GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2018 Spring Experiment Final Evaluation

Project Title: GOES-R and JPSS Proving Ground Demonstration at the 2018 Spring Experiment - Experimental Warning Program (EWP)

Organization: NOAA Hazardous Weather Testbed (HWT)

Evaluator(s): National Weather Service (NWS) Forecasters, Broadcast Meteorologists, Storm Prediction Center (SPC), National Severe Storms Laboratory (NSSL), University of Oklahoma (OU), Cooperative Institute for Mesoscale Meteorological Studies (CIMMS)

Duration of Evaluation: 30 April 2018 – 25 May 2018

Prepared By: Michael Bowlan (OU/CIMMS and NOAA/SPC) and Kristin Calhoun (OU/CIMMS and NSSL)

Submitted Date: 26 September 2018

Table of Contents

1. Executive Summary

This report summarizes the activities and results from the Geostationary Operational Environmental Satellite R-Series (GOES-R) and Joint Polar Satellite System (JPSS) Proving Ground demonstration at the 2018 Spring Experiment, which took place at the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) in Norman, OK from 30 April to 25 May 2018. The Satellite Proving Ground activities were focused in the Experimental Warning Program (EWP). A total of 14 National Weather Service (NWS) forecasters representing four NWS regions and an additional three broadcast meteorologists participated in the EWP experiment. They evaluated seven major baseline, future capability, and experimental GOES-R and JPSS products (Table 1) in the real-time simulated short-term forecast and warning environment of the EWP using the second generation Advanced Weather Interactive Processing System (AWIPS-II).

Some of the products demonstrated in 2018 were involved in previous HWT experiments and have received updates based on participant feedback from the HWT and other demonstrations. GOES-R products demonstrated in the 2018 EWP Spring Experiment included: GOES-16 Advanced Baseline Imager (ABI) Cloud and Moisture Imagery, baseline derived products and numerous multispectral Red Green Blue (RGB) products, the Geostationary Lightning Mapper (GLM) Lightning Detection, the Probability of Severe statistical model (ProbSevere), and the Convective Initiation Probability product. Additionally, GOES-16 provides 1-minute imagery via two 1000-km x 1000-km mesoscale sectors, and its value was also assessed in monitoring convective storm life cycles. As a JPSS Proving Ground activity, the NOAA Unique Combined Atmospheric Processing System (NUCAPS) temperature and moisture profiles were displayed using the AWIPS-II sounding analysis program. These soundings were created using data from three different polar orbiting satellites: the Suomi National Polar-orbiting Partnership (Suomi-NPP) and Europe's MetOp-A and MetOp-B. Additionally, a modified version of NUCAPS was also examined in which an automated correction incorporating surface observations was applied to the boundary layer to improve the accuracy of the sounding. Also, participants were able to view the NUCAPS derived parameters in a plan or cross-section view. Several visiting scientists attended the EWP over the four weeks to provide additional product expertise and interact directly with operational forecasters. Organizations represented by those individuals included: The University of Wisconsin Cooperative Institute for Meteorological Satellite Studies (UW/CIMSS), The University of Oklahoma Cooperative Institute for Mesoscale Meteorological Studies (OU/CIMMS), the National Severe Storms Laboratory (NSSL), the NASA Short-term Prediction Research and Transition Center (SPoRT), The University of Alabama in Huntsville (UAH), Science and Technology Corporation (STC), Lockheed Martin, Vaisala, and NWS. The Storm Prediction Center (SPC) and HWT Satellite Liaison, Michael Bowlan (OU/CIMMS and NOAA/SPC), provided overall project management and subject matter expertise for the Satellite Proving Ground efforts in the HWT with support from Kristin Calhoun (OU/CIMMS and NOAA/NSSL).

Forecaster feedback during the evaluation was collected using several different methods, including daily surveys, weekly surveys, daily debriefs, weekly debriefs, blog posts, informal conversations in the HWT and a weekly "Tales from the Testbed" webinar. Typical feedback included: suggestions for improving the algorithms, ideas for making the displays more effective for information transfer to forecasters, best practices for product use, suggestions for training, and situations in which the tools worked well and not so well. Forecasters' favorite aspect of this year's experiment was being able to evaluate real time GOES-16 imagery and products, and seeing the benefits of certain products or RGBs to their operations. The ProbSevere model continues to provide useful guidance, especially when applied to discrete storms, though improvements to performance are needed for multicellular/linear convective modes, and particularly when wind and tornadoes are the main hazard. Forecasters were also especially interested in evaluating GLM data and getting an early view of the data before others in operational offices. Finally, participants found the NUCAPS information to be helpful in filling spatial and temporal gaps that exist in atmospheric sounding information, and liked that the plan view displays provided a quick look at certain parameters and levels in a NUCAPS swath.

2. Introduction

GOES-R Proving Ground (Goodman et al. 2012) demonstrations in the HWT have provided users with a glimpse into the capabilities, products and algorithms that are and will be available with the GOES-R satellite series, beginning with GOES-16 which launched in November 2016. The education and training received by participants in the HWT fosters interest and excitement for new satellite data and helps to promote readiness for the use of GOES-R data and products. Additional demonstration of JPSS products introduces and familiarizes users with advanced satellite data that are already available. The HWT provides a unique opportunity to enhance research-to-operations and operations-to-research (R2O2R) by enabling product developers to interact directly with operational forecasters, and to observe the satellite-based algorithms being used alongside standard observational and forecast products in a simulated operational forecast and warning environment. This interaction helps the developer to understand how forecasters use the product, and what improvements might increase the product utility in an operational environment. Feedback received from participants in the HWT has proven invaluable to the continued development and refinement of GOES-R and JPSS algorithms. Furthermore, the EWP facilitates the testing of satellite-based products in the AWIPS-II data processing and visualization system currently used at NWS Weather Forecast Offices (WFOs).

