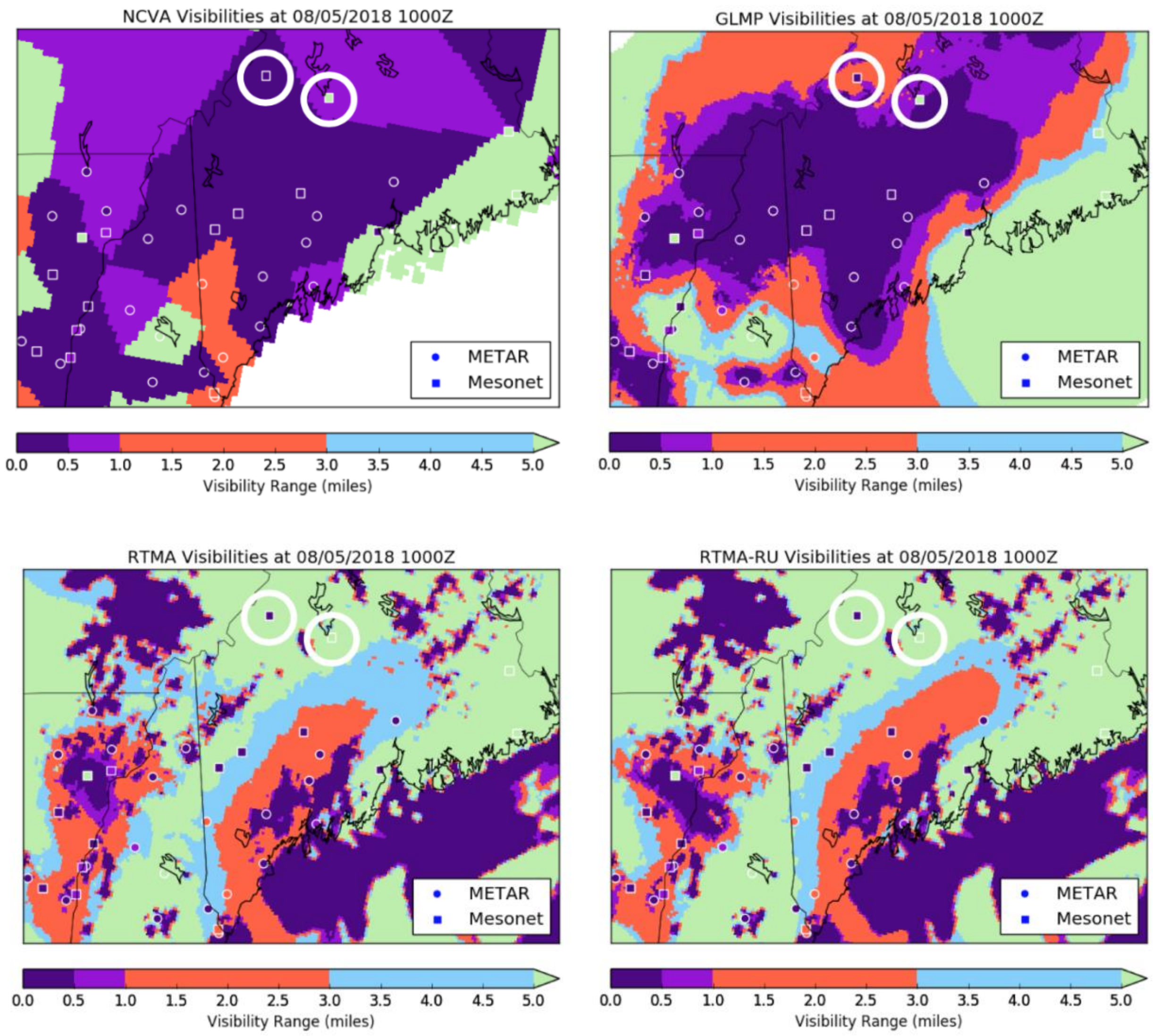


Figure 51. Visibility output from the NCVA (upper left), GLMP (upper right), RTMA (lower left), and RTMA-RU (lower right) in Maine during a fog event on August 8, 2018 at 1000Z. METAR and mesonet observations from the previous 15-minute period are overlaid using the same color bar for comparison to the products.

8 Analysis and Summary

8.1 Ceiling

Overall, the warm season results were similar to the cool season findings, with slightly lower skill scores. The ceiling analyses of NCVA and GLMP clearly separated themselves as superior to RTMA and RTMA-RU. The frequencies of NCVA and GLMP better matched the METAR frequencies at every flight category in the distributions. NCVA and GLMP scored higher than RTMA and RTMA-RU at every flight category for the CSI measured against METARs and ASOS. NCVA and GLMP also scored higher at the VFR, MVFR, IFR, and upper LIFR categories for the CSI measured against mesonet ceilings. However, RTMA and RTMA-RU did score higher than NCVA and GLMP at the lower LIFR category for the CSI measured against mesonet ceilings. It is worth noting that NCVA and GLMP attained higher PODs measured against METARs at every flight category while having lower frequencies than RTMA and RTMA-RU at the IFR, upper LIFR, and lower LIFR categories. These results held for both the meteorological and operational verification periods.

The RTMA products did not closely match the assimilated METARs and instead relied too heavily on model data in areas where observations were present.

Between NCVA and GLMP, the results were generally very close. Overall, GLMP analyzed lower occurrence rates of low ceilings at all flight categories than NCVA. The lower occurrence rates were reflected in areas of lower ceilings that generally were smaller for GLMP than for NCVA. GLMP was a close second to NCVA in CSI measured against METARs at the VFR, MVFR, and IFR categories, and bested NCVA at the upper and lower LIFR categories. For the CSI measured against mesonet ceilings, GLMP was second to NCVA at the MVFR and IFR categories, while bettering NCVA at the upper LIFR category. GLMP's poorest performance was at the lower LIFR categories measured against mesonet ceilings, which was tied to its substantially lower coverage and frequency of that flight category than the other products. Overall, GLMP strikes a good balance of limiting coverage areas while maintaining higher accuracy rates.

8.2 Visibility

NCVA and GLMP scored higher in CSI measured against METARs and ASOS than RTMA and RTMA-RU at every flight category. At each category, NCVA scored the best, at substantially higher values in the MVFR, upper LIFR, and lower LIFR categories. The results were much closer measured against mesonet visibility observations. NCVA was able to achieve the top score in the MVFR, IFR, and upper LIFR flight categories, while RTMA had the highest scores in the VFR and lower LIFR categories. GLMP was able to score close to the other products at the LIFR categories measured against mesonet observations despite having lower occurrence rates in the distributions at those categories. Similarly, the RTMA products scored close to the other products at the IFR category while having higher frequencies.

9 Appendix

9.1 Ceiling and Visibility Products Operational Timelines

The following graphics show the ceiling and visibility analysis products as they are available from the perspective of the operational decision maker.

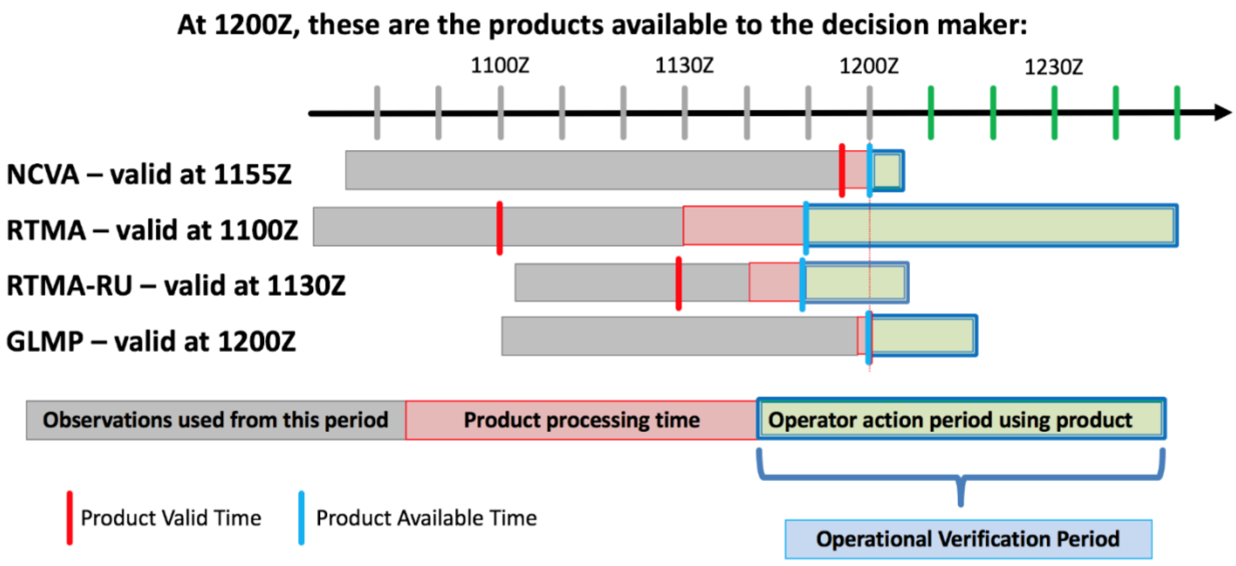


Figure 52. Operational verification period at 1200Z.

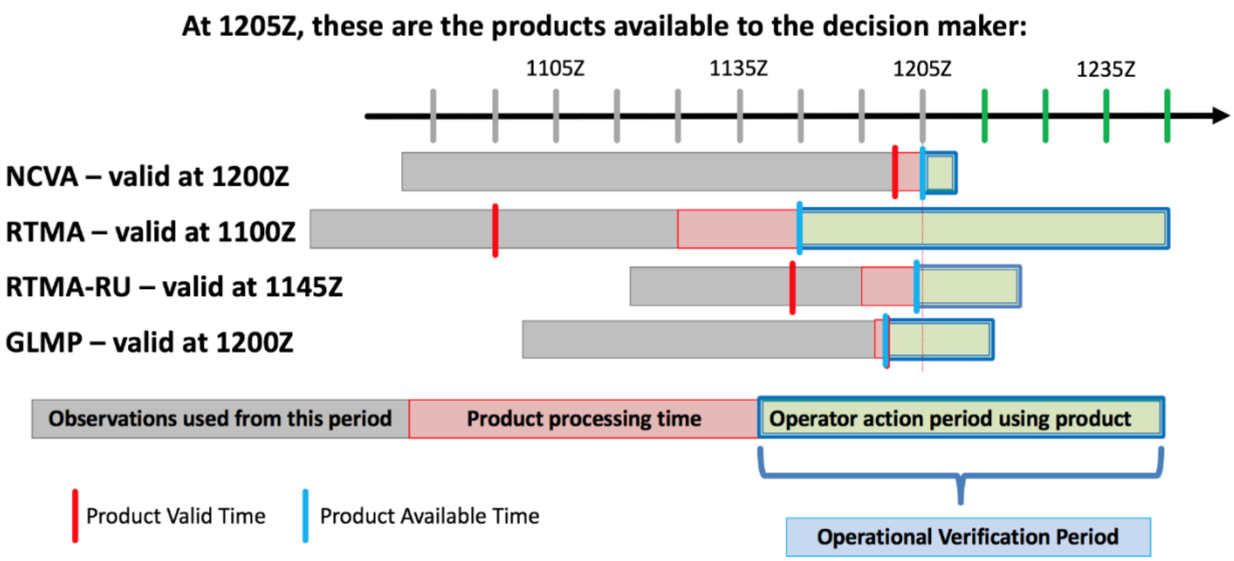


Figure 53. Operational verification period at 1205Z.

9.2 NCVA Timelines

The following figures show the timeline of the NCVA and the application of the Meteorological Verification Period and Operational Verification Period.

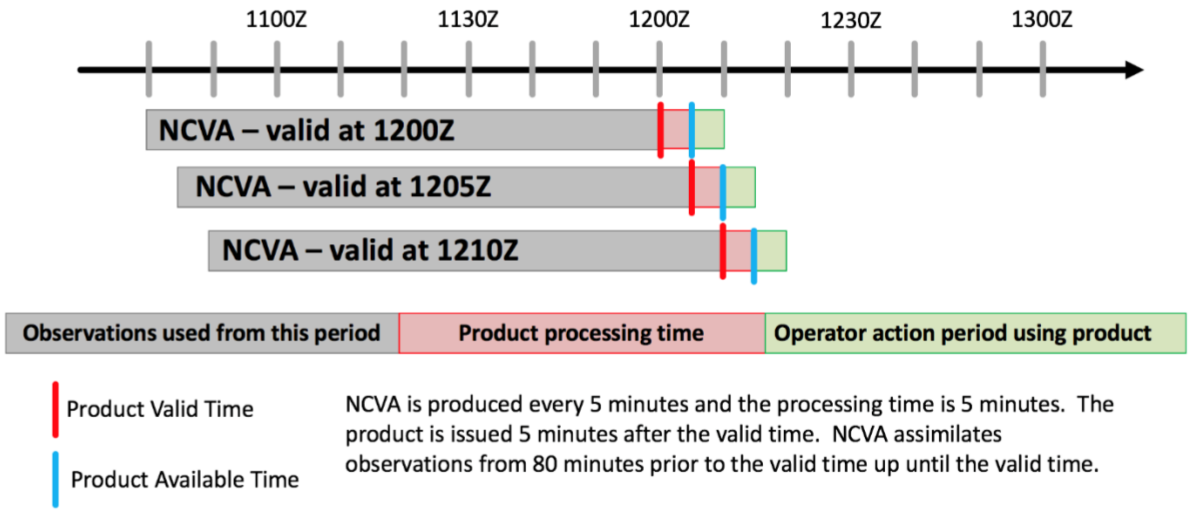


Figure 54. NCVA availability timeline.

