9.5 COMPARISON OF RADAR-BASED PRECIPITATION MOSAICS

Joseph A. Grim*, Matthias Steiner, James O. Pinto, Ken Stone and Daniel Megenhardt
National Center for Atmospheric Research – Research Applications Laboratory, Boulder, CO

1. Introduction

Current and accurate observations, as well as quality near-term forecasts, of intense convection are vital to aviation safety and efficiency (e.g., Stern et al. 1994, Evans and Ducot 2006, JPDO 2007, Zhang et al. 2011). Vertically integrated liquid water (VIL) has often been used as an indicator of storm intensity (Shafer et al. 2000, Robinson et al. 2002) and in predicting the presence of hail, which is a significant aviation hazard (e.g., Kitzmiller et al. 1995, Billet et al. 1997). Greene and Clark (1972) first defined the term VIL as the integral of the total mass of liquid water in the air above an area on the ground, as detected by radar and assuming a Marshall-Palmer (Marshall and Palmer 1948) water droplet size distribution. Radar echo tops (ETs) are of importance to pilots in determining if they can fly over, or have to deviate around, a storm (e.g., Evans et al. 2004, DeLaura and Evans 2006). In the early days of weather radar, ET was defined as the maximum height of the minimum detectable echo (Donaldson 1964), though it has since come to be more precisely defined as the uppermost altitude above a point at which the reflectivity equals a given threshold (e.g., 18 dBZ).

There are several sources available for current and short-term forecasts of VIL and ETs. The focus of this study is on two such two suites of VIL and ET products, each of which have been created for different primary purposes. One suite is from the Corridor Integrated Weather System (CIWS, Evans and Ducot 2006), created at the Massachusetts Institute of Technology – Lincoln Laboratory (MIT-LL), and intended mainly for use by the aviation community. The other suite of products is from the Multiple-Radar / Multiple-Sensor (MRMS) dataset (Zhang et al. 2011), developed by the NOAA National Severe Storms Laboratory (NSSL) and produced at the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC). The primary focus of the MRMS products is on quantitative precipitation estimation for use as guidance for flood and flash flood warnings. From both product suites, we have focused on comparisons of VIL and 18 dBZ echo tops (hereafter ET refers to 18 dBZ echo tops) for the contiguous United States, for both observations and forecasts up to 2 hours into the future.

2. Description of Datasets

Klingle-Wilson and Evans (2005) describe how the CIWS VIL and ET products are created, while Zhang et al. (2011) describe much of how the MRMS VIL and ET products are created. Here, we'll presently only a brief summary of their techniques, particularly focusing on notable differences. First, reflectivity data are obtained from several radar sources (NEXRAD, Terminal Doppler Weather Radars, and Canadian Radars). Next, the reflectivity data are quality controlled (QCed). At this point, the techniques diverge. CIWS calculates VIL and ETs for each individual radar and then combines each field from all radars onto a mosaic covering the U. S. and southern Canada (Fig. 1a). MRMS, on the other hand, interpolates the QCed reflectivity from all radars onto a single three-dimensional grid, before calculating mosaic grids of VIL and ETs for roughly the same area as CIWS (Fig. 1b). VIL is calculated from reflectivity using the empirical formula developed by Greene and Clark (1972):

$$VIL = 3.44 \cdot 10^{-6} * \left[\frac{Z_i + Z_{i+1}}{2}\right]^{\frac{4}{7}} \Delta h, \qquad (1)$$

where Z is the radar reflectivity, i is the index of the level and Δh is the depth of the layer. ETs are calculated by determining the highest altitude above a given location that equals or exceeds 18 dBZ. Data (VIL and ETs for CIWS, and reflectivity for MRMS) from storms observed at slightly different times are advected to a common time, before they are combined to create the mosaics.

The CIWS VIL and ET products are stored on a Lambert Equal Area projection, so that each grid square has an area of exactly 1 km^2. On the other

* *Corresponding author address*: Joseph A. Grim, National Center for Atmospheric Research, Research Applications Laboratory, P.O. Box 3000, Boulder, CO 80307-3000; e-mail: grim@ucar.edu.

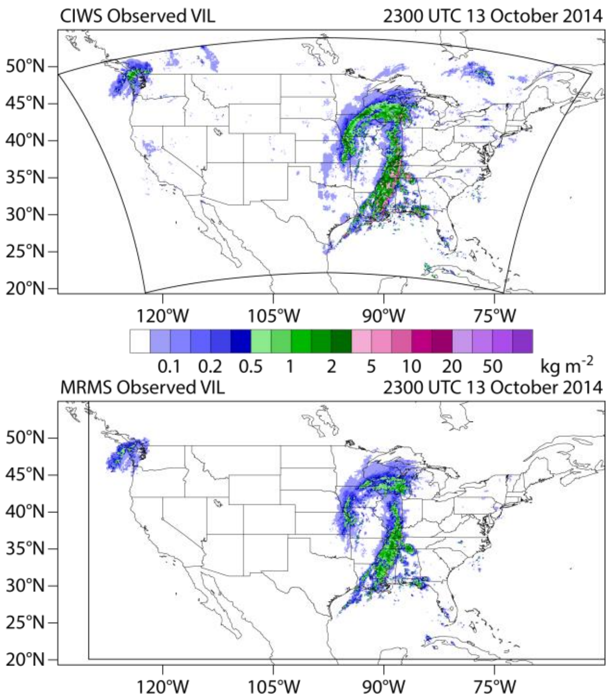

Figure 1. (top) CIWS and (bottom) MRMS VIL (kg m-2) at 2300 UTC 13 October 2014, displayed on a Cylindrical Equidistant projection. Outlines depicts limits of their native grids.

hand, MRMS products are stored on a cylindrical equidistant projection, where each grid square has an area of 0.01° × 0.01° (~1 km^2). New files are created at 2.5 minute intervals for CIWS and 2.0 minutes for MRMS. Example plots of CIWS and MRMS VIL, which also depicts their domains, are shown in Fig. 1, while Fig 2 shows example plots of ETs.

Forecast VIL and ET products for both CIWS and MRMS are created by advecting the observed VIL and ETs up to 2 hours into the future, using estimates of storm motion. For more information on how the CIWS forecast VIL and ET fields are advected, see Klingle-Wilson and Evans (2005) and Rappa and Troxel (2009). Lakshmanan et al. (2006) provide a description of how the MRMS fields are advected. Examples of the CIWS and MRMS forecast VIL and ET fields, along with their initial observed fields are shown in Figs. 3 and 4; these plots are zoomed in on a particular area, in order to be able to better show the advection and evolution of the fields.

3. Data Management and Quality Control

We had data feeds set up for both CIWS and MRMS, for observed and forecast fields of both

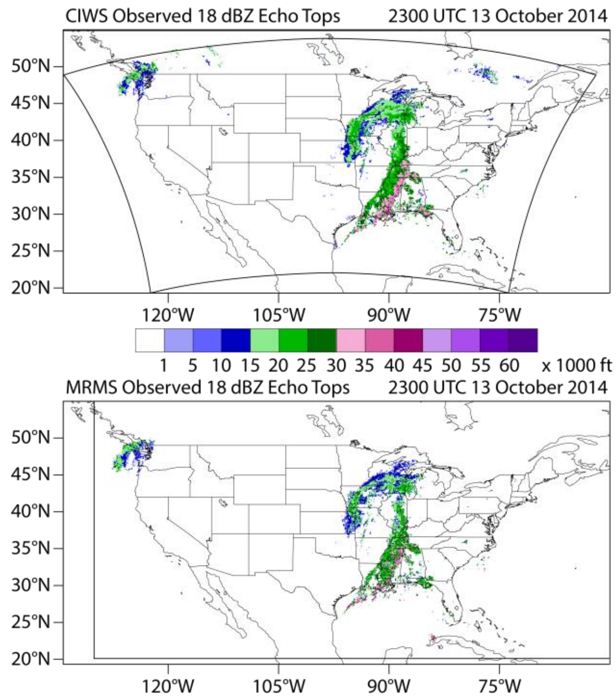

Figure 2. Same as Fig. 1, except for 18 dBZ ETs (× 1000 ft).

VIL and ETs. These data were received in near real time from late 2013 to early 2015. Occasionally, there would be data outages of some or all products from a certain dataset, and we did not have the option of retrieving data missed during these time periods. Because of the substantial disk space required for the CIWS and MRMS forecast datasets for an entire year's worth of data (over 5 TB), we subsampled the forecast data to only those generation times on the hour and half hour, and only kept forecast output at 30 minute intervals. Figure 5 shows that most days had nearly 100% CIWS availability for both observations and forecasts, while the overall availability including data outage periods was ~97%. MRMS observations, on the other hand, generally had daily availability between 75-100%, until a significant update was instituted in June 2014, after which the MRMS availability was near 100% on most days when the feed was not down. Because of the more frequent MRMS data outages, the availability of each of its products was between 81 and 85%.