In 2018, the GOES-R/JPSS Proving Ground activities were conducted during the weeks of April 30, May 7, May 14, and May 21 with five NWS forecasters participating for the first week and three NWS forecasters and one broadcast meteorologist participating each week for the following three weeks. In an effort to extend the satellite knowledge and participation to the broader meteorological community, and to recognize the critical role played by the private sector in communicating warnings to the public, broadcast meteorologists sponsored by the GOES-R Program participated in the Spring Experiment for the fifth year in a row, working alongside NWS forecasters. Training modules in the form of Articulate Power Point presentations for each demonstrated product were sent to and completed by participants prior to their arrival in Norman. Each week, participants arrived in Norman on Sunday, worked eight hour experimental warning shifts Monday-Thursday and a half-day on Friday before traveling home Friday afternoon.

Much of Monday was a spin-up day that included a one-hour orientation, familiarization with the AWIPS-II system, and one-on-one hands-on training between participants, product developers, and the Satellite Liaison. The shifts on Tuesday, Wednesday and Thursday were "flex shifts", meaning the start time was anywhere between 9 am and 3 pm, depending on when the most active convective weather across the Contiguous United States (CONUS) was expected to occur. The next-day start time was determined the previous evening by the Weekly Coordinator. The Friday half-day involved a weekly debrief and preparation and delivery of the "Tales from the Testbed" webinar.

Shifts typically began a couple of hours before convective initiation was expected to occur as many of the products demonstrated this year have their greatest utility in the pre-convective environment. At the start of each Monday-Thursday experimental warning shift, the Satellite Liaison and forecasters interrogated the large scale weather pattern across the CONUS and determined where to operate for the day. Forecasters, working in pairs, provided experimental short-term forecasts for their assigned County Warning Area (CWA) via a blog. Early in the shift, these were primarily mesoscale forecasts discussing the environment, where convection was expected to occur, and what the applicable demonstration products were showing. Once convection began to mature, one forecaster in the pair would switch to issuing experimental warnings for their CWA while the other forecaster would continue to monitor the mesoscale environment and compose blog posts. Blog posts regarding the use of demonstration products in the warning decision-making process were written during this period along with continued updates on the mesoscale environment. If severe convective activity in a CWA ceased or was no longer expected to occur, the Satellite Liaison would transition the pair of forecasters to focus on a more convectively active CWA.

At the end of each week, forecasters participated in the "Tales from the Testbed" webinar, prepared by the Satellite Liaison, via GoToMeeting. These 22-minute presentations gave participants an opportunity to share their experience in the HWT with an average of greater than 30 remote locations each week, including NWS Headquarters, NWS WFOs and research scientists at satellite cooperative institutes nationwide, providing widespread exposure for the GOES-R and JPSS Proving Ground products. Topics for each of the four webinars were chosen based on the particular week's weather. Sixteen minutes were allowed afterward for questions and comments from viewers on the webinar.

Feedback from participants came in several forms. During the short-term experimental forecast and warning shifts, participants were encouraged to blog their decisions along with any thoughts and feedback they had regarding the products under evaluation. Over 300 GOES-R and JPSS related blog posts were written during the four weeks of the Spring Experiment by forecasters, product developers, and the Satellite Liaison. At the end of each shift (Monday-Thursday), participants filled out a survey for each product under evaluation. The Tuesday-Thursday shifts began with a "daily debrief" during which participants discussed their use of the demonstration products during the previous day's activities. Friday morning, a "weekly debrief" allowed product developers an opportunity to ask the participants any final questions, and for the participants to share their final thoughts and suggestions for product improvement. Additionally on Friday morning, participants completed one last "end-of-the-week" survey. Feedback from the GOES-R and JPSS demonstrations during the 2018 Spring Experiment is summarized in this report.

3. Products Evaluated

Table 1. List of GOES-R and JPSS products demonstrated within the HWT/EWP 2017 Summer Experiment

Demonstrated Product	Category
Advanced Baseline Imager (ABI) imagery, baseline derived products	GOES-R Baseline
RGB Composites and Channel Differences	National Weather Service
ProbSevere Model	GOES-R Risk Reduction
GLM Lightning Detection	GOES-R Baseline
NUCAPS Temperature and Moisture Profiles	JPSS
All-Sky LAP Stability Indices, Total Precipitable Water, and Layered Precipitable Water Products	GOES-R Risk Reduction
Convective Initiation Probability	GOES-R Risk Reduction
Category Definitions: **GOES-R Baseline Products** – GOES-R Level 1 Requirement products that are funded for operational implementation **GOES-R Risk Reduction** – New or enhanced GOES-R applications that explore possibilities for improving Algorithm Working Group (AWG) products. These products may use the individual GOES-R sensors alone, or combine with data from other in-situ and satellite observing systems or NWP models with GOES-R **National Weather Service** – Products created within AWIPS-II **JPSS** – Products funded through the JPSS program	

3.1 Advanced Baseline Imager (ABI) Imagery, Baseline Derived Products

National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), National Environmental Satellite, Data, and Information Service (NESDIS), and GOES-R Program

This was the second year that real time GOES-16 ABI imagery and associated derived products were evaluated in the HWT. The primary focus of evaluation for this year's experiment was on the derived products, particularly those associated with convection and severe weather forecasting. Since the imagery from GOES-16 has been declared operational, that imagery was used heavily, but there was not a lot of emphasis on evaluation. The emphasis for the imagery primarily focused on the use and importance of the one-minute mesoscale sectors versus five-minute data in the forecasters' decision making capabilities. Feedback from this experiment on performance and display of the derived products and usefulness of the one-minute imagery is presented in this section.