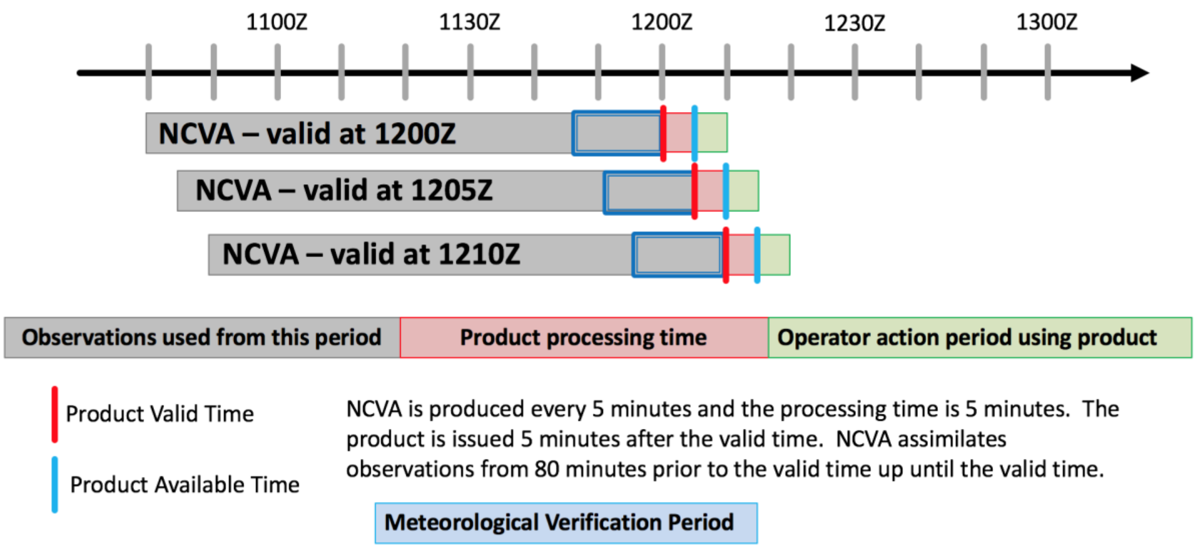


Figure 55. Meteorological verification period applied to NCVA.

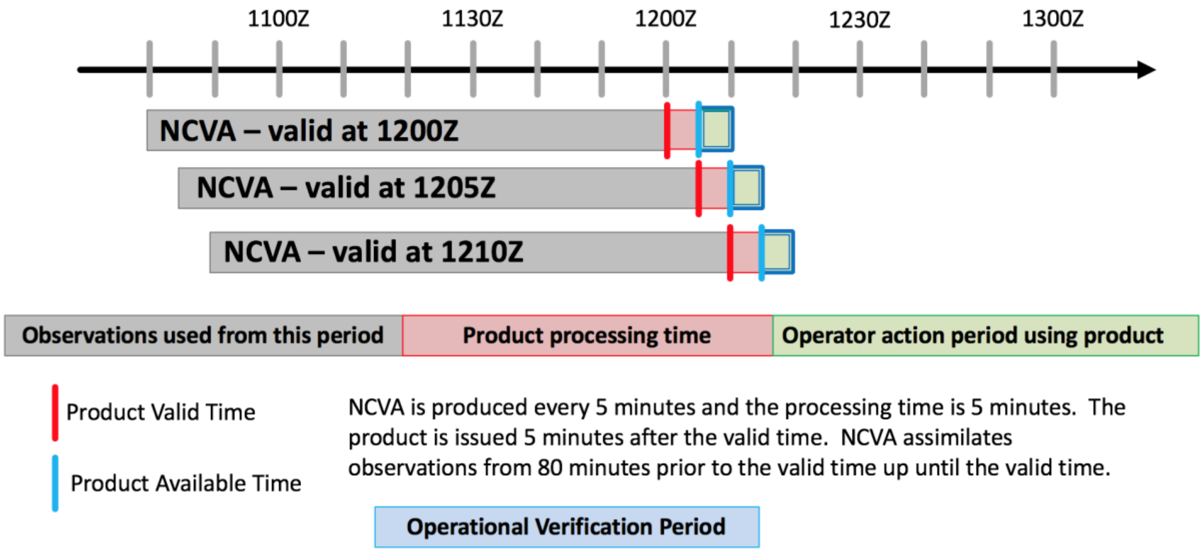


Figure 56. Operational verification period applied to NCVA.

9.3 GLMP Timelines

The following figures show the timeline of the GLMP and the application of the Meteorological Verification Period and Operational Verification Period.

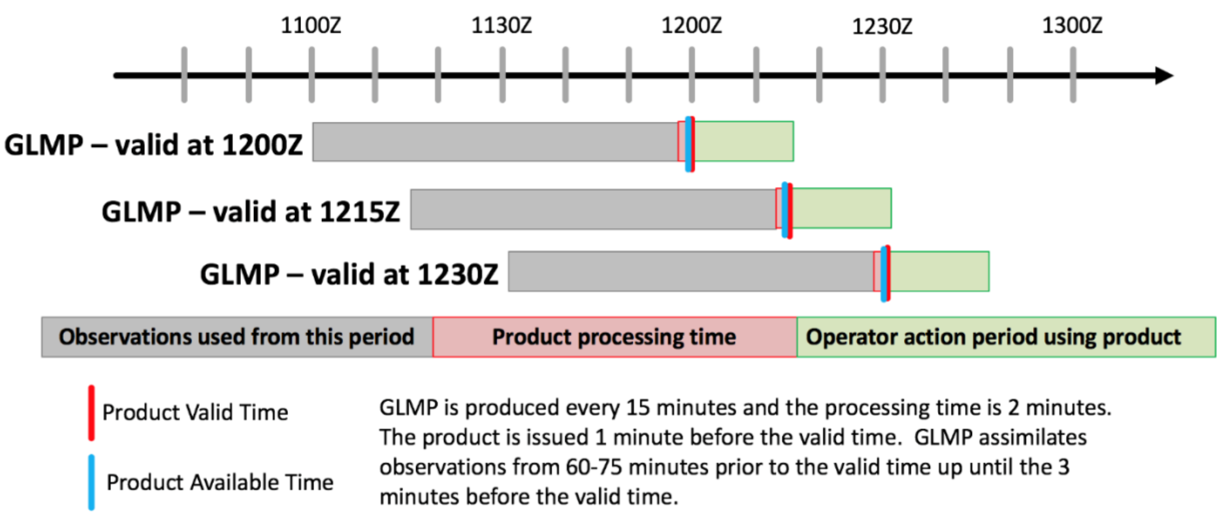


Figure 57. GLMP availability timeline.

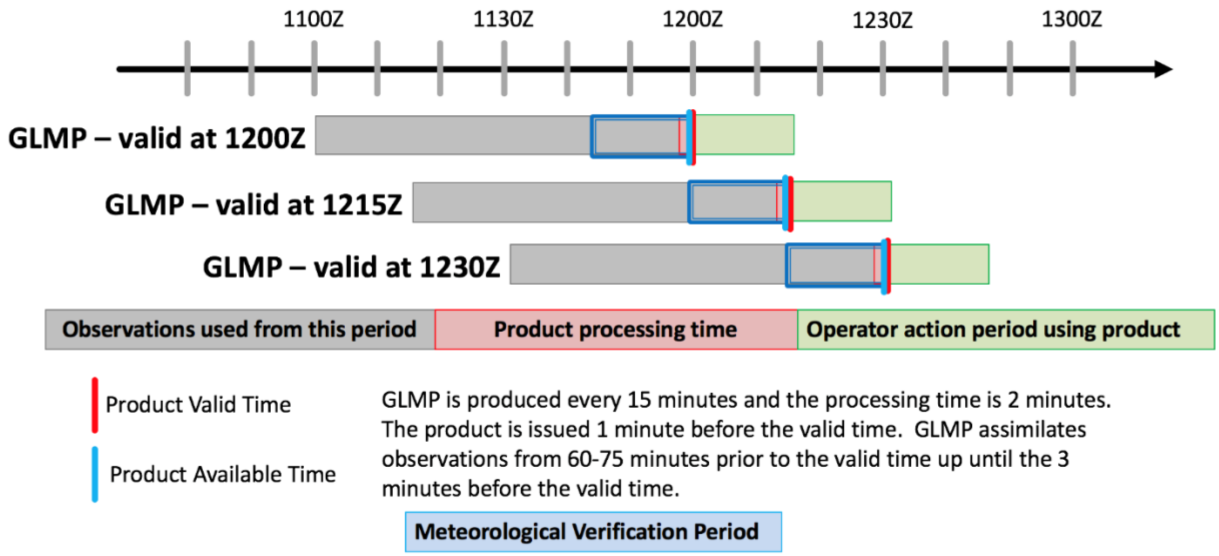


Figure 58. Meteorological verification period applied to GLMP.

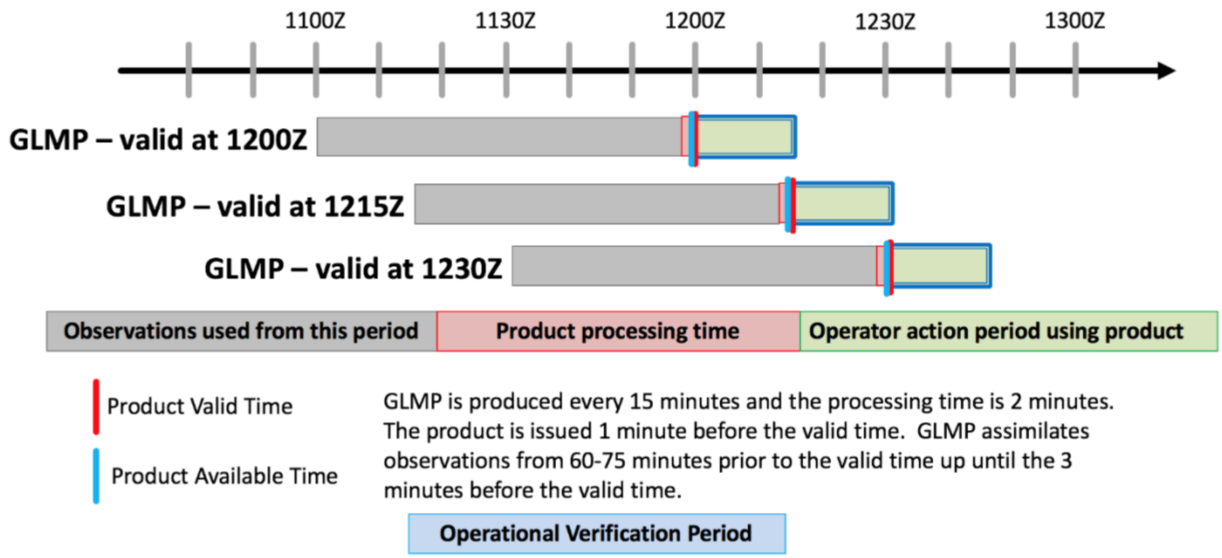


Figure 59. Operational verification period applied to GLMP.

9.4 RTMA Timelines

The following figures show the timeline of the RTMA and the application of the Meteorological Verification Period and Operational Verification Period.

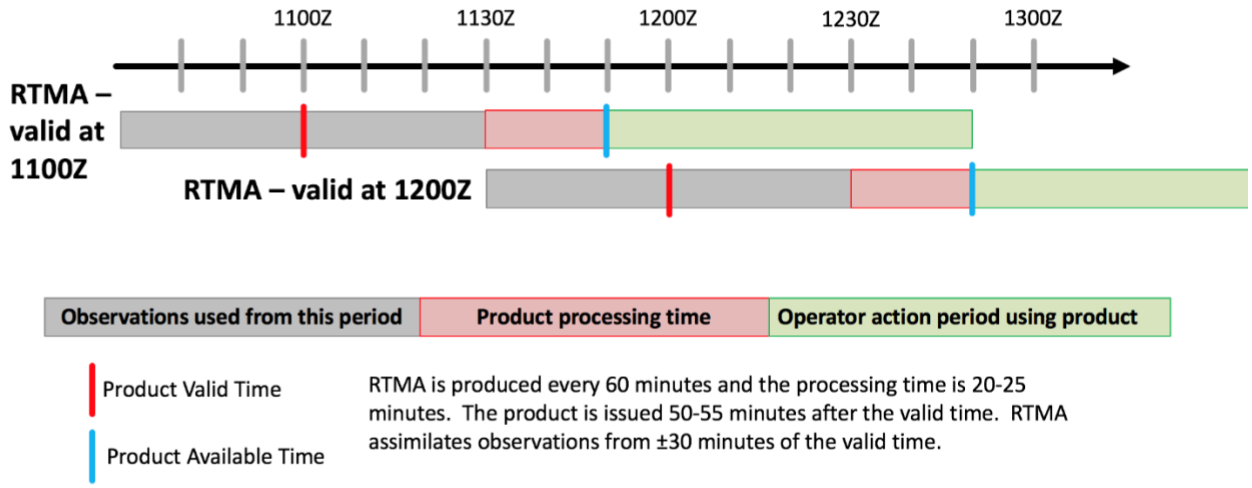


Figure 60. RTMA availability timeline.