In addition to the occasional periods of MRMS data outages, data was also sometimes missing from a significant portion of radars (Grim et al. 2015); this happened most often during business hours on weekdays, indicating that the bandwidth at the FAA WJHTC was likely unable to handle the

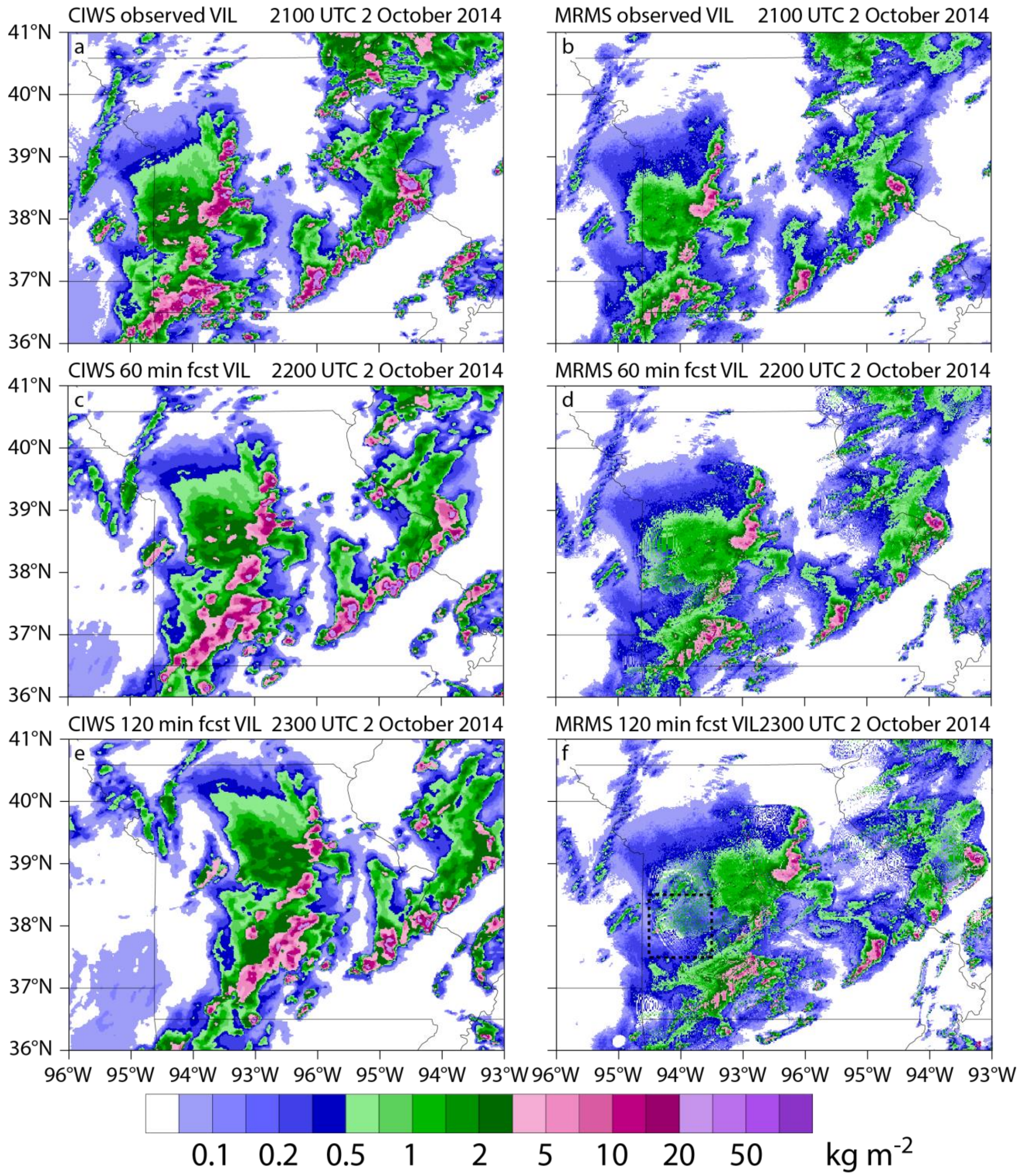

Figure 3. (left) CIWS and (right) MRMS observed and forecast VIL (kg m-2) over Missouri, with initial time at 2100 UTC 2 October 2014. The black dashed box in panel f indicates the zoomed in area shown in Fig. 7.

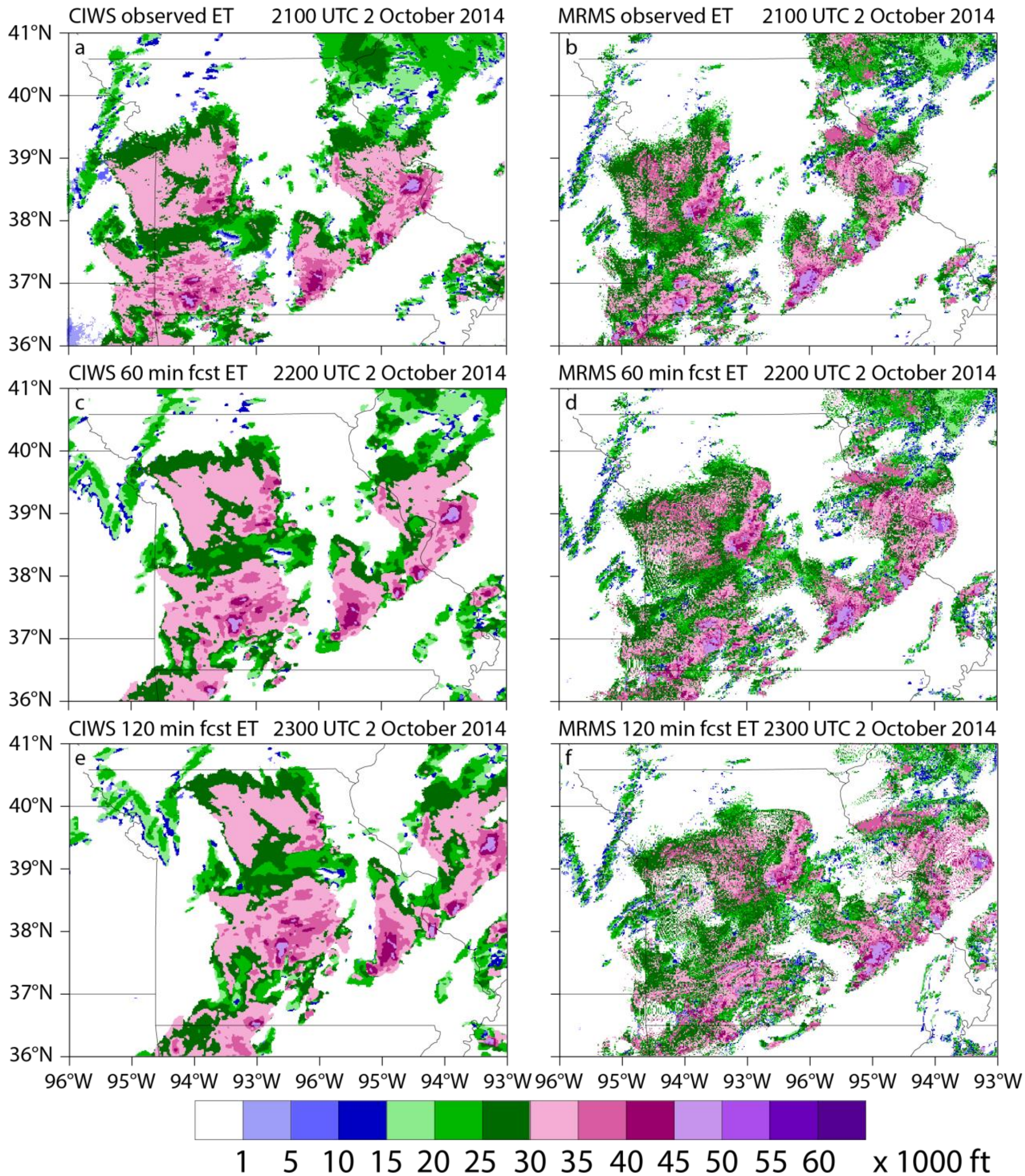

Figure 4. (left) CIWS and (right) MRMS observed and forecast ETs (×1000 feet) over Missouri, with initial time at 2100 UTC 2 October 2014.

data ingest at these times. In addition, MRMS data was sometimes missing from one or more of the four binary files (Grim et al. 2015), which are combined to create the MRMS netCDF files that we received.

The primary purpose of this study has been to compare the CIWS and MRMS data in the form in which we received them, as well as to identify strengths and weaknesses of each dataset. Although it wasn't our intent to QC the data, since we do not have the raw or intermediate data used

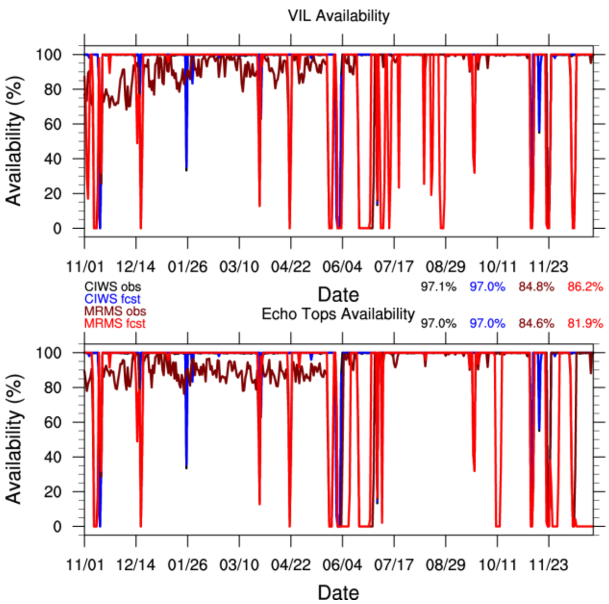

Figure 5. Availability of the CIWS and MRMS observed and forecast datasets for (top) VIL and (bottom) ETs, from 1 November 2013 – 31 December 2014. The colored numbers between the panels indicate the availability of each dataset throughout the period of study.

to create the final products, we did perform some QC wherever possible, to enable better comparisons of the two datasets. One of the primary issues observed with the MRMS dataset, and to a lesser extent from the CIWS dataset, were rays of bad data extending outward from radar sources, such as caused by microwave radiation from the sun near sunrise and sunset. These rays typically have low VIL values, but very high echo top values (Fig. 6). Since many of the erroneous MRMS ETs had values of 18 km (59,055 ft, the maximum possible for MRMS), all MRMS ET values of exactly 18 km were removed from subsequent analysis. We also utilized the QC fields that accompanied each dataset: the MRMS RQI (Zhang et al. 2012, Grim et al. 2015), and the CIWS VIL and ET FLAG arrays, though we found that eliminating data with low QC values had a statistically insignificant effect on final results (Grim et al. 2015).