Use of ABI imagery in the HWT

This was the second year that forecasters were able to use in-orbit GOES-16 ABI data in the testbed. The GOES-16 ABI imagery was declared operational in the GOES East position earlier this year and has been widely used throughout the NWS for the better part of a year before evaluations took place for the 2018 Spring Experiment in the HWT. Because the data had been declared operational, the primary focus of the evaluation of ABI data was in the usefulness and importance of the one-minute updating imagery and whether it helped a forecaster in his/her severe storm forecasting or warning decision-making process. In this year's experiment there were a number of different types of severe storms and forcing mechanisms that allowed forecasters to examine the usefulness of one-minute imagery in a variety of different situations across the country. At the end of each day, forecasters were asked if the one minute imagery aided them in their warning decision making that day and if so, in what ways did it help. They were also asked if the one-minute imagery provided them with significant information that was not captured in the routine five-minute imagery, and approximately 74% of responses stated that the one-minute imagery did provide them with significant information not captured by the five-minute imagery. Overall, at the end of each week they were asked to rate the impact of the one-minute imagery on their operations, and every forecaster responded that it had a very positive or extremely positive impact. The one-minute imagery was found to be very valuable to convective warning operations. Some responses and examples are shown below.

> "The 1-minute data, particularly the cloud top temperature (CTT) data was used in conjunction with ProbSevere and GLM data to provide confidence in earlier warning issuance than would otherwise be present."
> *Forecaster, End-of-Day Survey*

> "Aided in seeing rapidly intensifying convective updrafts within VIS/IR data."
> *Forecaster, End-of-Day Survey*

> "It helped show me a storm was maintaining intensity despite radar coverage indicating otherwise. It helped prepare me to issue a SVR warning."
> *Forecaster, End-of-Day Survey*

> "The 1-min visible and infrared imagery aided in identifying convective initiation, under the high cirrus cloud layer."
> *Forecaster, End-of-Day Survey*

> "It helped to identity brief overshooting tops as indicators of increases in updraft strength. These were often missed between 5-min scans."
> *Forecaster, End-of-Day Survey*

> "Evolution of the overshooting tops was more fluid in the 1-min imagery vs 5-min. Confidence in convective initiation was had earlier when viewing the 1-min imagery versus the 5-min imagery."
> *Forecaster, End-of-Day Survey*

> "During convective warning operations, I prefer to have a 2-panel VIS and IR display for monitoring storm top convective trends. Persistent features such as overshooting tops and above anvil cirrus plumes indicate particularly strong updrafts and long-lived storms."
> *Forecaster*, "1-min VIS/IR during Warning Ops (1 May 2018)", *GOES-R HWT Blog*

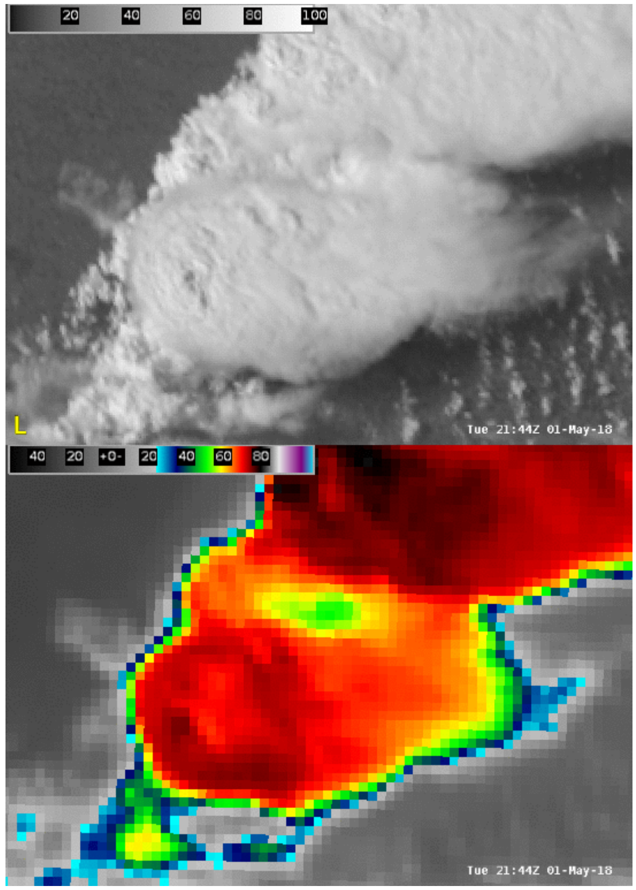

Figure 1: 2144 UTC 01 May 2018 GOES-16 0.64um “Red” Visible Imagery with 10.35um “Clean” IR Window Imagery over Southwestern Kansas.