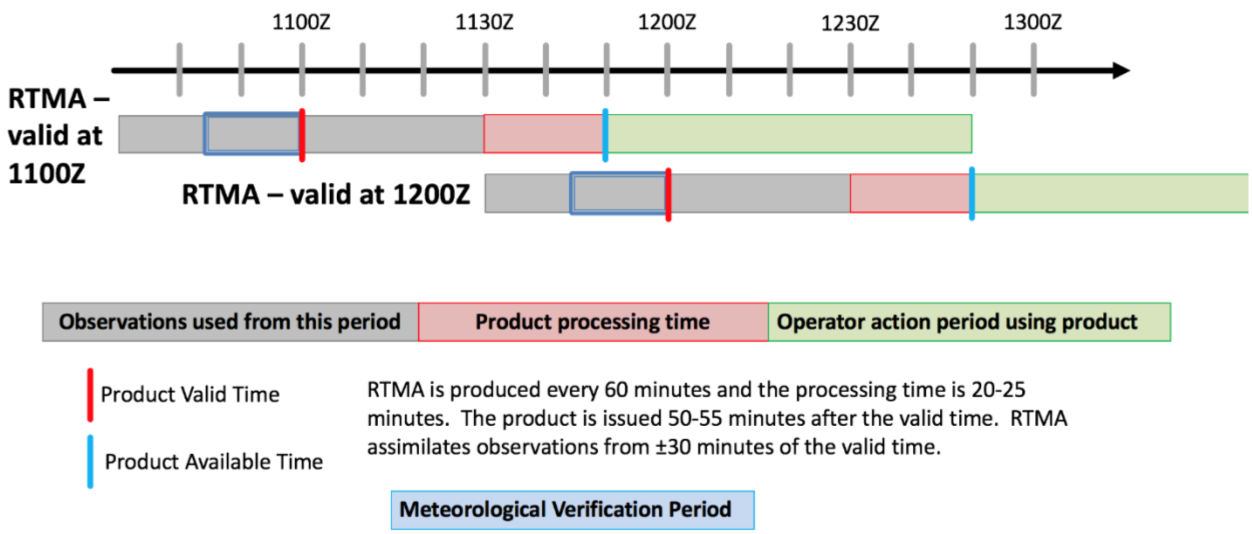


Figure 61. Meteorological verification period applied to RTMA.

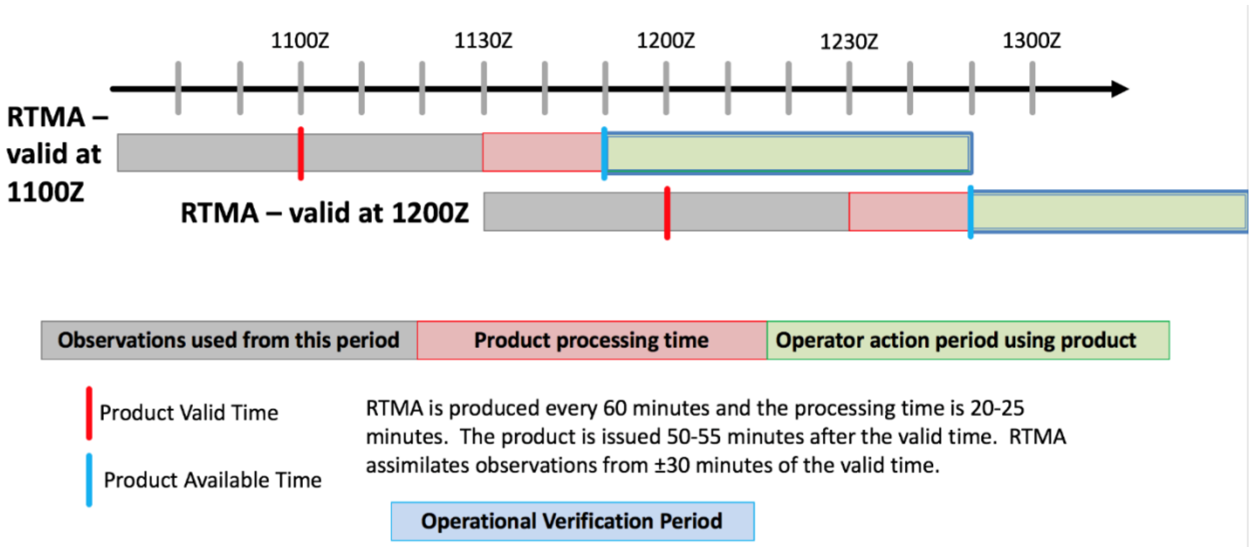


Figure 62. Operational verification period applied to RTMA.

9.5 RTMA-RU Timelines

The following figures show the timeline of the RMTA-RU and the application of the Meteorological Verification Period and Operational Verification Period.

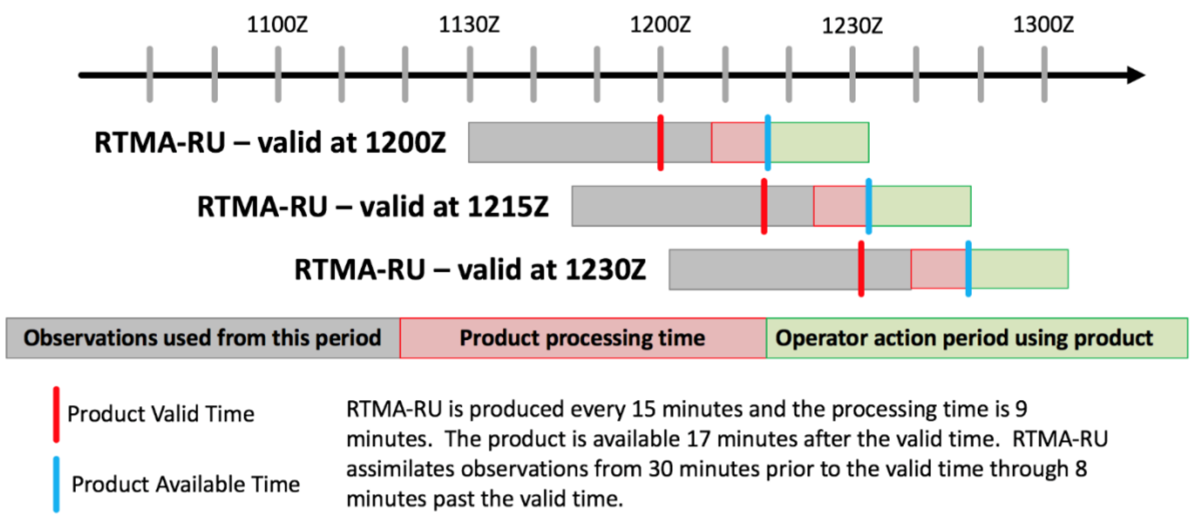


Figure 63. RTMA-RU availability timeline.

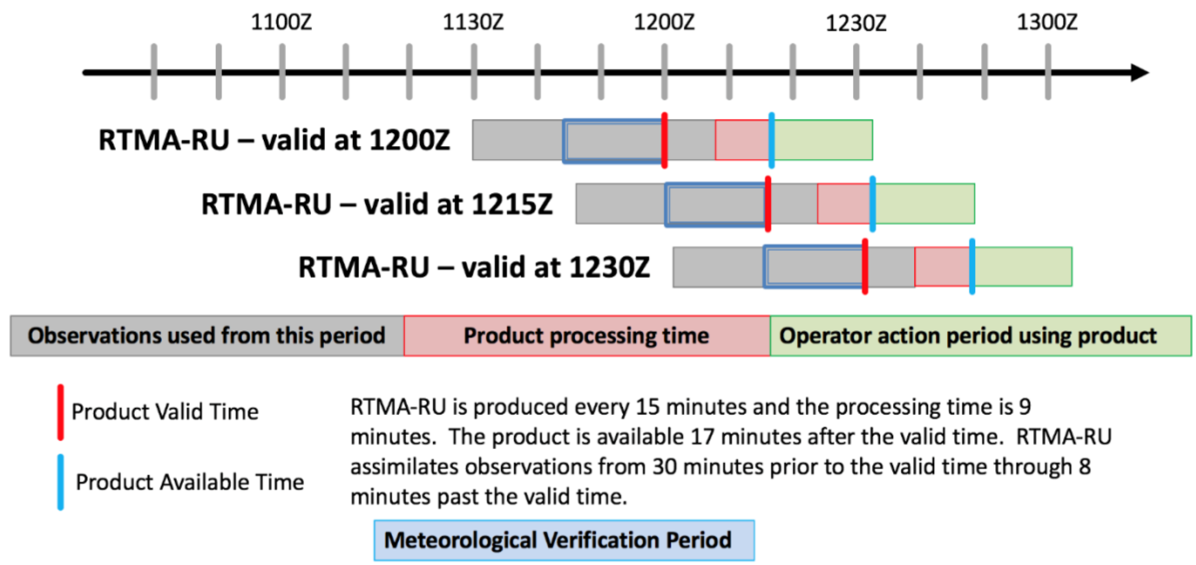


Figure 64. Meteorological verification period applied to RTMA-RU.

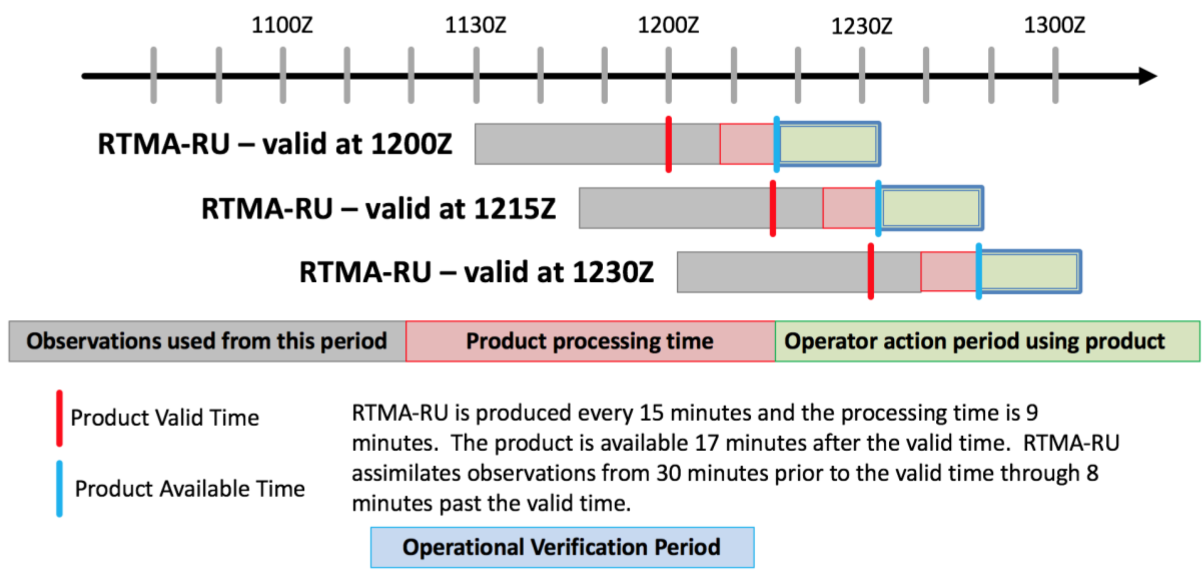


Figure 65. Operational verification period applied to RTMA-RU.

9.6 Geographic Distributions – Ceiling

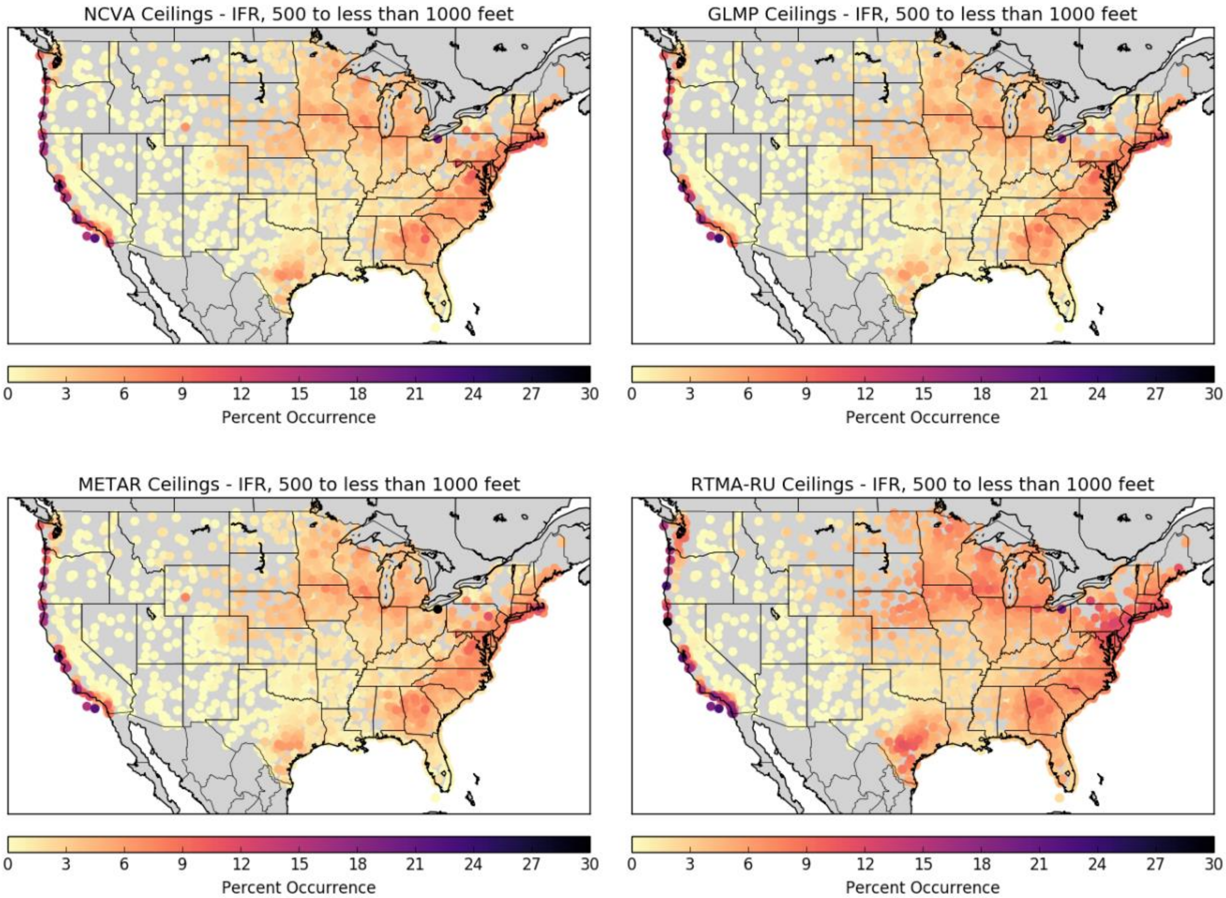


Figure 66. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the IFR ceiling flight category.