One important item that should be noted is that many of the MRMS QC issues we identified in the observational data from our data feed from the FAA WJHTC were not seen in the version of MRMS graphics on the NSSL website.* Therefore, it is likely that the FAA MRMS data we received

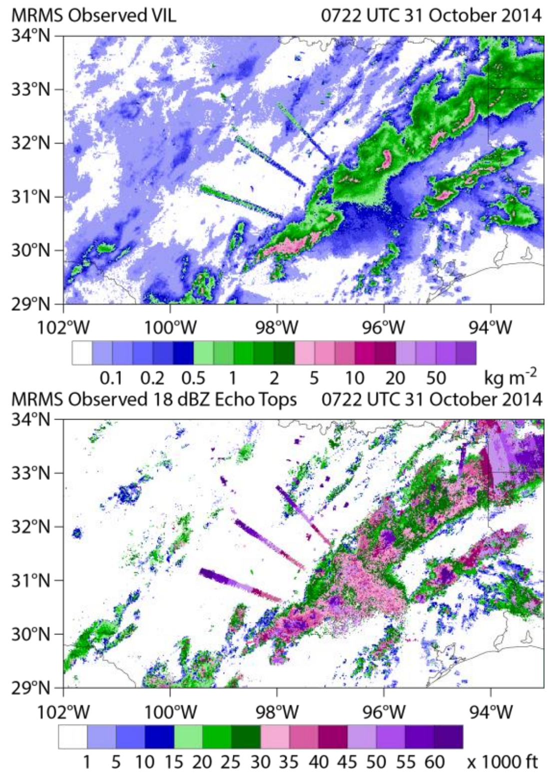

Figure 6. (top) MRMS VIL (kg m-2) and (bottom) MRMS ETs (ft) on 31 October 2013 at 0722 UTC, showing example of the “rays” of erroneous VIL and ETs.

did not always pass through the same QC tests, and suffer from the same data outages, as the version from NSSL. In addition, visual inspections of the MRMS forecast VIL and ET fields revealed some interesting artifacts, likely the result of their advection scheme. If one looks very closely at the MRMS forecast VIL and ET plots in Figs. 3 and 4, especially at the 120-minute forecast time, a “wave” pattern of concentric circles can be seen throughout the images. This is best shown in Fig. 7, zoomed in on an area identified in Fig. 3 for the 120-minute forecast. These advection artifacts were not seen in visual inspections of the CIWS forecast products.

4. Comparisons of VIL Analyses

One way of comparing the observed CIWS and MRMS products was to do a point-by-point comparison of the two at the exact same time. Pairs of CIWS and MRMS VIL values were selected, if the grid points in each product’s native

* http://nmq.ou.edu

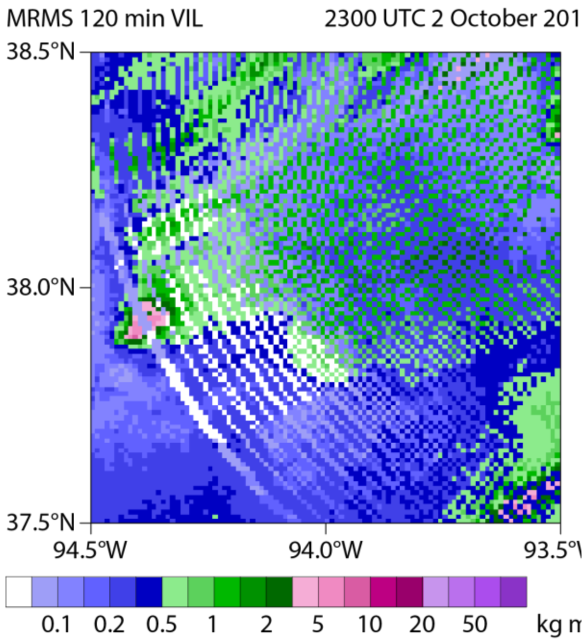

Figure 7. MRMS 120-min forecast VIL for the area outlined in a black box in Fig. 3f.

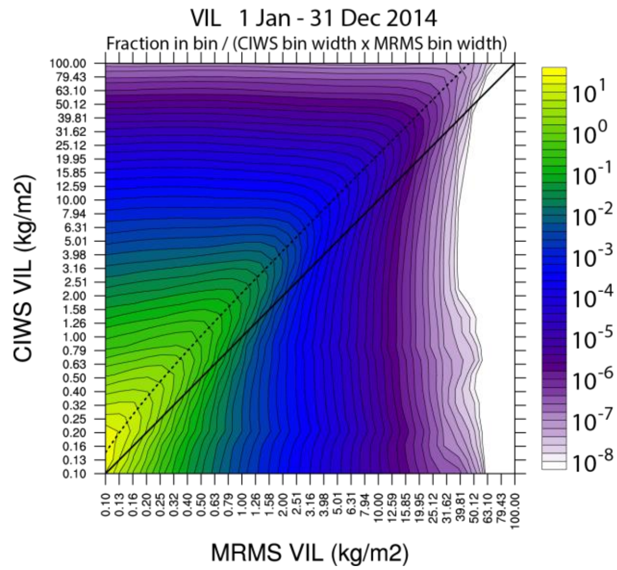

Figure 8. Two-dimensional histogram of point-by-point comparisons between CIWS and MRMS VIL. Bin limits are indicated along the x and y axes, and increase exponentially. The magnitude in each bin is the fraction in that bin, divided by the respective CIWS and MRMS bin widths. The solid black line is the 1:1 line, while the dashed line indicates the linear best fit line.

projection were within 75 seconds (half of the CIWS time resolution) and 0.5 km (half of the CIWS spatial resolution) of each other. In order to best visualize the results from this comparison, a two-dimensional (2D) histogram plot was created (Fig. 8), where the normalized counts within each bin are color shaded. Since low VIL values are exceptionally more frequent than high VIL values, the bin limits were calculated in an exponential fashion, according to the equation:

$$bin = 0.1 \times 10^{(0.1 \times i)}, \qquad (2)$$

where *bin* is the bin limit, and *i* is the bin number. To provide a value that is unbiased by the bin width, the count in each bin was divided by its respective CIWS and MRMS bin widths. In addition, an exponential contouring scale was used for the plots to span the very wide range of values. The comparison of all CIWS and MRMS VIL pairs reveals a wide spread in values, though the greatest frequency lies along the linear best fit line for the data. This best fit line indicates that CIWS VIL values are roughly double those of their MRMS counterparts, as the line had a slope of 2.17 and a y-intercept of -0.07; the negative y-intercept indicates that the ratio of CIWS to MRMS VIL increased slightly, with increasing VIL.

When comparing CIWS and MRMS VIL pairs, separated by season, a very similar pattern is evident amongst each season (Fig. 9). The most notable difference is that there are slightly higher frequencies for low VIL values (lower left side of plots) in the winter and slightly lower frequencies in the summer, while the opposite is true for higher VIL values (upper right side of plots). In order to better determine differences between seasons, Fig. 10a shows the slope and y-intercept values for each. (Note that slope and y-intercept values are dependent on each other, which is why they nearly mirror each other.) The range of slope values was from 2.09 to 2.25, with lowest slope values during the spring and summer and the highest in fall and winter. CIWS and MRMS VIL pairs were also subdivided into regions (Fig. 11) and time of day. The magnitude of variations (1.96 – 2.22) in regional differences was comparable to the seasonal differences, with the West region having the lowest slope value (Fig. 10b). When separating the VIL comparisons by time of day (Fig. 10c), a much smaller range of slope values was seen (2.14 – 2.18) than for the seasonal and regional subsets.