Not only did the one minute imagery aid in the convective initiation and updraft monitoring phase of warning operations, but it also aided forecasters in null cases. It was helpful in identifying areas that were more stable and where initiation was not immediately expected. The imagery was also helpful in identifying storms with weaker updrafts and subsequently might be sub-severe when compared to other surrounding storms. Sometimes, radar characteristics weren’t as obvious and/or radar coverage wasn’t adequate to be able to pick out some of those details evident in the rapidly updating satellite imagery. Similarly, there were instances where the one-minute imagery captures characteristics of convective decay and the downward trends in convection.

> “The 1-min vis and IR imagery helped to identify convective initiation. The 1-min visible imagery was useful after storms had matured to visualize the sheared nature of the convection, and that storms were not developing updrafts robust enough to overcome the very strong shear.”
> *Forecaster, End-of-Day Survey*

> “I was able to witness outflow boundaries from a couple of isolated severe thunderstorms, and I noticed a progression of overshooting CB tops that progressed in a direction different from the advancing line of thunderstorms.”
> *Forecaster, End-of-Day Survey*

> “Good depiction of orphan anvils along dryline with failed convective attempts. These increased in # right before initiation.”
> *Forecaster, End-of-Day Survey*

> “There are parts of the country with marginal or poor radar coverage. I am able to use the 1-minute satellite imagery to fill in those gaps. Additionally, having the 1-minute imagery helped show me trends in convection that 5-minute radar coverage would be slightly slower to pick up and alert me to.”
> *Forecaster, End-of-Week Survey*

> “1-minute imagery proves useful in my CWA, as we have a lot of pulse type convection. The 1-minute imagery distinguishes sustained updrafts from pulse updrafts.”
> *Forecaster, End-of-Week Survey*

> “The 1-minute data may actually be MORE valuable for the marginal/less obvious severe weather days because those are the ones where the warning decisions are much tougher and therefore you rely more on supplemental datasets (aside from radar) to help you make your decision.”
> *Forecaster, End-of-Week Survey*

Furthermore, the forecasters were also asked about what ABI channels they used and which were found most useful in their daily operations. A large majority of the responses focused on the standard “Red” visible and “Clean” IR channels as the most useful in convective operations. An assortment of the three water vapor channels were also found to provided usefulness, particularly in the pre-convective analysis phase of the experiment. Many of the other channels were deemed to not have been used at all during operations. The only other channels to have been used a little more and found to be useful in some cases aside from those five mentioned were the “Blue” visible and the “Cloud Top Phase” channel. Many forecasters did comment on the fact that they find the other channels valuable in the creation of RGB composites, which can provide the forecaster with important information.

> “In the nowcast and warning environment, standard 0.64 um VIS and 10.3 um IR channel will still be the most valuable channels. We are using the single band imagery to monitor trends in the cu fields, convective initiation, and mature cloud top trends, and these channels provide that information best. Of course, water vapor imagery is important for tracking synoptic scale features. The new LL water vapor channel has shown some use in identifying lower/mid-level features of interest, but not in most situations.”
> *Forecaster, End-of-Week Survey*

> “The Visible & IR channels (especially 2 & 13) are essential for monitoring convective development. The WV channels are useful in a pre-convective environment to identify

subtle features that may be the forcing mechanisms for eventual storms. While I do not often use many of the other channels on their own they are very useful when combined into RGBs to monitor the environment & convective trends."
Forecaster, End-of-Week Survey

"I don't often use more than view ABI bands on a stand-along basis, but the fact that many of the RGBs and difference channels use the various bands makes them extremely important. The use of RGBs is a game-changer for warning operations in my opinion as it allows one to key in on various aspects of the convective environment that aren't as easily seen on the legacy Vis/IR images."
Forecaster, End-of-Week Survey

Use of ABI Baseline Derived Products in the HWT

Forecasters also evaluated the GOES-16 baseline derived products in the HWT experiment this year. There was a heavier emphasis on the derived products this year than the imagery since the imagery had already been declared operational. The forecasters mainly focused on products of relative importance to convective initiation and forecasting, as well as those which could be useful for warning operations. So, not every product available from GOES-16 was formally evaluated during this experiment. Each day forecasters were asked which products they used and what they were used for, and at the end of the week, they were asked to explain which products would have the most impact on improving warning operations in the field. Overwhelmingly, the products used the most and which provided the most impact were the derived stability indices and the Total Precipitable Water products. These products, as expected, provided the most utility in a convective environment, particularly prior to initiation when looking for areas where storms could initiate. These products, most notably the CAPE, tended to underdo the values quantitatively when compared to RAP model derived values, but in a qualitative sense were seen as helpful for forecasting convective development and in some cases decay by identifying regions of greater relative instability and air mass gradients. Forecasters did note that they would rather have the gaps in data filled in, as in the All-sky version, as it helps with filling in the gradients and analyzing the pre-convective environment in the case of more cloud cover. Another product used at times was the derived-motion winds. This product was used primarily in a forecast sense and looking for jet maxes and areas of deep-layer shear that would subsequently affect the storm initiation and growth. Some other products were also examined, such as the Cloud Top Temperature, Phase, and Pressure products. These products were used after storms had initiated while forecasters were monitoring for new updraft growth or decay within the storm. Feedback on these derived products is presented in this section.