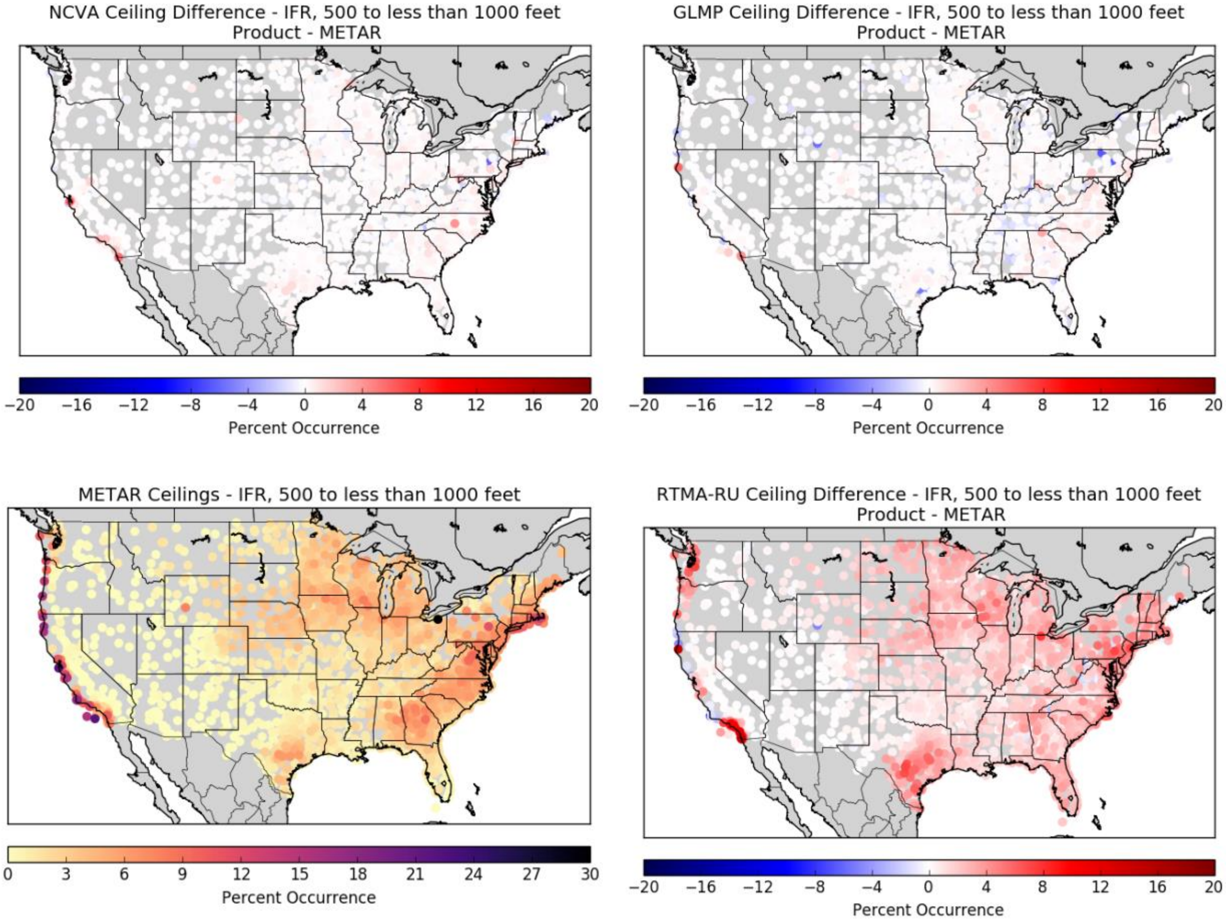


Figure 67. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the IFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

9.7 Geographic Distributions – Visibility

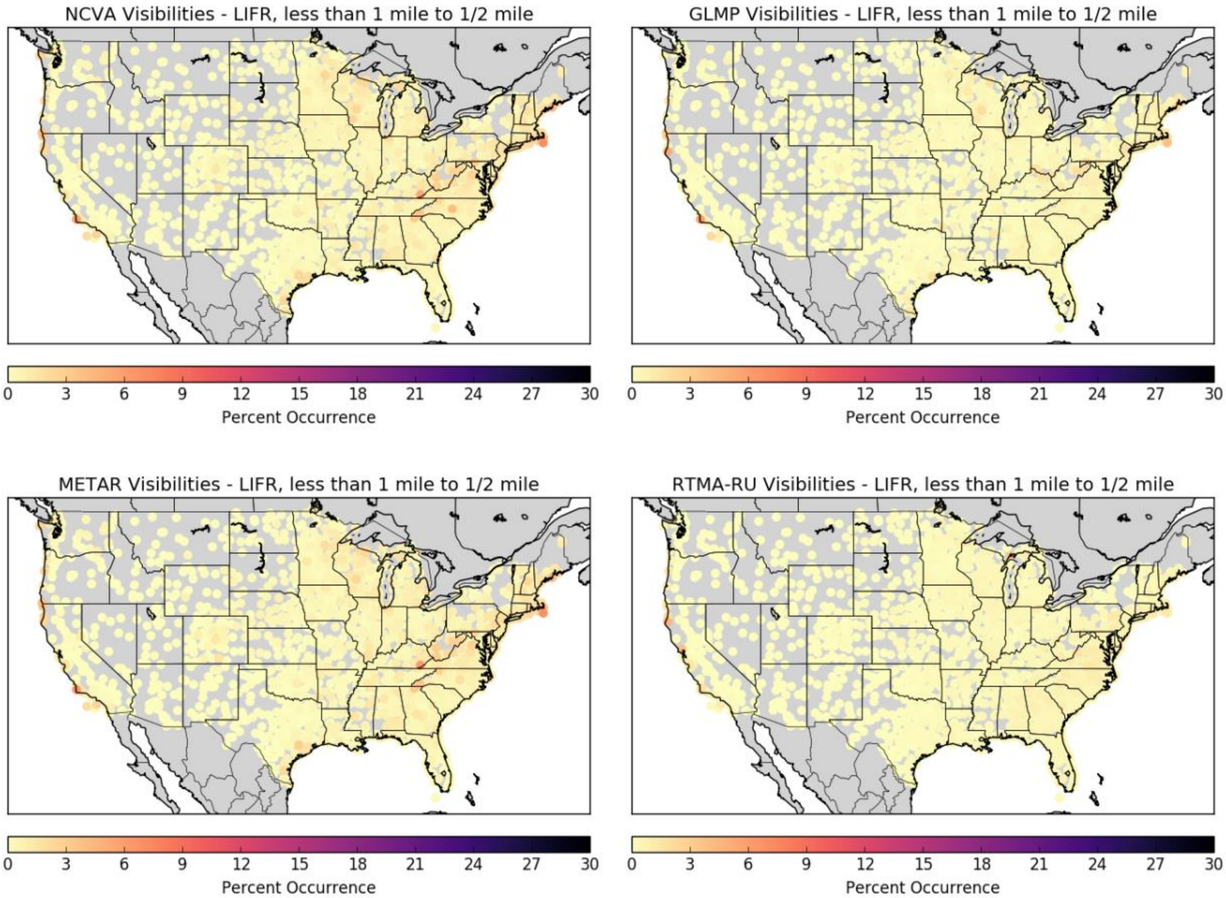


Figure 68. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the upper LIFR visibility flight category.

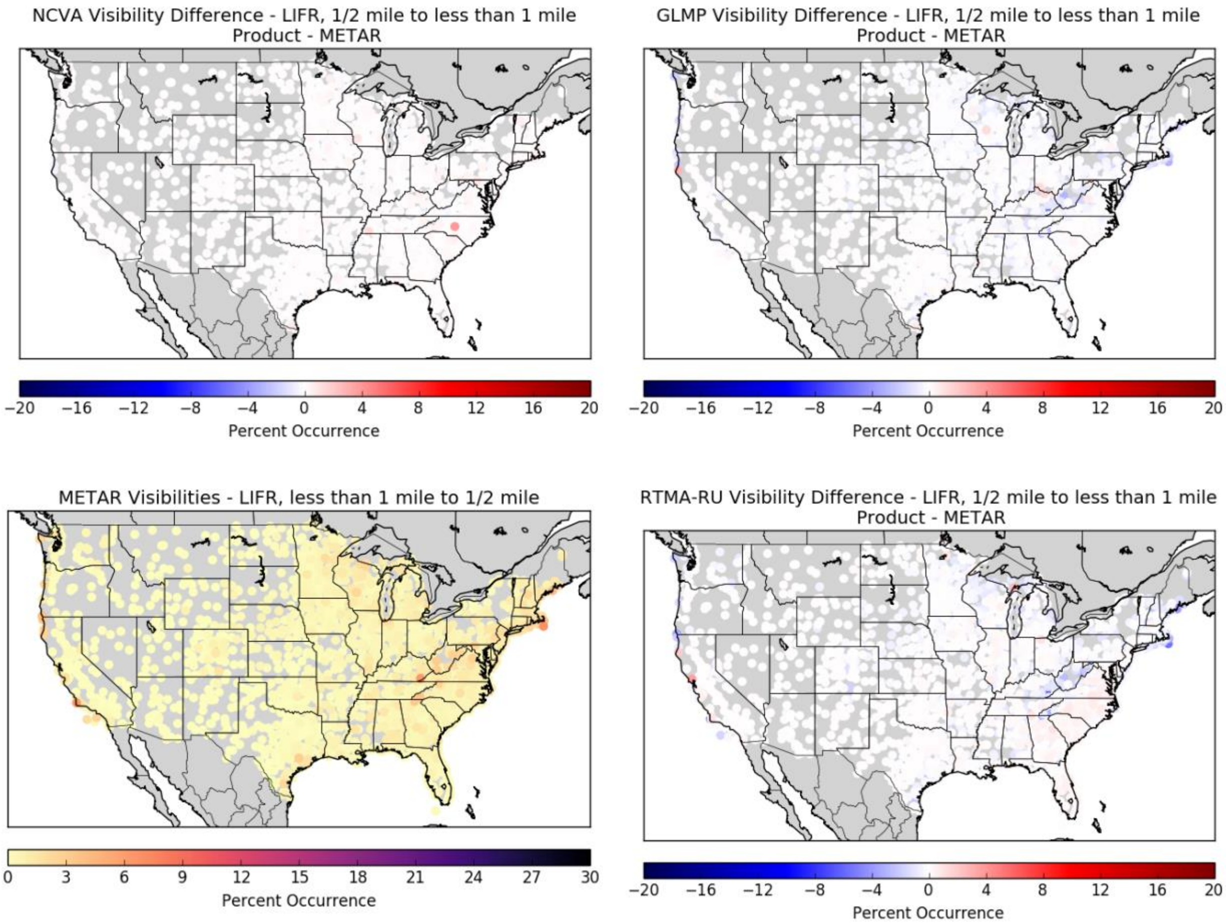


Figure 69. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the LIFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

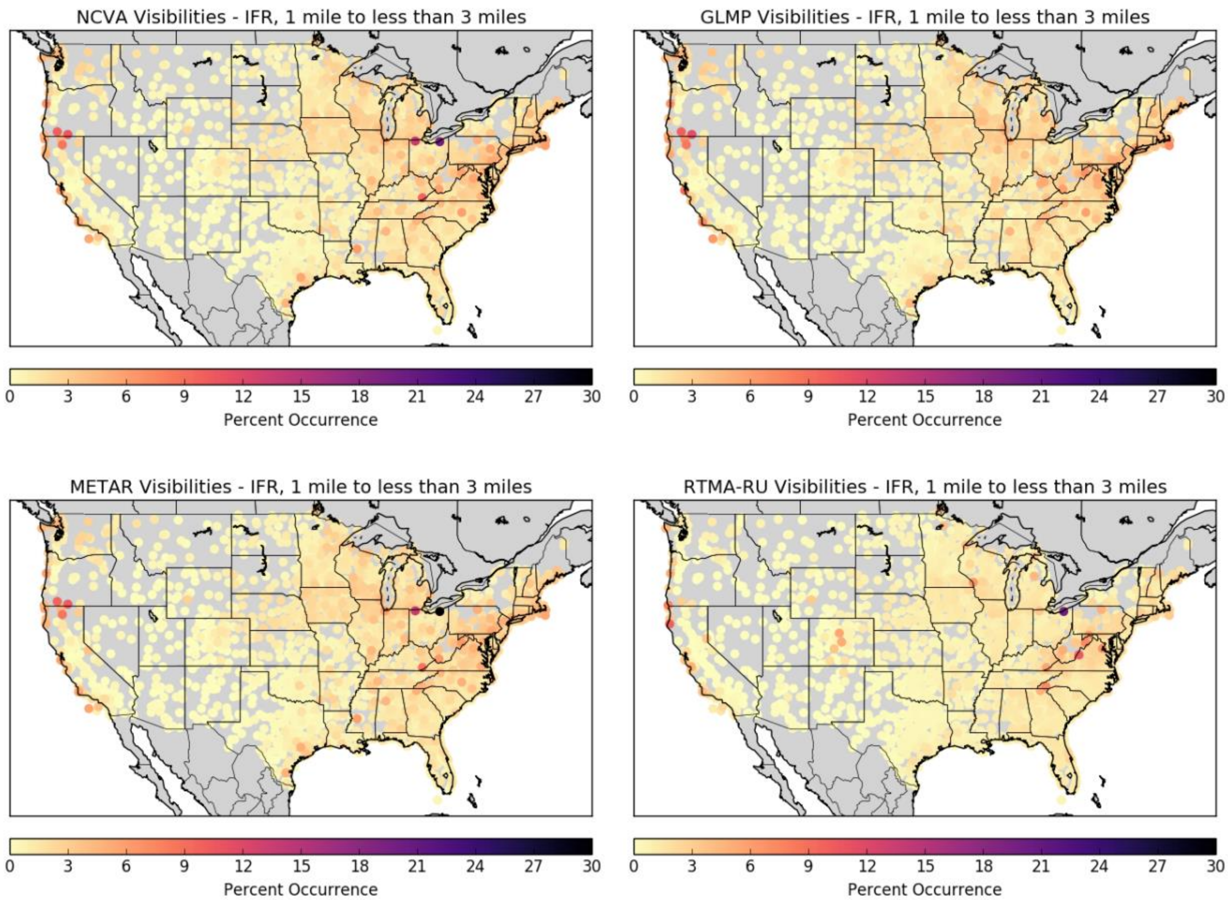


Figure 70. Geographic distributions for NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) compared against METARs (lower left) for the IFR visibility flight category.