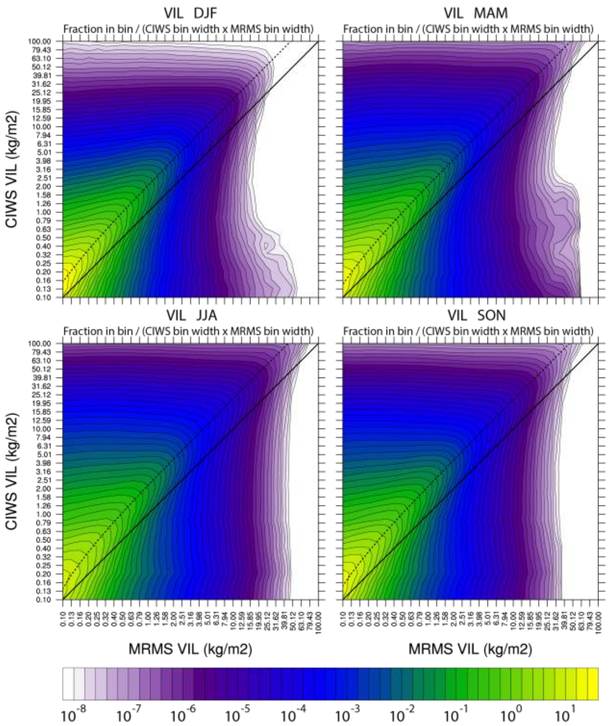

Figure 9. Same as Fig. 8, except divided into seasons.

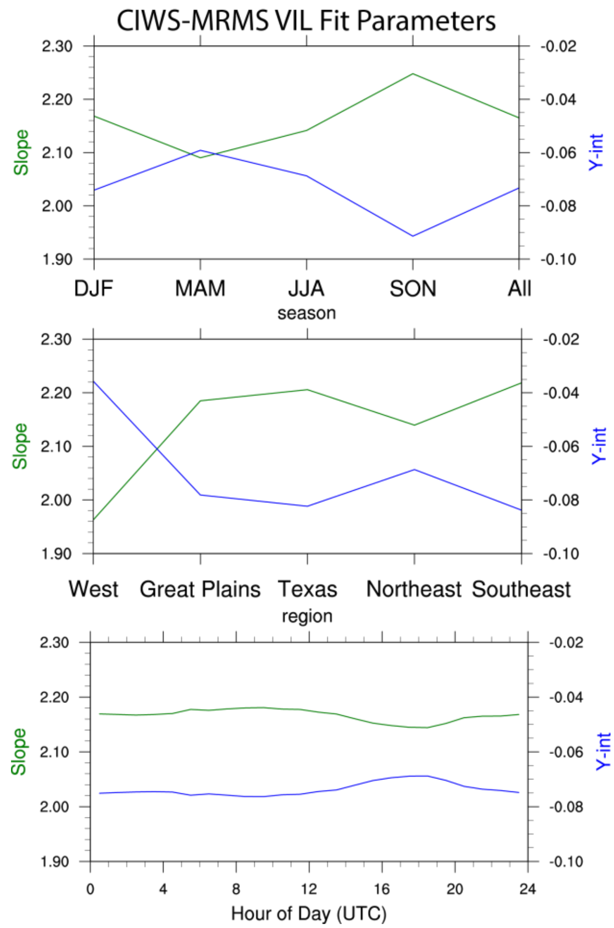

Figure 10. Slope and y-intercept values for linear best fit lines for VIL histograms, differentiated by (top) season, (middle) region, and (bottom) time of day.

5. Comparison of ET Analyses

Similar point-by-point comparisons were done for ETs, except that here the bin widths were kept the same for all heights (Fig. 12). However, the CIWS bin widths are slightly smaller than the MRMS bin widths due to the differing vertical resolutions of each dataset. The CIWS ETs are stored as integer multiples of 1000 ft (304.8 m), while the MRMS ETs are derived from a grid with varying vertical resolution, ranging from 0.25 at lower altitudes to 1 km at higher altitudes; therefore, a bin width of 3000 ft (914.4 m) is used for the CIWS (y) axis, while a similar bin width of 1 km (3280.8 ft) is used for the MRMS data. In addition, the minimum ET bin limit for both datasets was set at 10,000 ft (3048 m).

The 2D ETs histogram (Fig. 12) shows a more complicated relationship than the VIL 2D histogram (Fig. 8). The most notable item in the plot is that the major axis (indicated by the dashed black line on Fig. 12) shows that the average CIWS ETs are generally 3000 ft higher than their MRMS counterparts at low ETs, nearly identical around 35,000 ft (10,668 m), and lower above that level. It is interesting to note that the difference in CIWS vs. MRMS ETs changes most rapidly with height within the critical aviation cruising altitudes of 30,000 – 40,000 ft (9,144 – 12,192 m); therefore further research is needed to determine which dataset is more accurate at these levels. Another notable feature of the ET 2D histogram is that the spread is much wider than the VIL 2D histogram (Fig. 8), particularly on the lower-right-hand side of the plot. A visual inspection of a couple dozen individual CIWS and MRMS ET map plots for points with low CIWS ETs and very high MRMS ETs (lower-right-hand side of Fig. 12) indicated that nearly all of these pairs are from erroneous MRMS ETs. Likewise, visual inspection of around a dozen plots with very high CIWS ETs and low MRMS ETs (upper-left hand side of Fig. 12) indicated these pairs were mostly from erroneous CIWS data.

When separating the ET histograms by season, the most obvious difference is that in winter, there is a much lower fraction of high ETs, and a lot more low ETs, with the opposite being true in the summer (Fig. 13). In addition, the

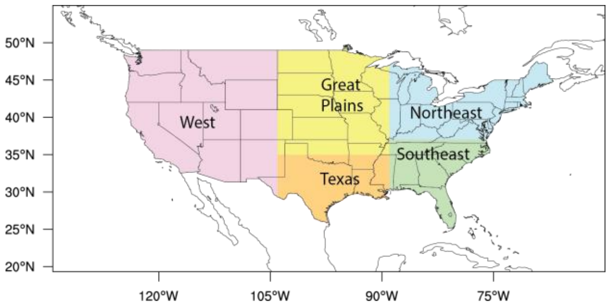

Figure 11. Map of analysis regions.

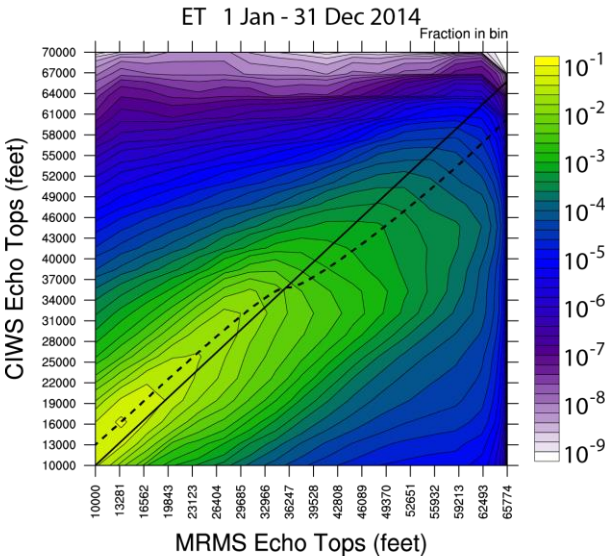

Figure 12. Two-dimensional histogram of point-by-point comparisons between CIWS and MRMS ETs. Bin limits are indicated along the x and y axes, and reflect the slightly different vertical resolution of the two datasets. The magnitude in each bin is the fraction in that bin. The solid black line is the 1:1 line, while the dashed black line indicates the major axis.

elevation where the major axis (dashed gray line in Fig. 13) crosses over the 1:1 line varies from ~32,000 ft in winter to ~36,000 ft in summer. This indicates that the proclivity of their relationship is not only related to altitude, but to season as well. Dividing the CIWS-MRMS ET pairs by region and time of day reveals more similarities amongst their 2D histograms than between seasons, with variations in the regional cross-over point of ~2,000 ft (610 m), and diurnal differences of less than 1000 ft (305 m, not shown).

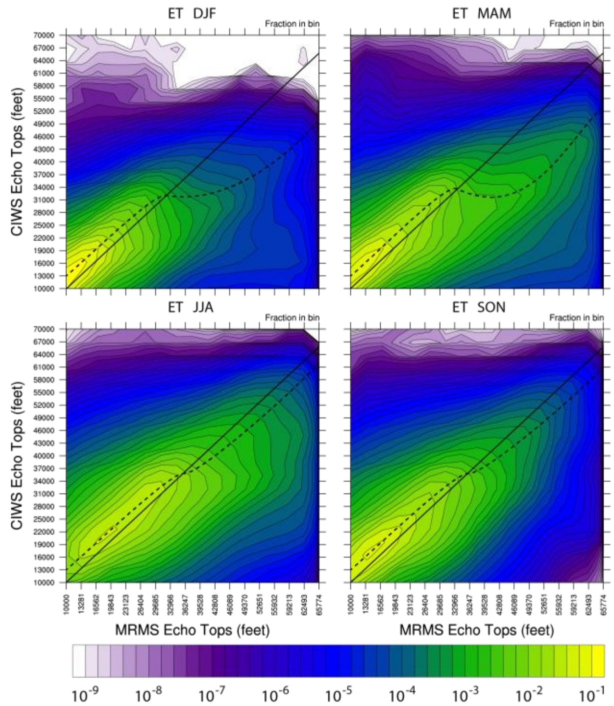

Figure 13. Same as Fig. 12, except divided into seasons.