"Yes, I looked at baseline derived CAPE and PWATS for assessment of downstream environment from ongoing convection. They provided additional confidence in what I was seeing on the SPC mesoanalysis fields."
Forecaster, End-of-Day Survey

"Yes. I used the baseline GOES 16 derived CAPE product. It wasn't the most useful given much of the CWA had data blackout due to thick cloud cover."
Forecaster, End-of-Day Survey

> "Yes, I utilized some of the derived CAPE data which was useful for diagnosing, to a degree, the environment far downstream of active convection. However, with the convective blow off, the environment immediately downstream of the convection could not be analyzed."
> *Forecaster, End-of-Day Survey*

> "Yes, All-Sky Lap CAPE, LI & TPW on the Mesoscale 1 sector over NC/VA These products are particularly useful in helping with highlighting pre-convective environments, gradients in particular, and they can also help the forecaster build a multi-dimensional perception of the atmosphere to begin looking at more specific severe threats."
> *Forecaster, End-of-Day Survey*

> "The satellite derived winds highlighted the upper level jet and diffluence aloft as it approached the forecast area. Storms developed under the area of diffluence as one would expect. Also, increasing low level southerly winds were observed across the region, indicating moisture advection and favorable shear given westerly/southwesterly mid/upper level winds."
> *Forecaster, End-of-Day Survey*

> "The Derived Motion Winds provide the most value for convective warning operations. In the absence of observed wind data both spatially (vertically and horizontally) and temporally, the satellite winds are provide added value. Uses include: 1) identification of low-level and upper level jets and comparison with the model jet depiction to see how well the model is handling the situation, 2) using surface ob to wind-level DMW to compute bulk layer shear, 3) identification of upper-level diffluence and low level convergence patterns."
> *Forecaster, End-of-Week Survey*

> "Cloud top temperatures and heights were used fairly extensively for monitoring storm development and growth. Certainly these were used in tandem with the more widely-utilized "standard" IR imagery. Yes, this imagery would be helpful, specifically prior to or leading up to warning decision processes."
> *Forecaster, End-of-Week Survey*

Some of these products played an important role in forecasting for a convective event in the Des Moines, Iowa, area on 3 May 2018. The forecaster noted a closed upper low moving across southern Nebraska early in the afternoon from upper-level water vapor imagery (Fig. 2). Visible imagery showed low clouds across much of the County Warning Area (CWA), with some more vigorous cumulus clouds beginning to develop across the far southern portions of the CWA in the clear warm sector. GOES-16 derived-motion winds revealed a 110+ knot southwest-to-northeast-oriented jet core at 300 mb across Iowa with another jet core advancing into southern Iowa throughout the afternoon from northern Missouri, which agreed with the GFS forecast for the area. Winds in the 750mb to 850mb layer were generally 20 knots on average across the area with slightly stronger winds advancing into the southern portions of the CWA implying increased low-level warm air advection (Fig. 3). These observations helped increase confidence for convective initiations across the southern portion of the CWA later in the afternoon with increasing lift arriving into the region. The forecaster also noted looking at the derived stability indices for the event, but due to large cloud cover there was not much to be determined, however the All-sky version did show some increasing values of CAPE also moving into the region from the south, though the values were underdone.

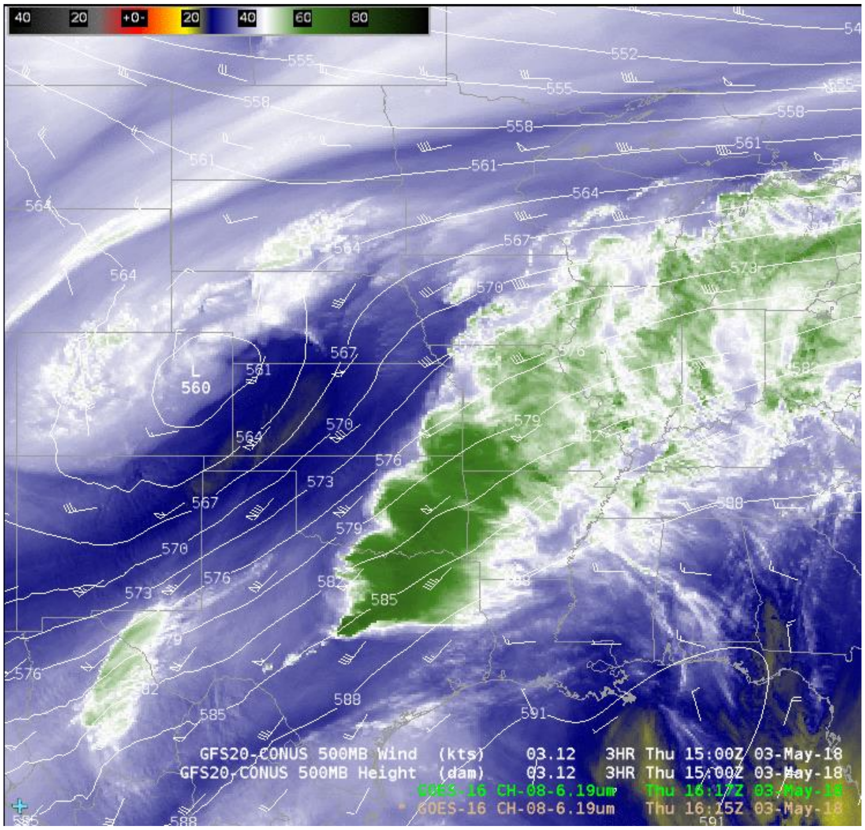

Figure 2: 1617 UTC 03 May 2018 GOES-16 6.19um “upper-level water vapor” imagery with GFS 500mb heights (white contour) and winds (white wind barbs).