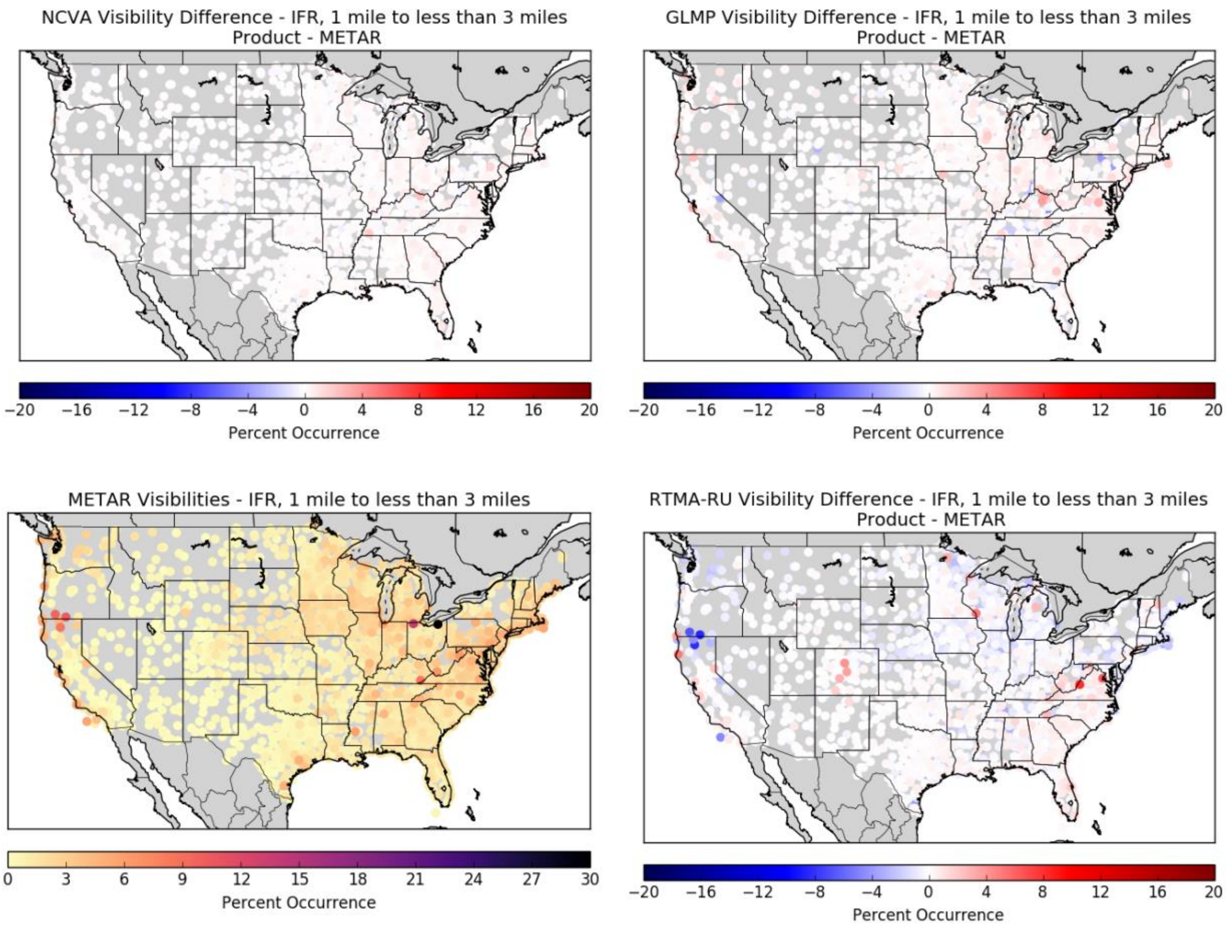


Figure 71. Difference maps of the geographic distributions of NCVA (upper left), GLMP (upper right), and RTMA-RU (lower right) for the IFR flight category. The METAR geographic distribution map (lower left) is shown for reference.

9.8 Skill Scores – Meteorological Period Ceiling

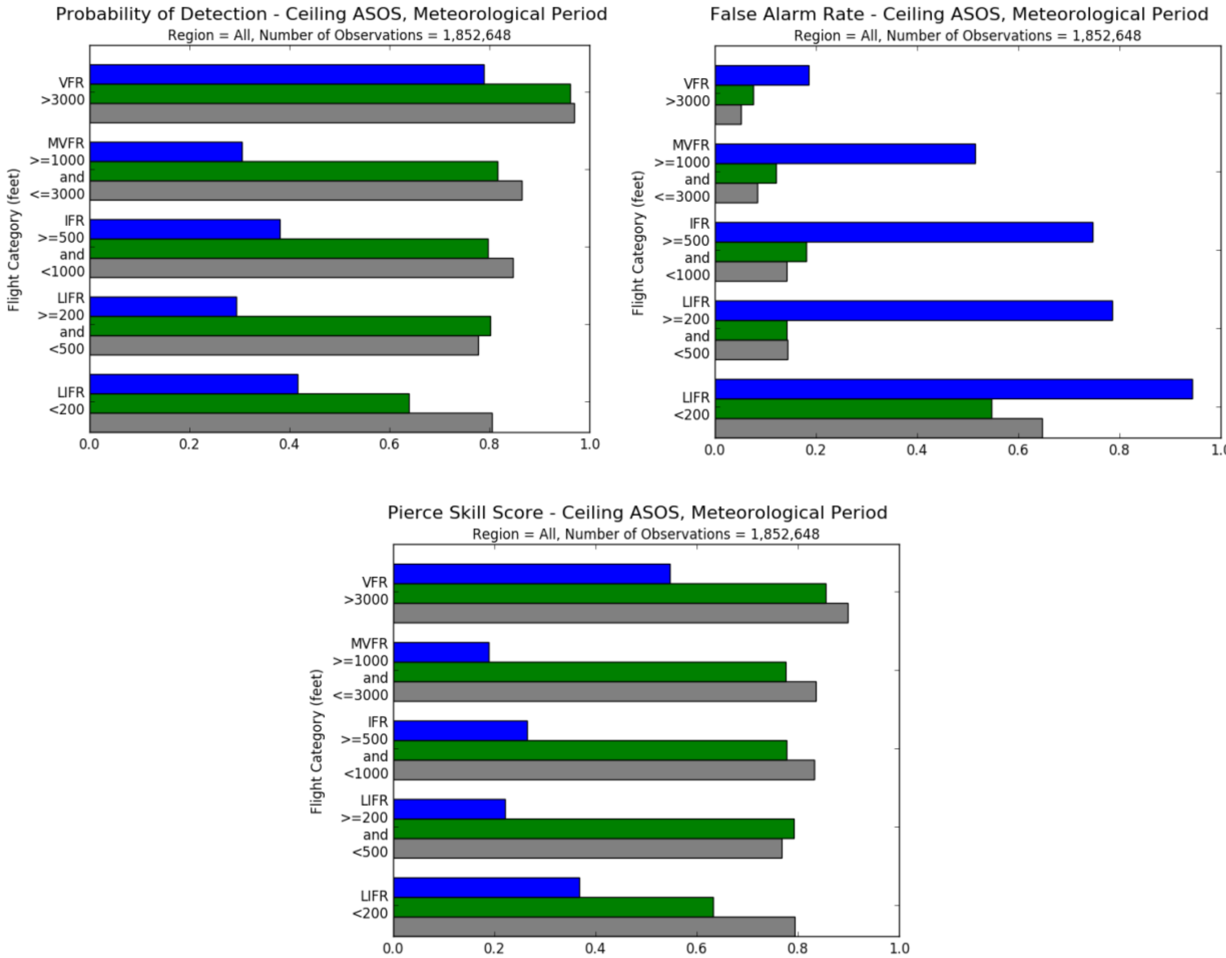


Figure 72. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS ceiling observations during the meteorological verification period.

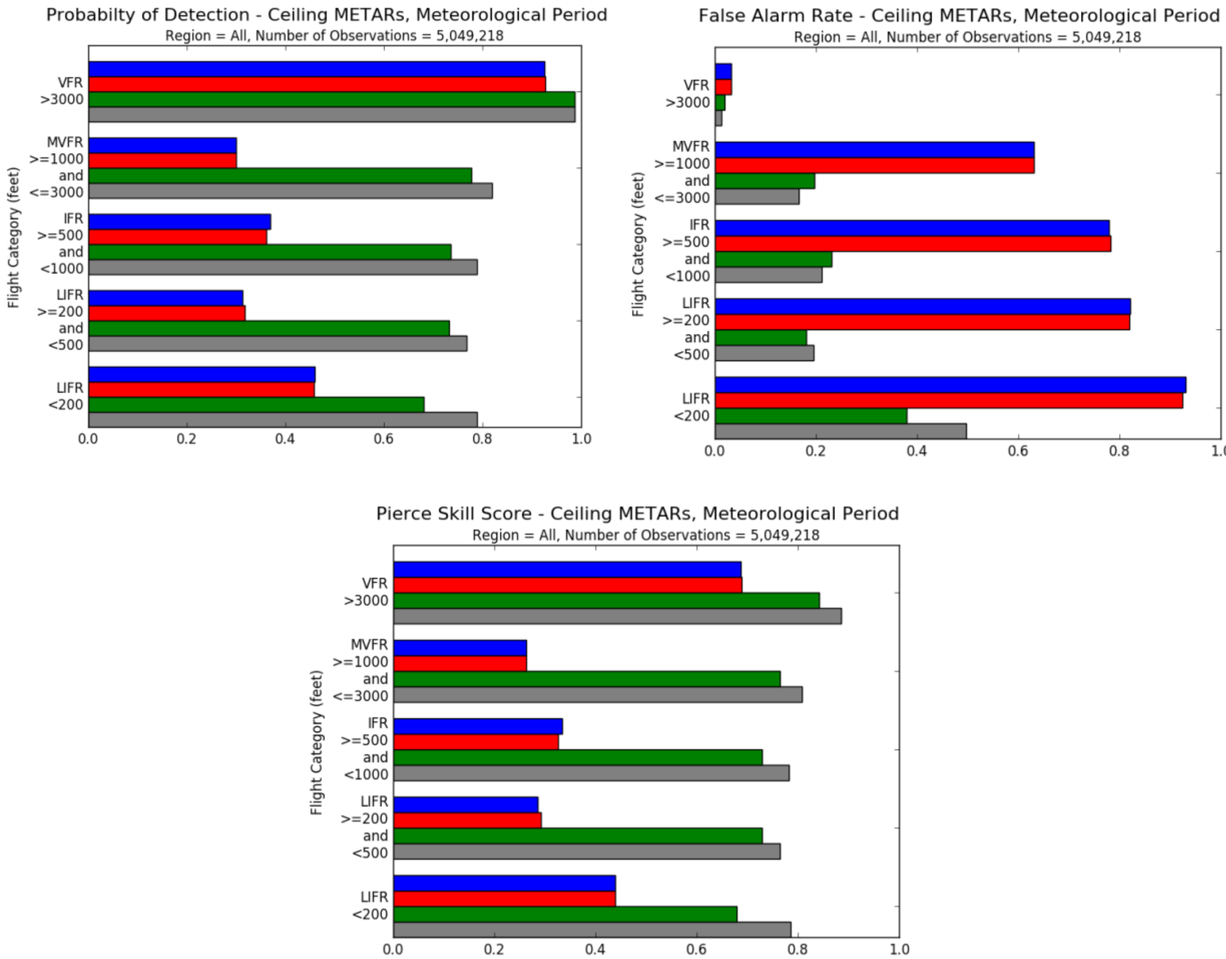


Figure 73. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet ceiling observations during the meteorological verification period.