6. Assessment of VIL Forecasts

In order to impartially assess the skill of the CIWS and MRMS forecast products, each forecast is compared to its respective observational analysis, valid at the forecast time. To assess the skill, each observed and forecast pixel was assigned a logical value (true or false), depending on whether its value exceeded a certain threshold (e.g., 3.5 kg m^{-2} for VIL). Then, each forecast pixel was assigned an attribute of “hit”, “miss”, “false alarm” (FA) or “null”, depending on the comparison of the observed and forecast logical values. A “hit” is defined where a true value is present for both the forecast and observed fields; a “miss” is defined where the observed value is true, but the forecast value is false; a “FA” is defined where the observed value is false, yet the forecast value is true; and a “null” attribute is defined where both the observed and forecast values are false. Finally, two skill scores are calculated using counts of the hits (N_{hit}), misses (N_{miss}), FAs (N_{FA}) and nulls (N_{null}) for all pixels and all times in the dataset:

$$POD = \frac{N_{hit}}{N_{hit}+N_{miss}} \text{ and} \tag{3}$$

$$Bias = \frac{N_{hit}+N_{FA}}{N_{hit}+N_{miss}}. \tag{4}$$

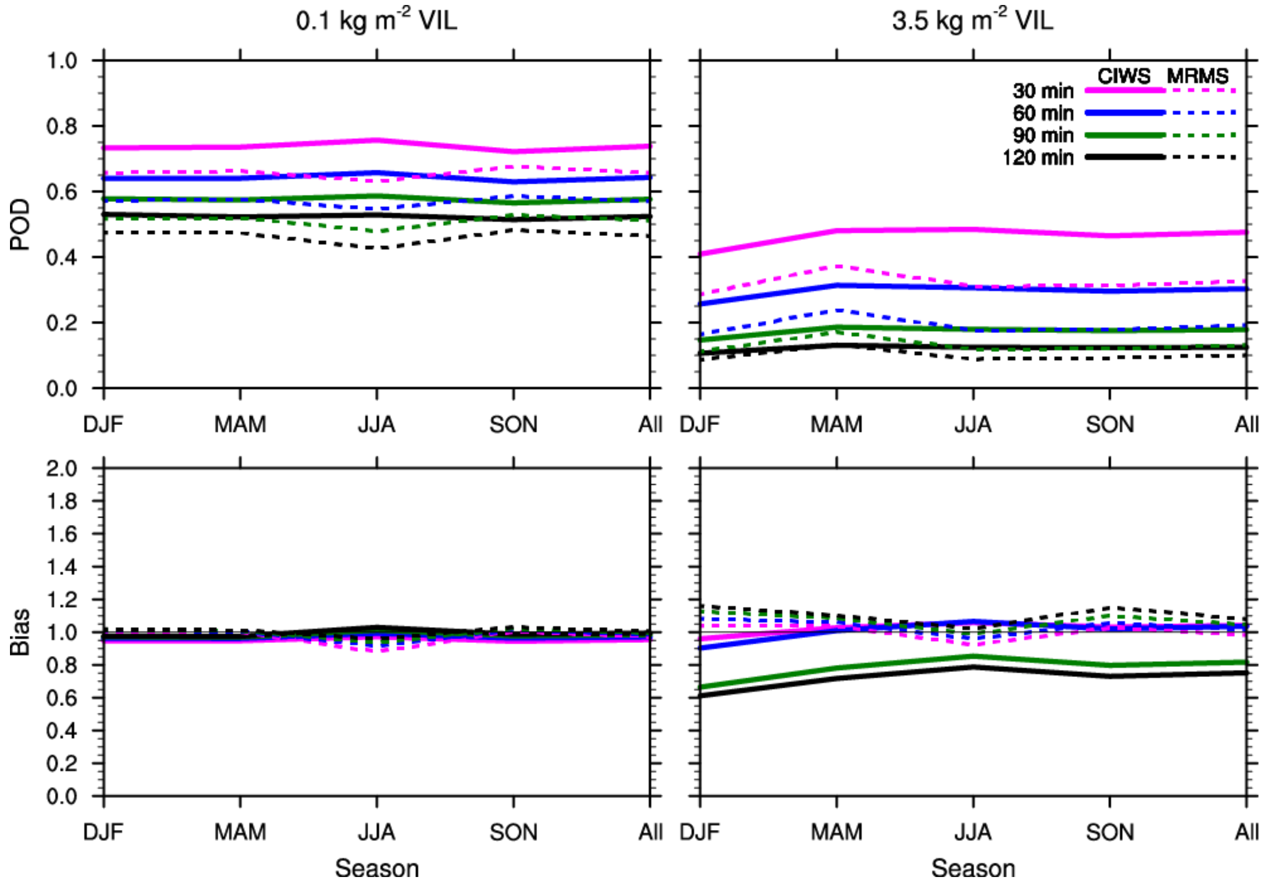

Figure 14. CIWS and MRMS forecast VIL scores for each season, compared to each dataset's observations, using a threshold of (left) 0.1 kg m^{-2} and (right) 3.5 kg m^{-2}. Statistics plotted are (top) probability of detection (POD) and (bottom) bias. The key to the line colors and dash patterns is in a corner of the upper-right panel.

Two VIL exceedance thresholds were used to define the precipitation area: 0.1 kg m^{-2} and 3.5 kg m^{-2}. The lower threshold was used to assess the forecast performance for nearly all precipitation, while the higher threshold was used to estimate the performance for largely convective-type areas of heavier precipitation. Since CIWS VIL values are generally double that of their MRMS counterparts, the MRMS skill scores shown here are for the equivalent thresholds of 0.05 kg m^{-2} and 1.66 kg m^{-2}, respectively, using the linear best fit line equation shown in Fig. 8.

When looking at the probability of detection (POD) plots (Fig. 14), as would be expected, the forecast skill degrades with increasing lead time for both datasets, as well as with the higher threshold level (3.5 vs. 1.0 kg m^{-2}). The POD scores are nearly equal amongst seasons, as the differences between the highest and lowest scores for any season are less than 0.06. In addition, the CIWS forecasts showed greater skill in each season than their MRMS counterparts, using both calibrated MRMS VIL thresholds (Fig. 14) and uncalibrated MRMS VIL thresholds (not shown). Using the calibrated thresholds increased the MRMS forecast POD slightly, but they still remained well lower (-0.01 to -0.15) than those of CIWS. Concerning the frequency bias (hereafter referred to as "bias") scores, both datasets had values near one for all seasons at the 0.1 kg m^{-2} threshold, meaning that the number of FAs and missed forecasts were nearly equal. For the 3.5 kg m^{-2} threshold, CIWS bias scores were generally less than one, especially at longer lead times, indicating that misses outnumbered FAs. On the other hand, the MRMS bias scores were generally a little larger than one, indicating a few more FAs than misses.

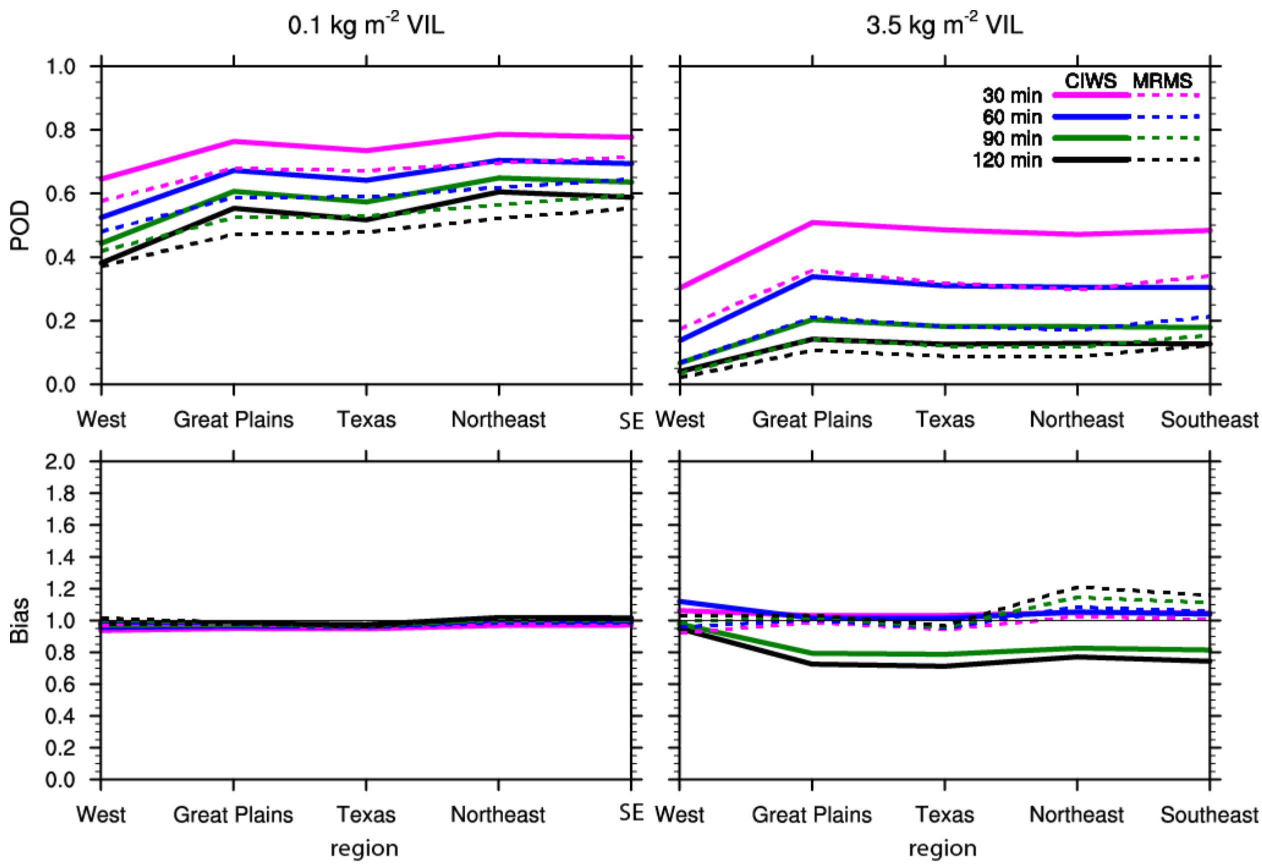

Figure 15. Same as Fig. 14, except for each region shown in Fig. 11.