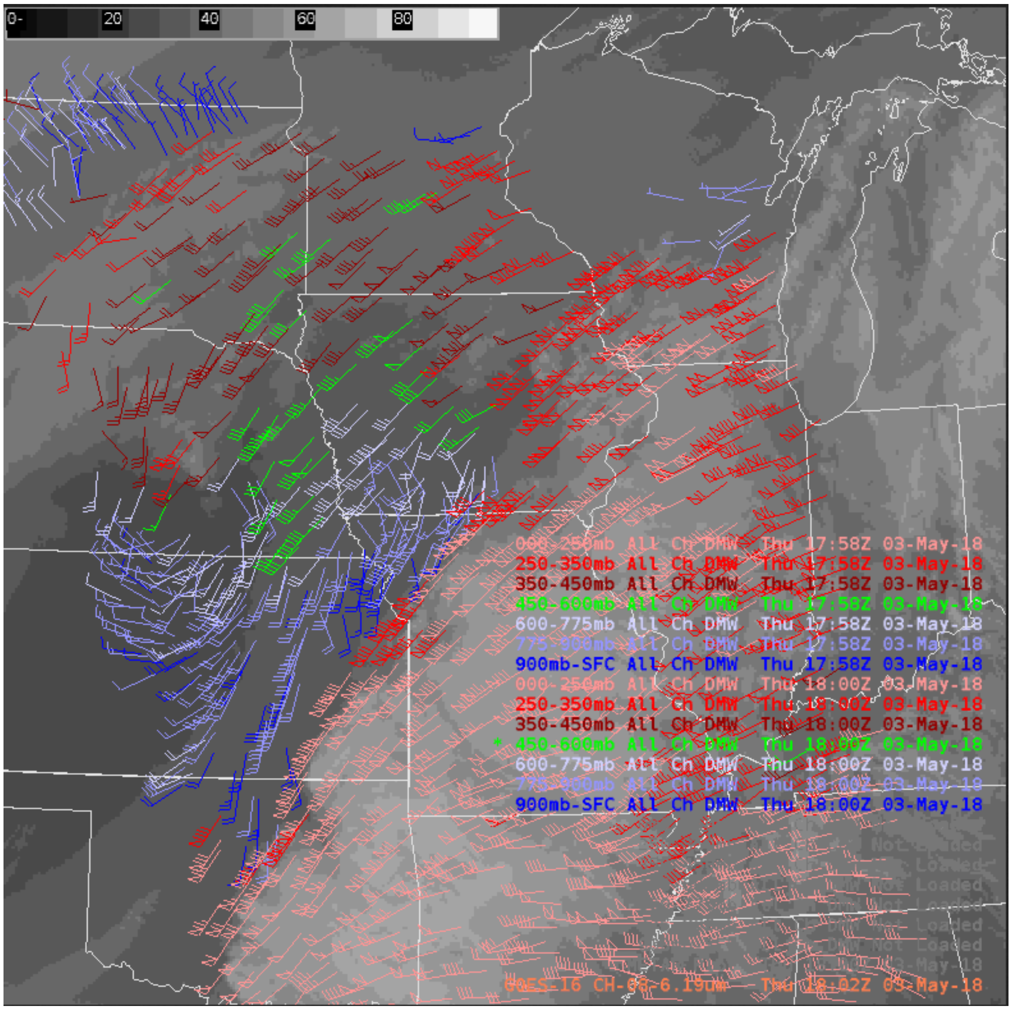

Figure 3: 1800 UTC 03 May 2018 GOES-16 Derived Motion Winds (colored by height, blue – low-level, green – mid-level, and red – upper-level) with GOES-16 6.19um "upper-level water vapor" imagery (gray scale)

Suggestions for Improvement of Derived Products

There were some suggestions for improvement of some of the derived products for greater usefulness in routine operations. The most common suggestion from the forecasters was that it was much more preferred to have the data gaps filled in on the stability indices and the precipitable water. Throughout the weeks, the forecasters continually noted how they found the All-sky version of these products easier to use and analyze instead of having to try to fill in gaps themselves. Those products have some flaws of their own that forecasters would like to see fixed, which will be discussed later, but they still overwhelmingly used the All-sky products more in their operations during the experiment. Also, work could be done to obtain values, particularly of CAPE, that are closer to what is seen in other data sources. Currently, the stability fields are consistently underdone, though the gradients and relative values are consistent with other datasets. More accurate values would help give the forecasters confidence in the use of the products. For the derived-motion winds, there was one area that was noted by forecasters to be fixed for easier use in operations. When looping the winds, different levels of wind data would show up at different time steps causing the loop to be jumpy and making hard to pick out trends in the data. Forecasters think that the wind data should all be time matched and synced so that when the data is looped it flows seamlessly from one image to the other without jumping around to different levels and different times. Some quotes and feedback from the forecasters is posted below.