9.9 Skill Scores – Meteorological Period Visibility

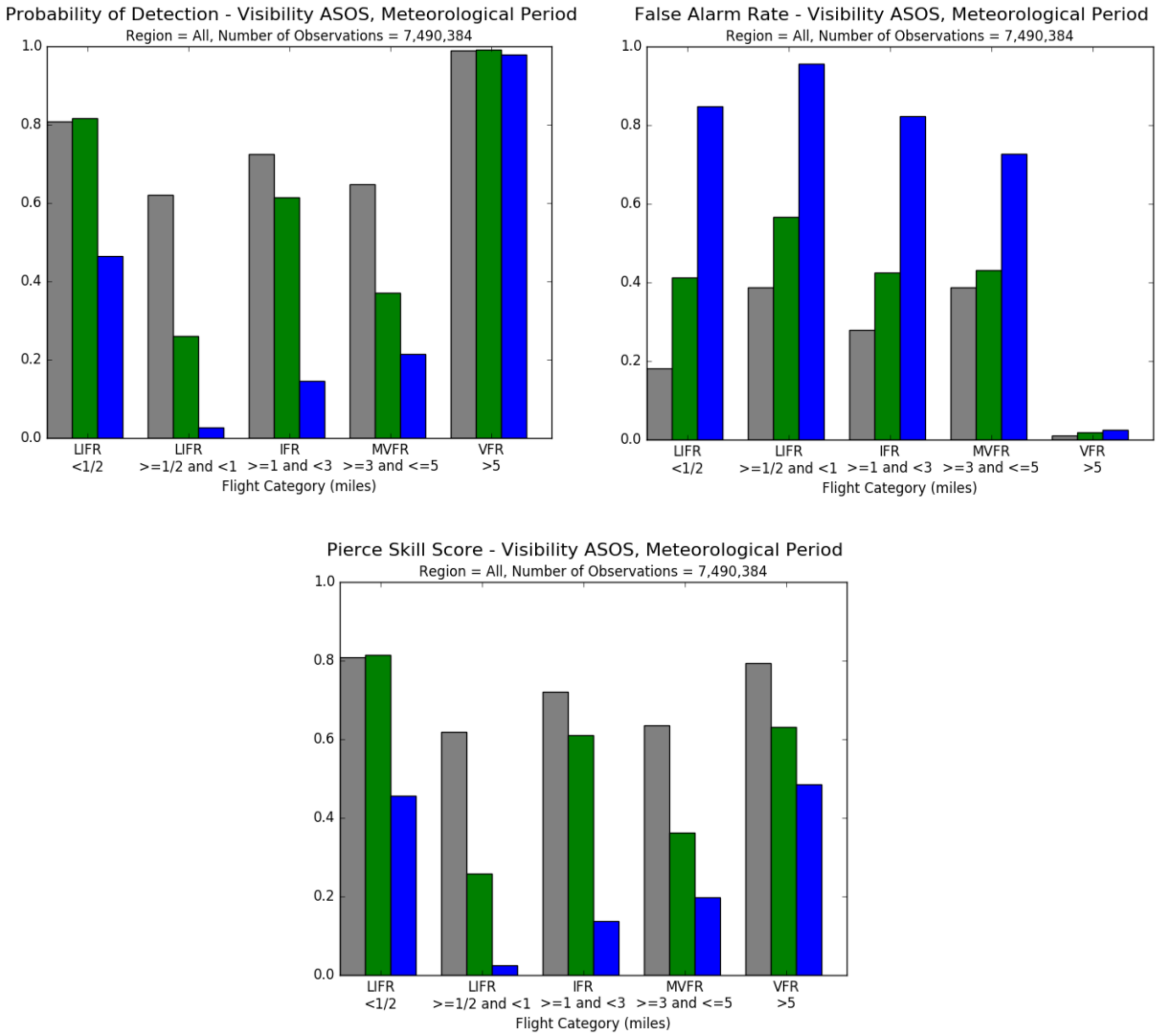


Figure 74. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS visibility observations during the meteorological verification period.

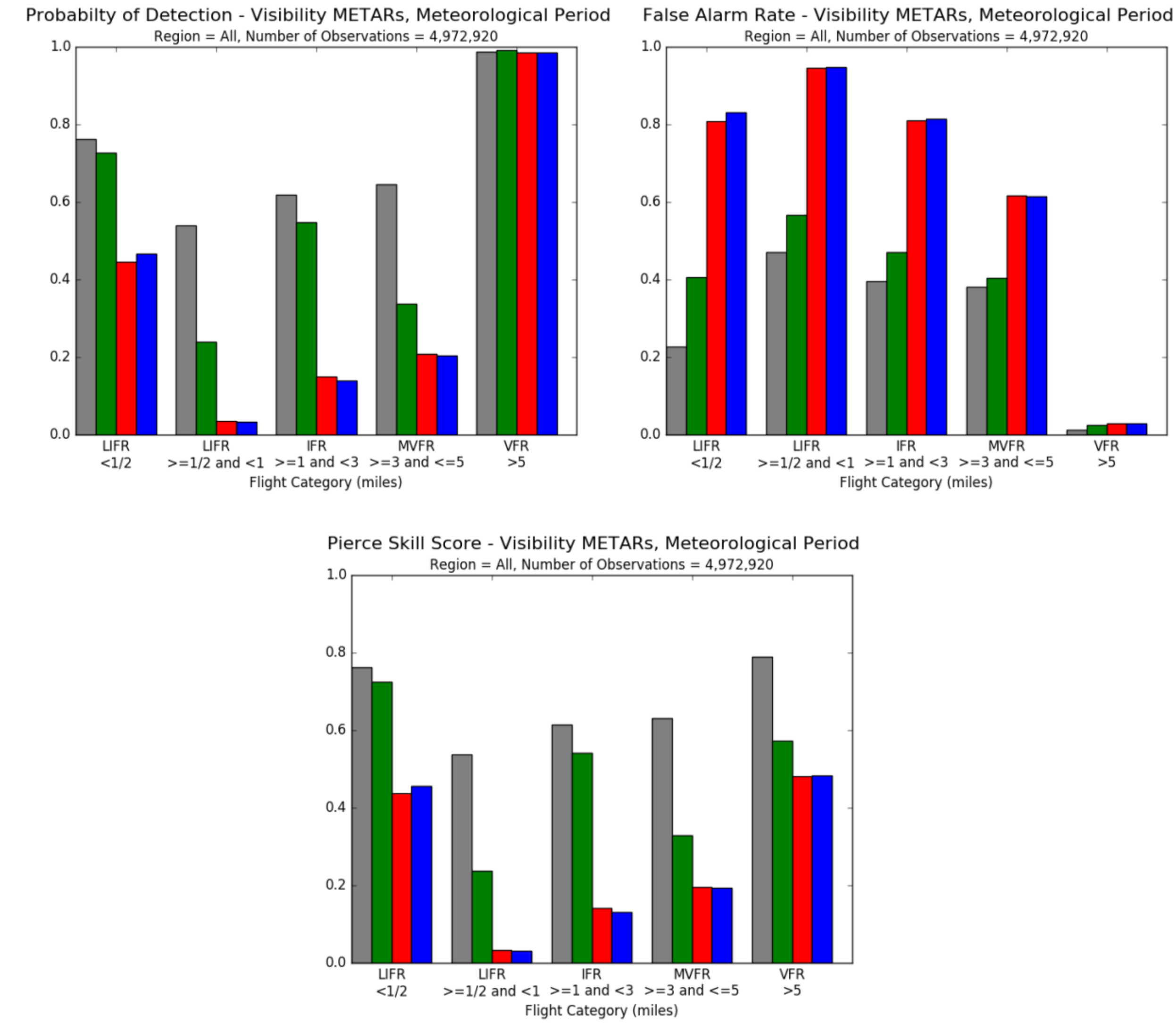


Figure 75. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against METAR visibility observations during the meteorological verification period.

9.10 Skill Scores – Operational Period Ceiling

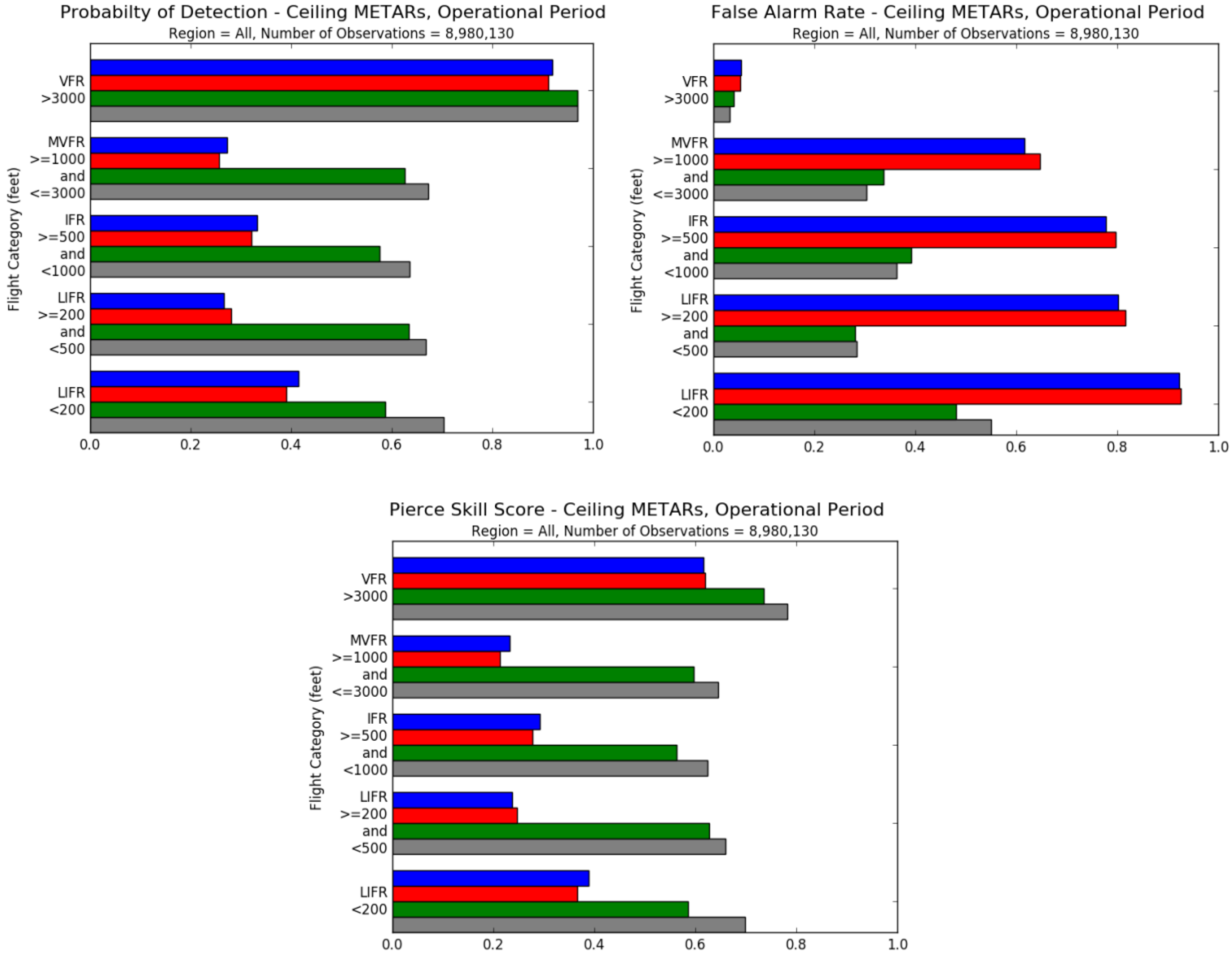


Figure 76. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against METAR ceiling observations during the operational verification period.

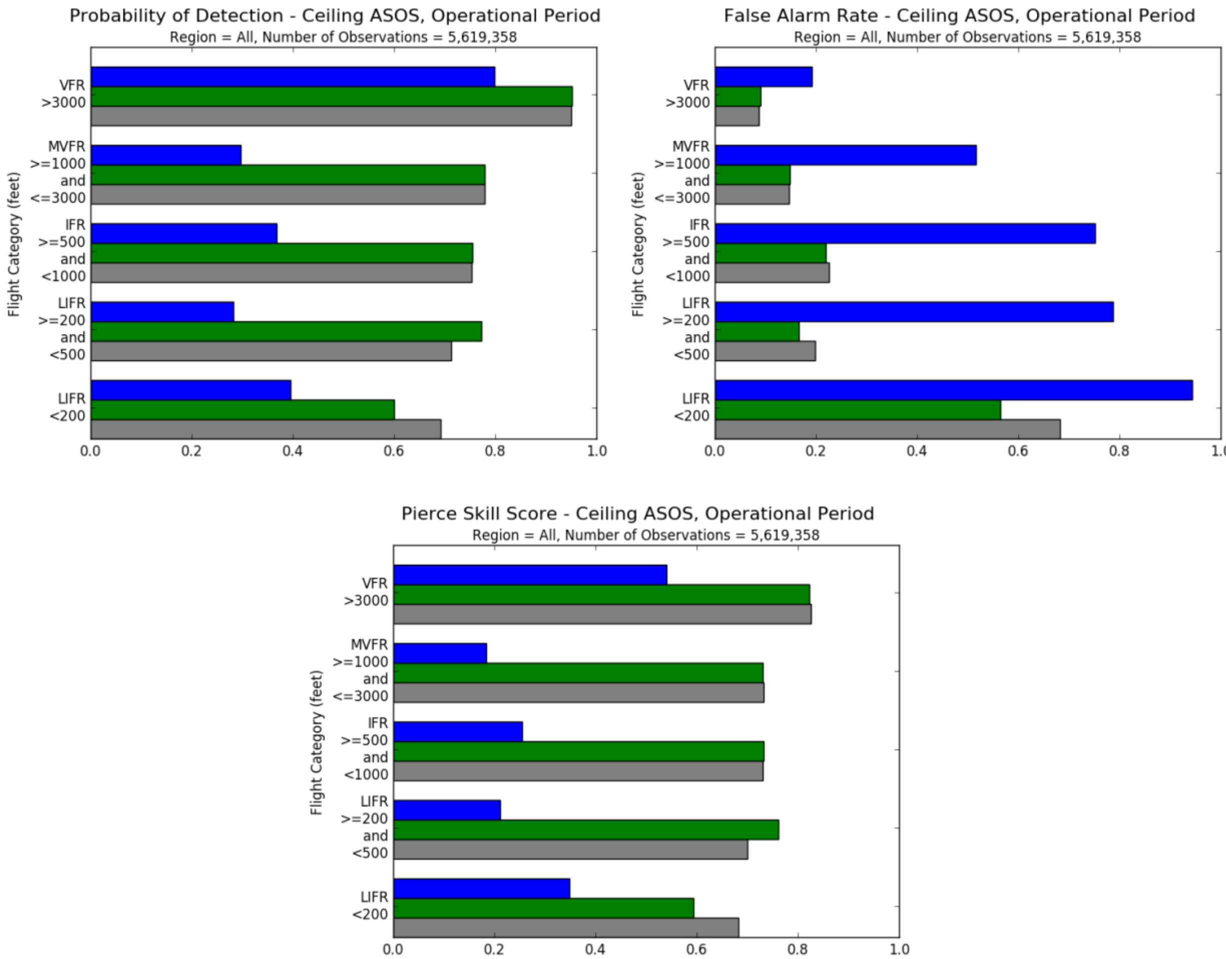


Figure 77. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS ceiling observations during the operational verification period.