When considering the performance of the CIWS and MRMS VIL forecasts by region, the POD lines shown in Fig. 15 are nearly parallel to each other, indicating that both products have proportionally varying levels of skill for each region. However, the lowest POD scores were for the West region, likely the result of incomplete and compromised radar coverage, as well as the more complicated nature of precipitation over mountainous terrain. As with the seasonal bias scores (Fig. 14), the regional bias scores were again very near one for the 0.1 kg m^{-2} threshold (Fig. 15); likewise, at the 3.5 kg m^{-2} threshold, the CIWS bias scores were generally less than one at longer lead time (misses > FAs), while the MRMS bias scores were generally equal to or slightly larger than one (FAs ≥ misses).

The POD scores showed a distinct diurnal variation for both CIWS and MRMS (Fig. 16), with generally equal scores at each hour of the day, except for a dip centered on about 1830 UTC (early afternoon) for CIWS and about an hour later (1930 UTC) for MRMS. As with the regional analyses (Fig. 15), the respective CIWS and MRMS POD lines are nearly parallel to each other. The bias scores for the 0.1 kg m^{-2} threshold are very near one throughout the day, though there is a dip in the early afternoon, with the MRMS bias minima about an hour later than the CIWS minima. At the 3.5 kg m^{-2} threshold, the diurnal variation in bias scores are more amplified than at 0.1 kg m^{-2}, with relative peaks at 0000 – 0100 UTC and ~1300 UTC, and minima near 0700 UTC, and especially at 1700 – 1900 UTC. Once again, the times of the MRMS minima trail the CIWS minima times. The high MRMS biases at the times of the peaks (as high as 1.5 at the 120-minute forecast time) are the result of there being a lot more FAs than misses. Conversely, the CIWS bias scores for the 3.5 kg m^{-2} threshold, which are mainly less than or equal to one, have minimum values near 1700 UTC, with the lowest bias scores for the 90- and 120-minute forecasts, when misses far outnumber FAs.

7. Assessment of ET Forecasts

Considering the skill of the CIWS and MRMS ET forecasts, divided by season, the CIWS POD

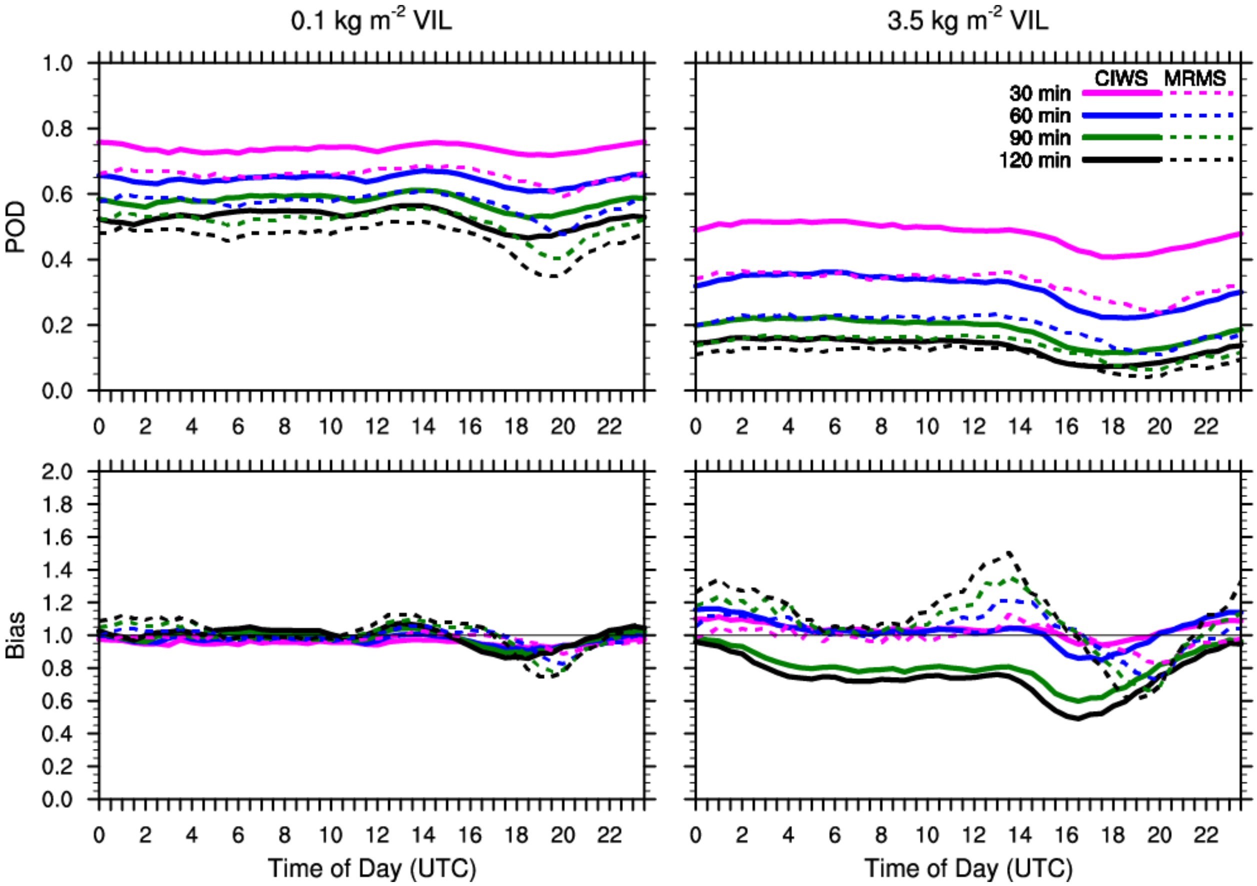

Figure 16. Same as Fig. 14, except for each hour of the day.

scores are all higher than their MRMS counterparts, for both the 10,000 ft (3048 m) and 30,000 ft (9144 m) thresholds (Fig. 17). For both datasets, the 10,000 ft threshold POD scores are slightly higher than their respective 30,000 ft counterparts. The difference between the 10,000 ft and 30,000 ft scores are dependent on season. CIWS summer POD scores are only 0.02 to 0.15 lower for 30,000 ft than for 10,000 ft, while the winter scores are 0.16 to 0.23 lower. The seasonal variation in POD scores between threshold altitudes are even more marked for MRMS, with summer scores 0.17 to 0.19 lower for 30,000 ft than 10,000 ft, and winter scores 0.25 to 0.28 lower. The bias scores are very near one (misses ≈ FAs) for both datasets, at all forecast times and all seasons, with the MRMS bias scores being generally slightly higher than their CIWS counterparts.

Dividing the ET forecast performance by region reveals that CIWS POD scores are 0.02 to 0.29 higher than their MRMS counterparts (Fig. 18). The highest scores are in the Southeast for both datasets at the 10,000 ft threshold, and over the Great Plains at the 30,000 ft threshold. The lowest scores are for the West region, just as they are for the VIL forecast scores. As for the bias scores, both datasets have values near one for the 10,000 ft and 30,000 ft thresholds, with the exception of the West in the CIWS product, where values were ~0.85. As for the 30,000 ft threshold, MRMS bias values were near one at earlier forecast times, to as high as 1.27 at later forecast times, with the highest bias values in the Northeast and Southeast. The CIWS bias scores at 30,000 ft were very near one at all lead times.