> "I used some baseline derived stability products today, but I preferred the All Sky LAP imagery with the cloudy gaps filled in."
> *Forecaster, End-of-Day Survey*

> "I did not rely heavily on the derived baseline products today given the widespread cloud cover across the region."
> *Forecaster, End-of-Day Survey*

> "I looked at the Total Precip Water and the Derived Stability Indices. They helped show the location of the dry line and the instability ahead of it. However, the indices did seem to be underdone in areas where there was a good retrieval."
> *Forecaster, End-of-Day Survey*

> "IT would be nice if the winds weren't so jumpy when looping the data. They should be time matched to make them flow more seamlessly."
> *Forecaster, Daily Debrief*

3.2 GOES-16 RGB Composites and Channel Differences

National Weather Service (NWS)

The last GOES-16 related products to be evaluated during the experiment were the numerous RGB composites and channel differences. These products are created on the fly within AWIPS-II and combine multiple channels together to highlight certain features related to a forecast problem or certain phenomena. Numerous convective RGBs were examined in the HWT to examine their value for convective forecasting and development. A couple of channel differences were also evaluated for their use in convective environments. Channel differences are also created within AWIPS-II by subtracting the values of one channel from another to pull out information that can be important and otherwise not noticed when looking at a single channel.

Use of RGB Composites in the HWT

First, there was an overall sentiment from the forecasters that better training is still needed for forecasters to be able to fully apply RGB use in operations. Reference sheets or "quick guides" were distributed to each forecaster to help them become more familiar with the RGBs and channel differences. After they became more familiar with the RGBs and learned what certain colors represent within the RGB, a number of them were used throughout the remainder of the each week of the experiment. A large number of forecasters focused in on two specific RGBs throughout the experiment: the Day Cloud Phase Distinction RGB and the Day Cloud Convection RGB (Fig. 4). These two RGBs aided in identifying areas of convective initiation by identifying clouds that have glaciated from those that have remained mostly water clouds. These two RGBs also allowed forecasters to distinguish between more mature deep convective clouds from those that are less mature. Many of the other RGBs were also examined throughout the weeks and the Simple Water Vapor RGB and Differential Water Vapor RGB were identified as having some utility for monitoring subtle larger-scale features, such as shortwave troughs and jets that were a little less discernable than the single-channel water vapor channels.

> "I continue to find the day cloud convection RGB incredibly useful in terms of visually separating more mature convection from developing or soon-to be convection."
> *Forecaster, End-of-Day Survey*

> "Yes, I utilized the Day Cloud Convection RGB which provided a great perspective on differentiation between the shallow cu field and the deeper convection just upstream of this shallow cu field. The ability to distinguish between low and high level clouds in a "mixed high/low" (i.e. cirrus above shallow cu) environment was very beneficial."
> *Forecaster, End-of-Day Survey*

> "Day Cloud Convection and Day Cloud Phase Distinction. Both helped to show convective growth ranging from the low, warm, water-based clouds all the way up to the high, could ice-based anvil. It was different displays and color schemes for each. But both worked well to show the convection. Differential Water Vapor. It helped give a general view of the moisture in the atmosphere."
> *Forecaster, End-of-Day Survey*

> "Yes, the Simple Water Vapor RGB which aided in verifying the saturated environment at least in the mid to upper levels. Also, looked at the Air Mass RGB but didn't really see much added value probably due to limb effects. Utilized the Day Cloud Convection RGB as well which highlighted in great detail the different stages of the cloud development and increasingly well-defined structure of the cells."
> *Forecaster, End-of-Day Survey*

> "Yes, the day cloud convection RGB was incredibly useful for identifying "feeder" low level cumulus into developing updrafts which contained ice."
> *Forecaster, End-of-Day Survey*

In the case shown below from 22 May 2018 in Southwest Texas, a forecaster utilized the Day Cloud Phase Distinction, Day Cloud Convection, and Simple Water Vapor RGBs to monitor for convective initiation across the region (Fig. 4). The forecaster also had a MRMS -20C isothermal reflectivity product up monitoring for the first radar echoes. He noticed that there was a large cumulus field across the region with some more vigorous growth taking place over the Davis Mountains in SW Texas. The Day Cloud Phase Distinction RGB shows the greenish towering cu turning to a yellow color indicating that glaciation is occurring and initiation is underway. The forecaster also noted that the yellow cu field turned to a white color in the Day Cloud Convection RGB, again indicating that convective initiation was underway as the first radar echoes began to appear on the -20C isothermal reflectivity scan. These RGBs helped alert the forecaster that initiation was occurring in an environment where storms were expected throughout the day and that warning operations could be starting soon. These RGBs were used in this way many times throughout the experiment to keep forecasters aware of when initiation was beginning across their area of responsibility.