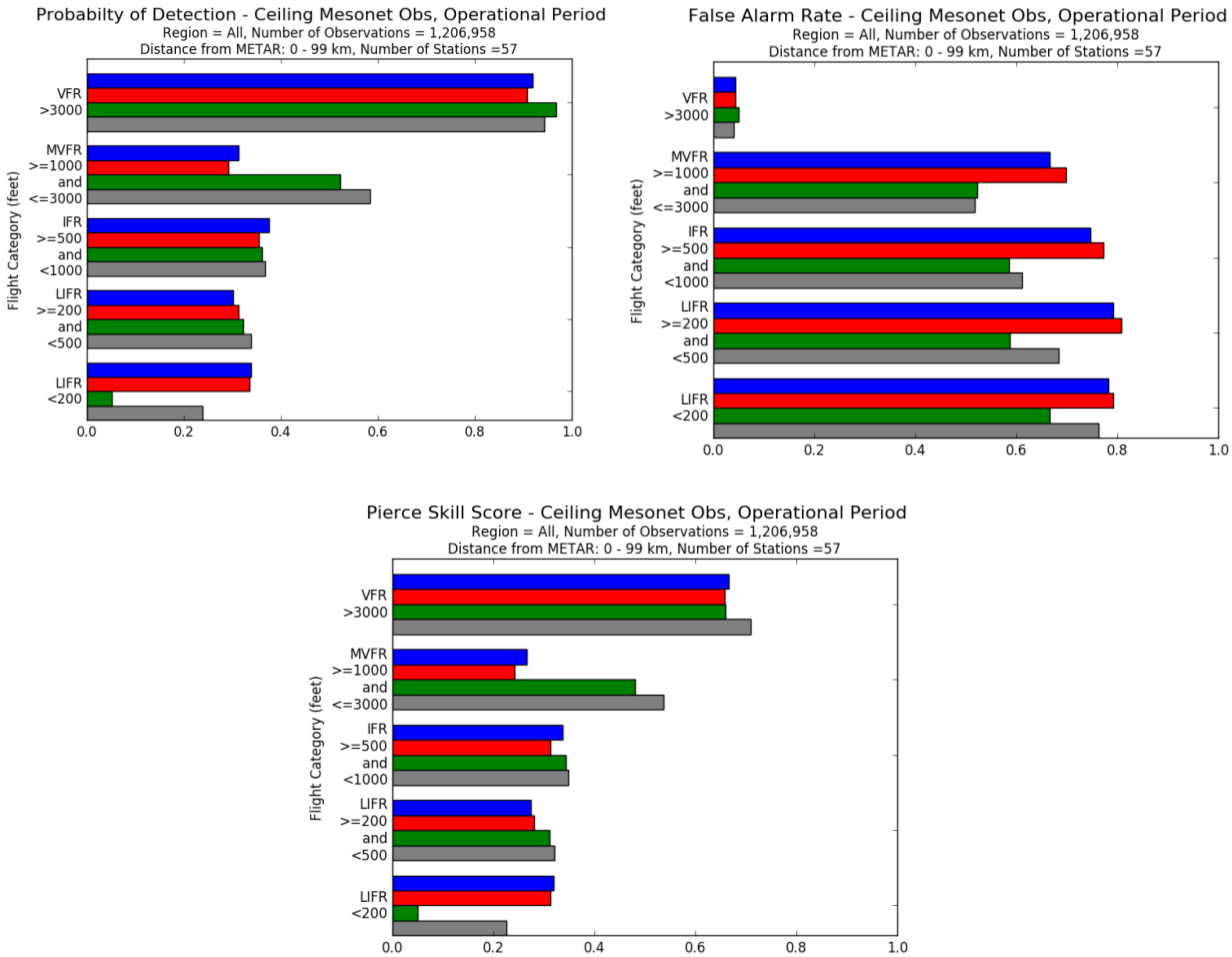


Figure 78. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet ceiling observations during the operational verification period.

9.11 Skill Scores – Operational Period Visibility

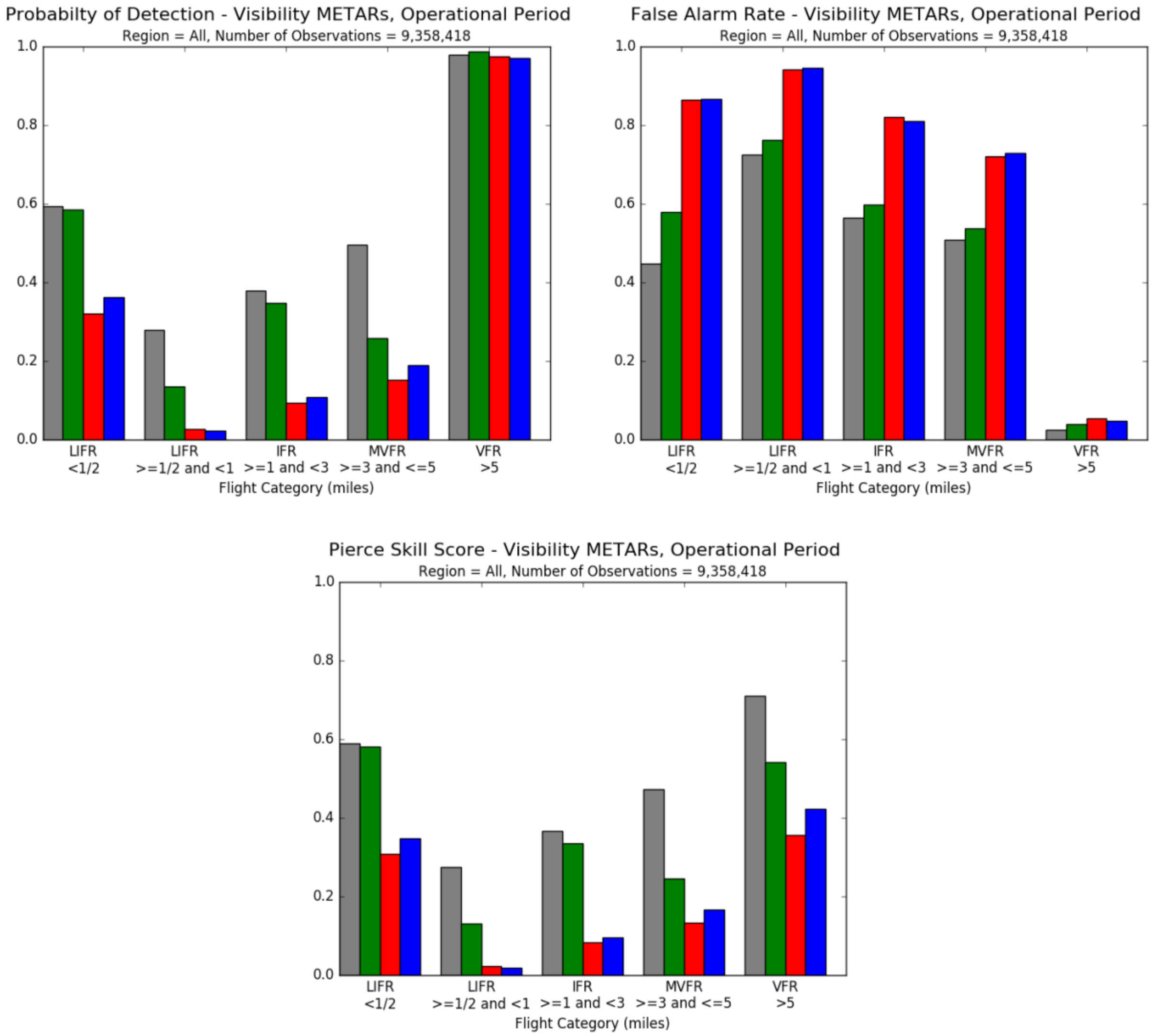
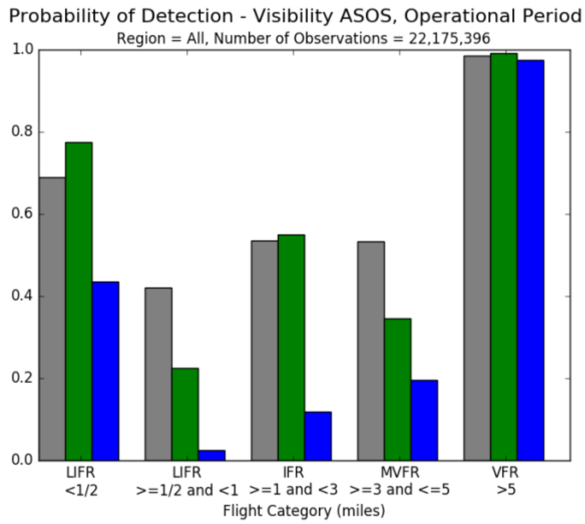
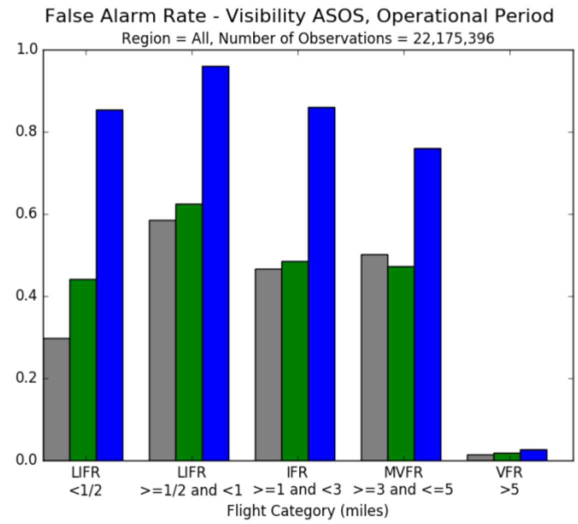


Figure 79. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against METAR visibility observations during the operational verification period.





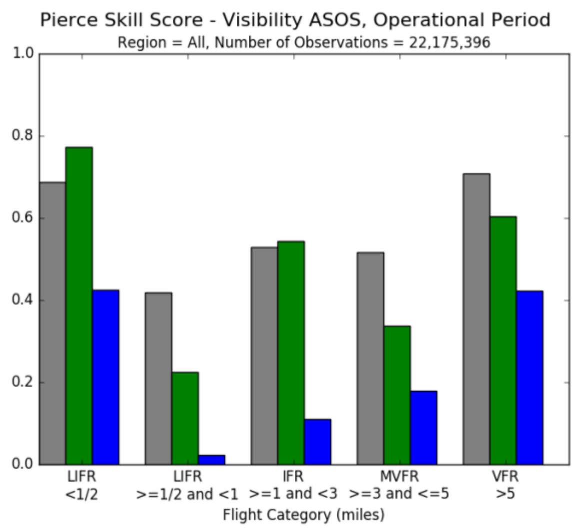


Figure 80. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), and RTMA-RU (blue) measured against ASOS visibility observations during the operational verification period.