The diurnal variations in ET forecast scores are shown in Fig. 19. Just as has been noted for the VIL forecast scores (Fig. 16), the CIWS and MRMS POD lines mostly parallel each other throughout the day, with CIWS scores ~0.2 higher than their MRMS counterparts. A small dip in their

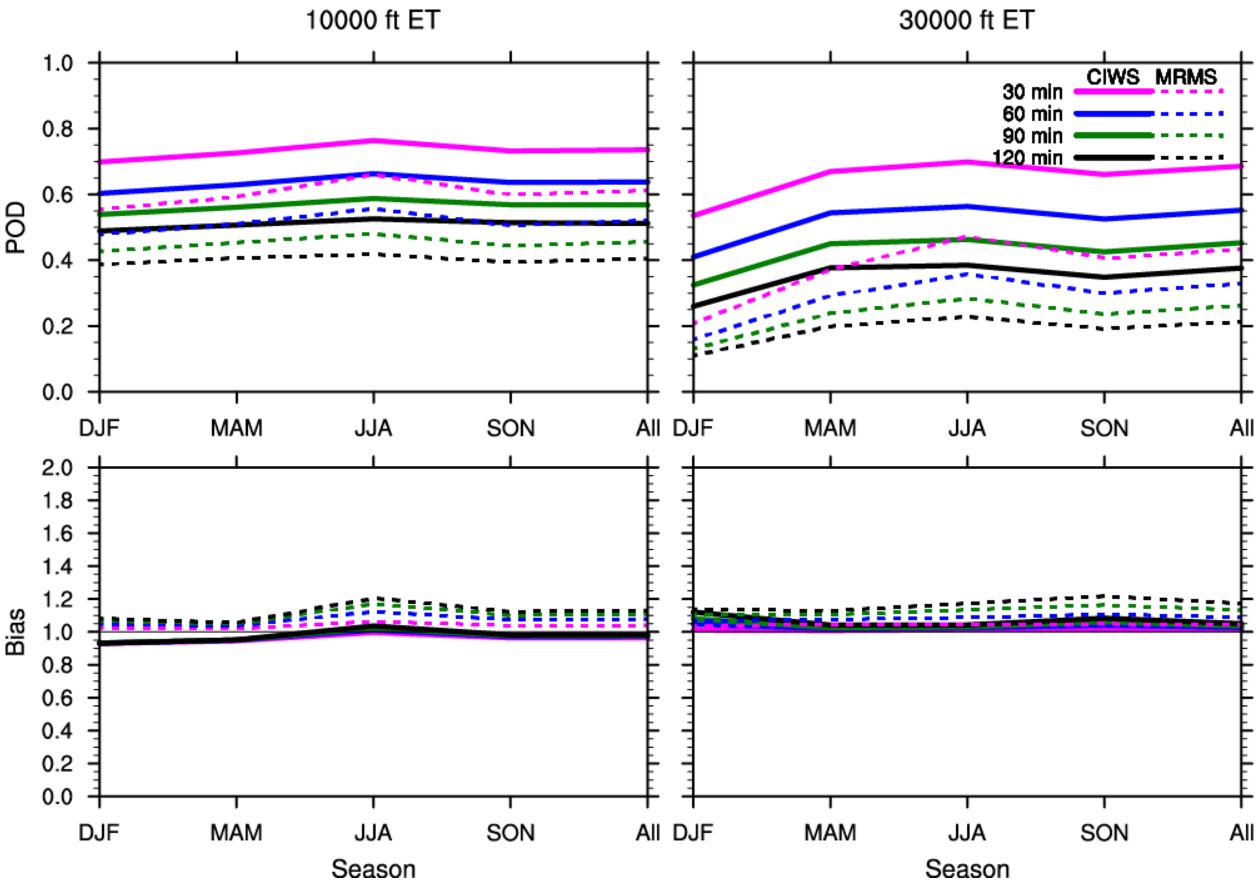

Figure 17. CIWS and MRMS forecast ETs scores for each season, compared to each dataset's observations, using a threshold of (left) 10,000 ft (3048 m) and (right) 30,000 ft (9144 m). Statistics plotted are (top) probability of detection and (bottom) bias. The key to the line colors and dash patterns is in a corner of the upper-right panel.

scores is observed near 1800-1900 UTC for CIWS, and 1 – 2 hours later for MRMS (Fig. 19). The most notable diurnal variations occur in the ET bias scores; at 10,000 ft, CIWS bias scores are near one, with only a very weak diurnal amplitude (~0.05) for CIWS, and mostly greater than one values for MRMS at nearly all hours, with a peak around 1400 UTC and a minimum around 1900 UTC. At 30,000 ft, the diurnal patterns observed in the bias scores are very similar, except that they are amplified. The MRMS bias scores are as high as 1.88 at 1330 UTC for the 120-minute forecast, with a minimum of 0.52 for the 120-minute forecast at 1930 UTC. The amplitude of the diurnal CIWS bias cycle for the 30,000 ft threshold is slightly smaller than that of MRMS, with peak values as high as 1.30 at 0100 and 1300 UTC, and as small as 0.57 at 1800 UTC. The times of the CIWS minimum and maximum bias values are about an hour or two earlier than those for MRMS.

9. Conclusions

This comparative study assessed many of the differences between the Corridor Integrated Weather System (CIWS) and Multiple-Radar / Multiple-Sensor (MRMS) vertically-integrated liquid water (VIL) and 18 dBZ echo tops (ETs) data sets: how each observed field is created, and the differences between the respective CIWS and MRMS products. The CIWS products are from the Massachusetts Institute of Technology – Lincoln Laboratory (MIT–LL), while the MRMS data came from the FAA William J. Hughes Technical Center (WJHTC). Both datasets are continually improving, so this report should be considered an

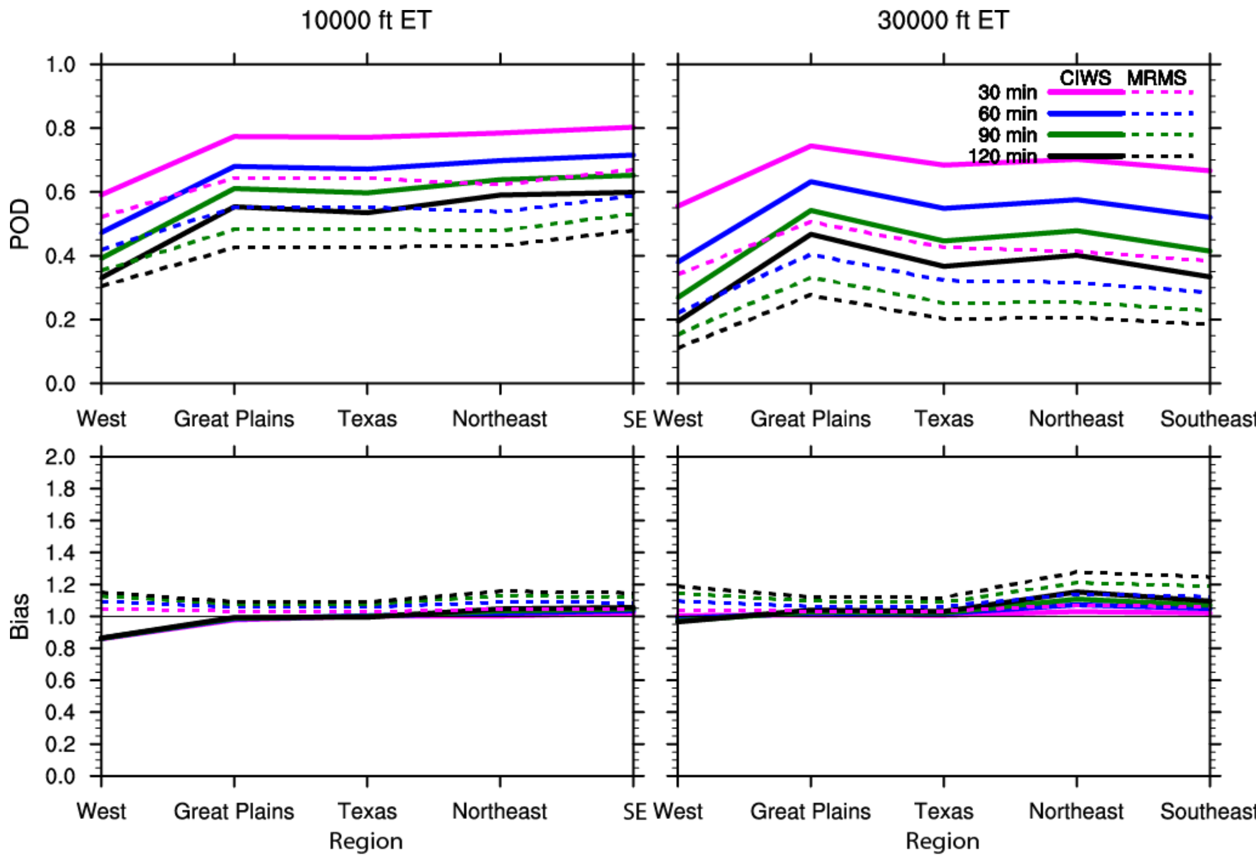

Figure 18. Same as Fig. 17, except for each region shown in Fig. 11.

assessment of the two during the time period of this project: fall 2013 – winter 2015.

Direct comparisons of contemporaneous CIWS and MRMS VIL products showed that CIWS VIL was on average 2.17 times as large as MRMS VIL (as defined by the slope of the best fit line to the data), though this varied seasonally, regionally and diurnally by up to ±0.16, 0.26 and 0.04, respectively. Direct comparisons of CIWS and MRMS ETs indicated that their relationship varied with height: at low altitudes, CIWS ETs were generally ~3000 ft (914 m) higher than MRMS ETs; this difference decreased with height so that they were nearly equal at a “crossover point” of ~35,000 ft (10668 m), while MRMS ETs were typically higher than CIWS ETs above this level. The difference in ETs changes most rapidly with height within the critical aviation cruising altitudes of 30,000 – 40,000 ft (9,144 – 12,192 m), so it will be important to determine which dataset is more accurate at these levels. The crossover point varied most by season (~4,000 ft = 1524 m), to a lesser extent by region (~2,000 ft = 610 m) and less than 1000 ft (305 m) by time of day.