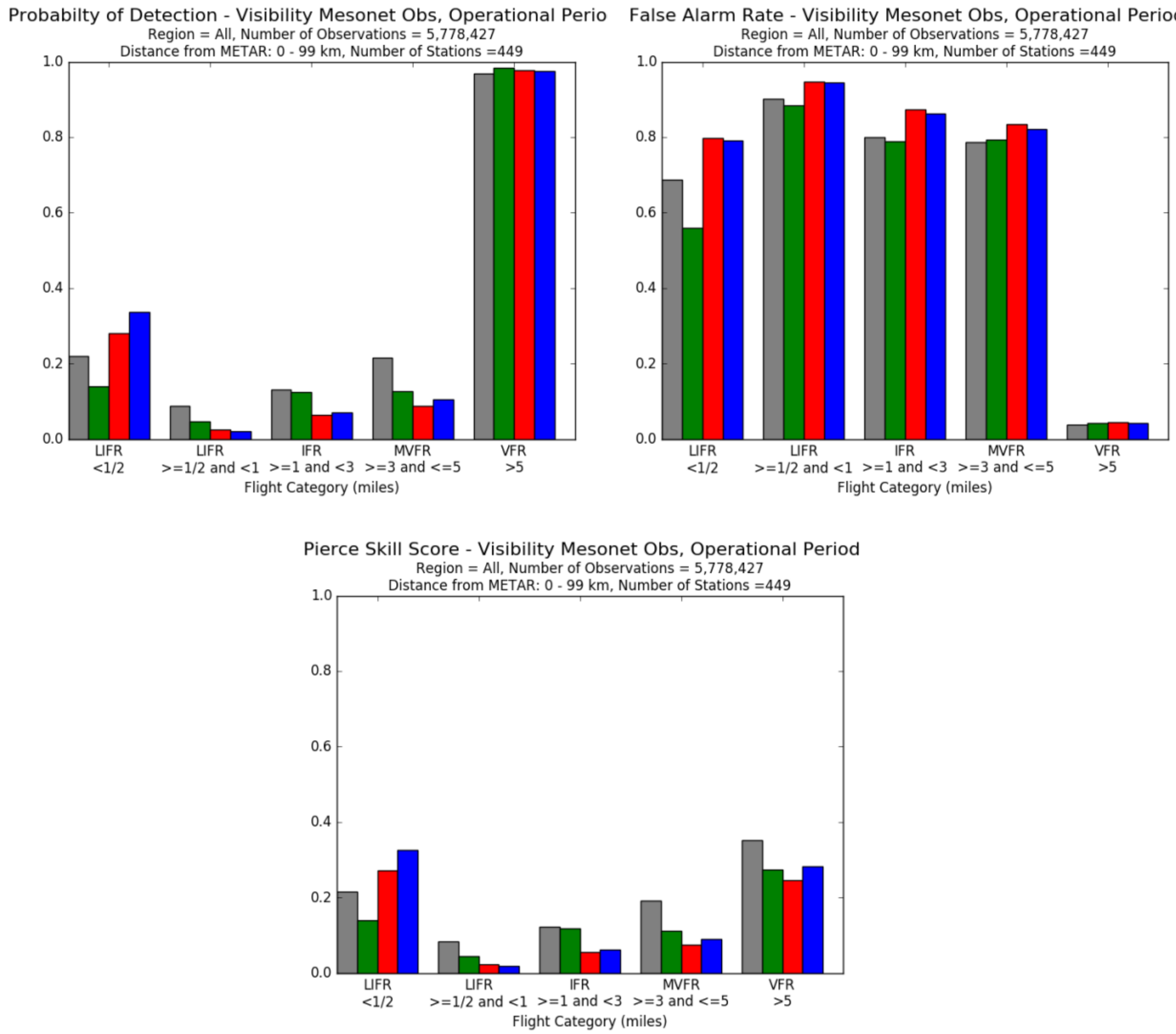


Figure 81. POD (upper left), FAR (upper right), and PSS (bottom) for NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) measured against mesonet visibility observations during the operational verification period.

9.12 Skill Scores – Regional

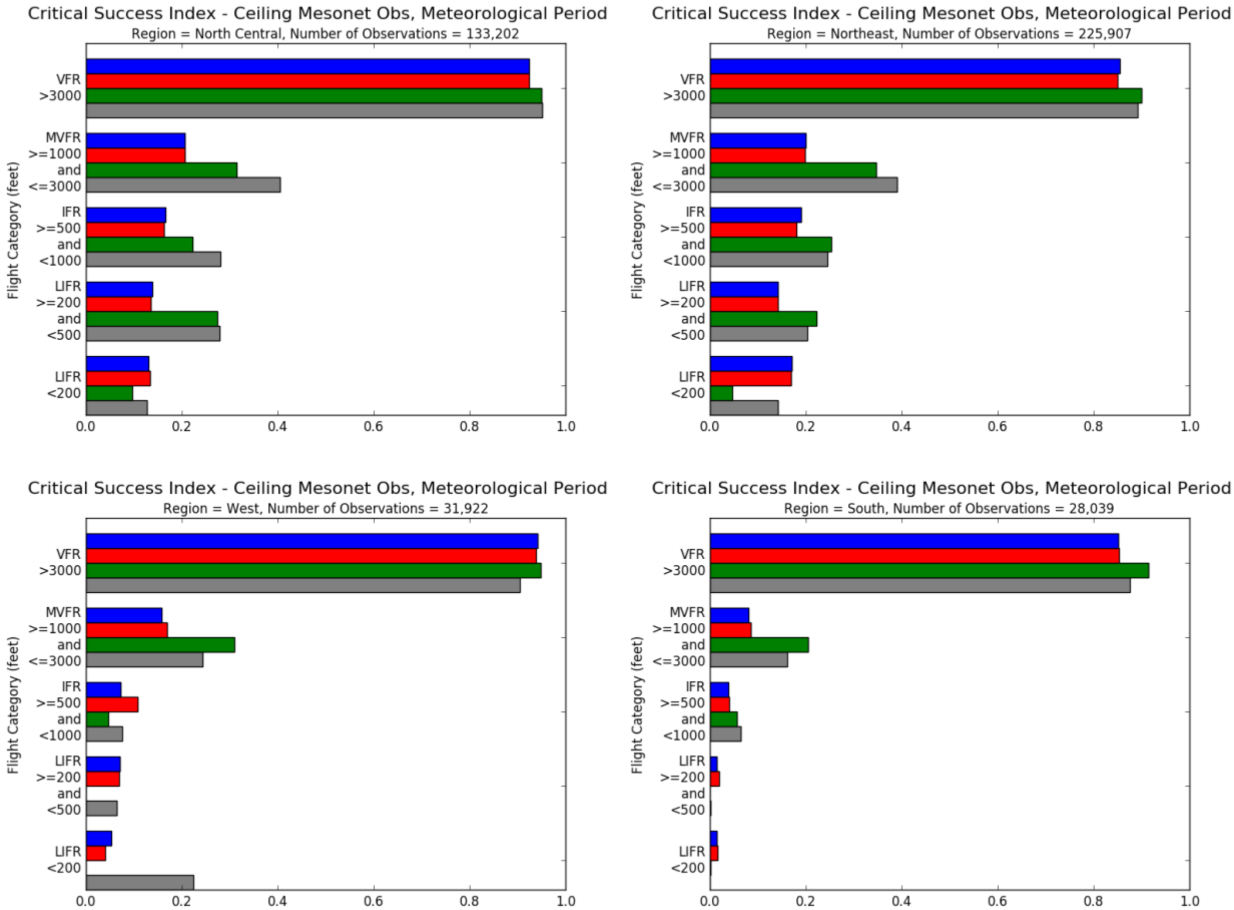


Figure 82. CSI in the North Central (upper left), Northeast (upper right), West (lower left), and South (lower right) regions of NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) during the meteorological verification period measured against mesonet ceiling observations. Note that the West and South regions have sample sizes an order of magnitude less than the North Central and Northeast.

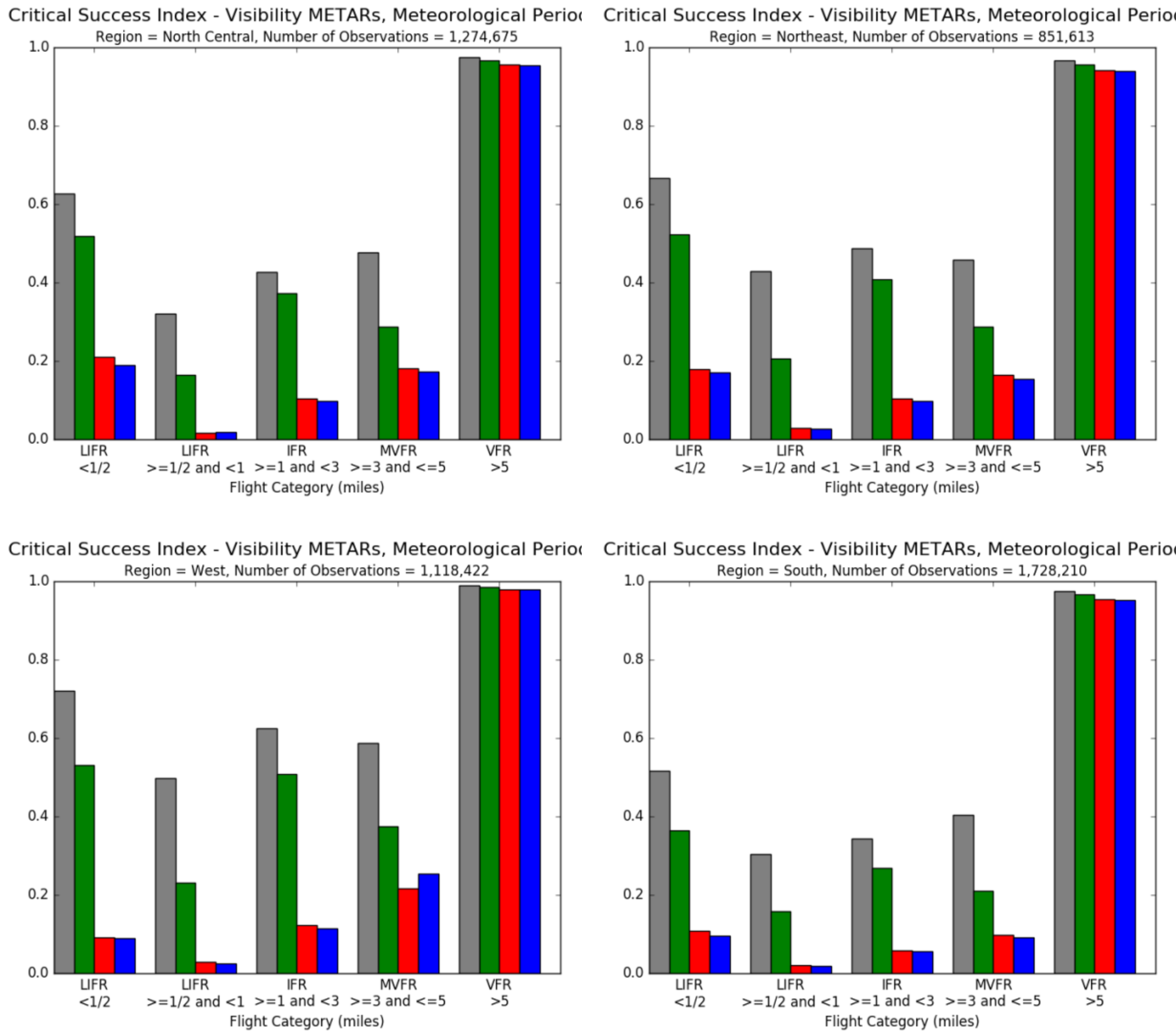


Figure 83. CSI in the North Central (upper left), Northeast (upper right), West (lower left), and South (lower right) regions of NCVA (gray), GLMP (green), RTMA (red), and RTMA-RU (blue) during the meteorological verification period measured against METAR visibility observations.

10 References

Ceiling and Visibility Display. http://www.aviationweather.gov/cva/help.

Fenton, K., G. Layne, and M. Petty, 2017. Investigation of Data Sources for the Verification of Ceiling and Visibility. Report submitted to the FAA.

Ghirardelli, J., J. Charba, J. Im, F. Samplatsky, and B. Glahn, 2015. Improving Localized Aviation MOS Program (LAMP) Guidance by Utilizing Emerging Forecast and Observation Resources.

Harris Corporation, 2017: Product Definition and User's Guide (PUG) Volume 5: Level 2+ Products. Product Definition and User's Guide (PUG), 677 pp, http://www.goes-r.gov/products/docs/PUG-L2+-vol5.pdf.

Heidinger, A., M. Pavolonis, P. Minnis, A. Walther, and S. Lindstrom, 2016: GOES-R Cloud and Microphysical Products, Fog and Low Stratus. University of Wisconsin-Madison CIMSS, accessed 5 December 2017, cimss.ssec.wisc.edu/goes/goes-r/training/working/GOES_R_CloudProducts.pptx.

GOES-R, 2018: GOES-16 ABI L2+ Clear Sky Mask Release, Provisional Data Quality: Read-Me for Data Users. Accessed 29 August 2018, https://www.ncdc.noaa.gov/sites/default/files/attachments/ABI_L2_CSM_Provisional_ReadMe.pdf.

Myrick, D., 2017. RTMA, URMA, and NAM DNG upgrade and Introduction of RTMA with Rapid Updates: Effective December 12, 2017.

National Ceiling and Visibility Analysis (NCVA) Product Product Description Document (PDD). http://www.nws.noaa.gov/infoservicechanges/NCVAPDD2.pdf.

Pondeca, M., S. Levine, J. Carley, Y. Lin, Y. Zhu, J. Purser, J. Mcqueen, J. Whiting, R. Yang, A. Gibbs, D. Parrish, and G. DiMego. Ongoing Improvements to the NCEP Real Time Mesoscale Analysis (RTMA) and UnRestricted Mesoscale Analysis (URMA) and NCEP/EMC.

Schmit, T., P. Griffith, M. Gunshor, J. Daniels, S. Goodman, and W. Lebair, 2017: A closer look at the ABI on the GOES-R series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, doi:https://doi.org/10.1175/BAMS-D-15-00230.1.

Yang, R., M. Pondeca, S. Levin, G. DiMego, J. Whiting, J. Carley, G. Manikin, and A. Gibbs. Development of a Rapid-Update Real-Time Mesoscale Analysis of Ceiling and Visibility.