For analysis of the forecast products, each forecast was compared to its own analysis at the valid time (e.g., CIWS forecast VIL compared to CIWS observed VIL). The CIWS forecast VIL and ET POD skill scores were always 0.01 to 0.15 higher than those from MRMS when compiled by lead time, VIL/ET threshold, season, region and time of day. The bias scores revealed that each dataset had different biases, depending on region, season and time of day. VIL bias scores were near 1 for both CIWS and MRMS at the 0.1 kg m^{-2} threshold, meaning false alarms (FAs) were nearly equal in number to misses. At the 3.5 kg m^{-2} threshold, bias scores were generally less than or equal to 1 for CIWS, indicating that misses were more frequent than FAs; while the opposite was true of MRMS bias scores which were generally greater than 1. For the diurnal cycle, MRMS bias scores peaked near 0100 and as high as 1.5 at the 120-minute forecast time; this indicates that there were many more FAs than misses at this

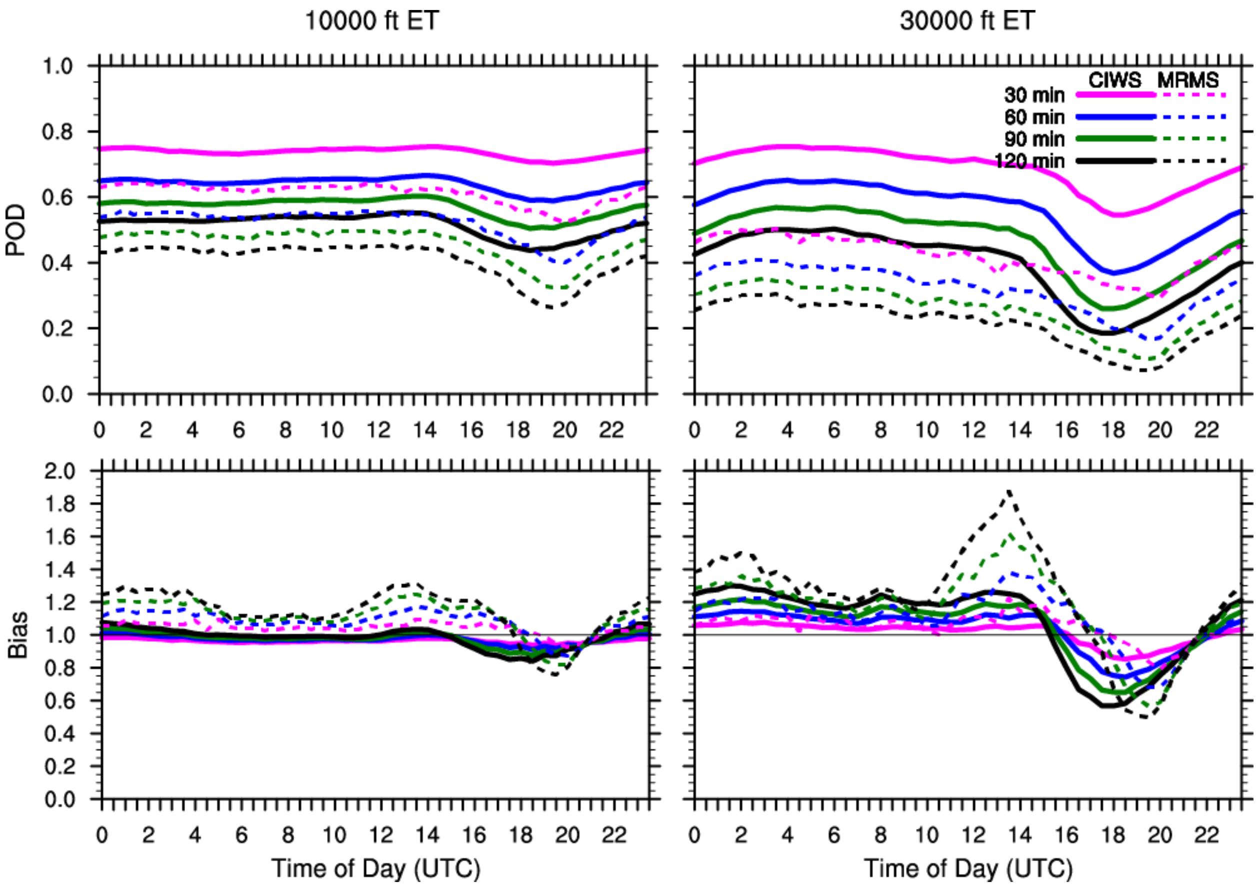

Figure 19. Same as Fig. 17, except for each hour of the day.

time of day. Concerning ET POD scores, CIWS scores are all higher than their MRMS counterparts by at least 0.02, for both the 10,000 ft (3048 m) and 30,000 ft (9144 m) thresholds and all seasons and regions. There was a weak diurnal cycle for CIWS and MRMS POD scores, with the minima in the MRMS cycle being one to two hours later than for CIWS. As for ET bias scores, CIWS and MRMS biases were near 1 for the 10,000 ft (3048 m) threshold for all seasons and regions, indicating neither dataset had a significant bias toward misses or FAs at this threshold. However, there was a pronounced diurnal cycle for both CIWS and MRMS bias scores, which was larger for the 30,000 ft threshold than the 10,000 ft threshold. The range of CIWS bias scores at 30,000 ft was 0.57 to 1.30, and slightly wider for MRMS at 0.52 to 1.88. Since the diurnal cycle shapes are similar between CIWS and MRMS, it is likely something meteorological that is causing these cycles (as opposed to a dataset bias.)

Instances of missing and insufficiently QCed data were more common in the MRMS data than for CIWS, although many of the MRMS issues observed from our feed from WJHTC were not visually observed in the concurrent MRMS products produced at the National Severe Storms Laboratory (NSSL). In addition, there were improvements to the data quality throughout the period of study. The most common form of persisting QC issues were: 1) occasional missing data from some radars in MRMS products, and 2) wavelike patterns in forecast MRMS products. Further investigation using the NSSL version of MRMS that has fewer QC issues is crucial, to determine its relative performance compared to CIWS.

Acknowledgements

This research is in response to requirements and funding by the National Weather Service (NWS), and access to CIWS data was provided via agreement with 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 NWS or FAA.

References

Billet, J., M. DeLisi, B. G. Smith, and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting,* **12,** 154–164.

DeLaura, R., and J. Evans, 2006: *An exploratory study of modeling enroute pilot convective storm flight deviation behavior*. Project Report NASA/A-6, MIT Lincoln Laboratory, Lexington, MA.

Donaldson, R., 1964: A demonstration of antenna beam errors in radar reflectivity patterns. *J. Appl. Meteor.,* **3,** 611–623.

Evans, J. E., and E. R. Ducot, 2006: Corridor Integrated Weather System. *Lincoln Lab. J.,* **16,** 59–80.

Evans, J., K. Caruson, M. Wolfson, M. Robinson, E. Ducot, and B. Crowe, 2004: Improving convective weather operations in highly congested airspace with the Corridor Integrated Weather System (CIWS). Preprints, *11th Conf. on Aviation, Range and Aerospace Meteorology,* Hyannis, MA, Amer. Meteor. Soc., P1.5.

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water – A new analysis tool. *Mon. Wea. Rev.,* **100,** 548–552.

Grim, J. A., M. Steiner, J. Pinto, K. Stone, and D. Megenhardt, 2015: Assessment and comparison of the CIWS and MRMS VIL and echo top observational and forecast products. Technical Report to the NWS.

JPDO, 2007: *Concept of operations for the next generation air transportation system*. Version 1.2, Joint Planning and Development Office, Washington, DC, 226 pp.

Kitzmiller, D. H., W. E. McGovern, and R. F. Saffle, 1995: The WSR-88D severe weather potential algorithm. *Wea. Forecasting,* **10,** 141–159.

Klingle-Wilson, D., and J. Evans, 2005: *Description of the Corridor Integrated Weather System (CIWS) weather products*. Project Report ATC-317, MIT Lincoln Laboratory, Lexington, MA.

Lakshmanan, V., T. Smith, K. Hondl, G. Stumpf, and A. Witt, 2006: A real-time, three-dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity, and derived products. *Wea. Forecasting,* **21,** 802–823.

Marshall, J. S., and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.,* **5,** 165–166.

Rappa, G., and S. Troxel, 2009: *CIWS Product Description Revision 1.0*. Project Report ATC-355, MIT Lincoln Laboratory, Lexington, MA.

Robinson, M., J. E. Evans, and B. A. Crowe, 2002: En route weather depiction benefits of the NEXRAD vertically integrated liquid water product utilized by the Corridor Integrated Weather System. *10th Conf. on Aviation, Range and Aerospace Meteorology (ARAM),* Portland, OR, Amer. Meteor. Soc.

Shafer, M. A., D. R. MacGorman, and F. H. Carr, 2000: Cloud-to-ground lightning throughout the lifetime of a severe storm system in Oklahoma. *Mon. Wea. Rev.,* **128,** 1798–1816.

Stern, A. D., R. H. Brady III, P. D. Moore, and G. M. Carter, 1994: Identification of aviation weather hazards based on integration of radar and lightning data. *Bull. Amer. Meteor. Soc.,* **75,** 2269–2280.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. Van Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D.-J. Seo, E. Wells, and C. Dempsey, 2011: National mosaic and multi-sensor QPE (NMQ) system. *Bull. Amer. Meteor. Soc.,* **92,** 1321–1338.

Zhang, J., Y. Qi, C. Langston, and B. Kaney, 2012: Radar quality index (RQI) = a combined measure for beam blockage and VPR effects in a national network. *Weather Radar and Hydrology Symposium,* Exeter, UK.