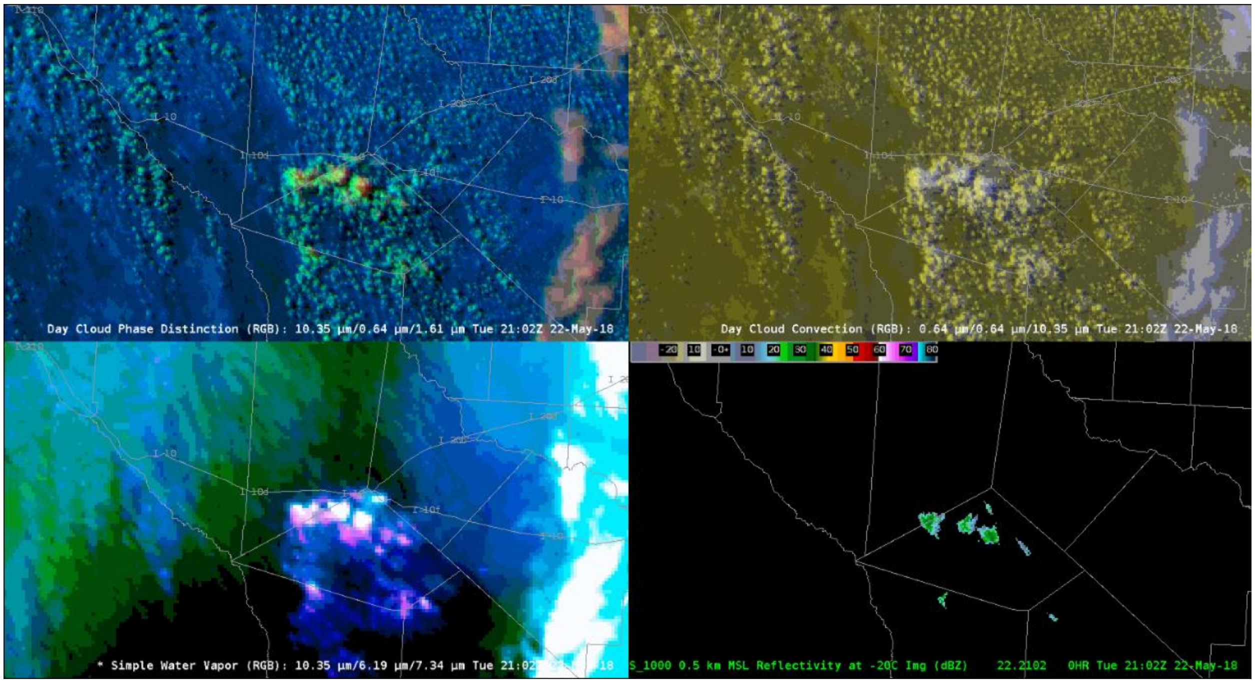

Figure 4: 2102 UTC 22 May 2018 Four Panel or Day Cloud Phase RGB (upper left), Day Cloud Convection RGB (upper right), Simple Water Vapor RGB (lower left), and MRMS -20C Reflectivity (lower right) denoting convective initiation occurring across SW Texas as towering cumulus clouds have begun to glaciate and become mixed phase clouds.

In addition to those RGBs mentioned above, forecasters also examined some of the other RGBs that were available during the experiment. Many of these other RGBs that were looked at were not used very heavily during the experiment and didn't provide a lot of additional value for the purpose of this experiment. A couple of RGBs that did provide some utility were the Air Mass RGB for locating synoptic scale features and the Fire Temperature RGB for identifying some hotspots.

> "Yes, I did use the Air Mass RGB. Verifying the location of the potential vorticity wrapping around the cusp of the comma head and from the cold conveyor belt wrapping around the shortwave trough."
> *Forecaster, End-of-Day Survey*

> "I looked a little at Fire temperature as we got started for wildfires to the east of our area"
> *Forecaster, End-of-Day Survey*

Use of Channel Differences in the HWT

The ABI channel differences were also evaluated during the experiment this year. The Split Window Difference was used most throughout the experiment, mainly for trying to identify low-level moisture pooling ahead of a dryline or surface boundary in the absence of clouds. This difference did provide some value at times in some of these specialized circumstances, largely before widespread cumulus fields were present. The other channel differences were also examined and were not really seen to provide any value for this convective experiment except for a couple of instances where the Split Water Vapor Difference was used to highlight the position of the dryline.

> "I did use a split water vapor product, which provided a value in terms of being able to see where the low level dry air was oriented with convective initiation just east of the advancing drier air."
> *Forecaster, End-of-Day Survey*

> "I looked at the split difference window and also placed the layer precipitable H20 and the mid-level layer precipitable H20. The split difference imagery at least verified the moisture transport."
> *Forecaster, End-of-Day Survey*

> "The Split Window Difference provides the most value for convective warning operations, but only under the right conditions (clear skies, decreasing temperature with height). In most cases, it doesn't provide much value. None of the other channel differences provide value to warning operations that I have experienced."
> *Forecaster, End-of-Week Survey*

> "The split window difference proved useful to me, as I was able to see moisture pooling, which is useful to see where a CU field and ultimately convection may fire."
> *Forecaster, End-of-Week Survey*

> "Yes, split water vapor. Again, it showed where the dryline was and where storms would fire up and how much moisture they had to work with."
> *Forecaster, End-of-Day Survey*

Limitations and Recommendations for Improvement of RGBs and Channel Differences

The major downfall particularly to the use of the RGBs and Channel Differences for the forecasters this year remained the insufficient training available to them to learn how to apply these products in their operational workflow. While training was provided for them during their week here, the forecasters became much more proficient in their understanding and use of the products, but more is needed for proper interpretation and application in a fast-paced operational environment. There were also many comments on the color tables being used for the channel differences, especially the Split Window Difference, being hard on the eyes and hard to discern anything in the imagery. The use of so many bright colors together in the Split Window was seen as a hindrance to the analysis of the imagery. Many suggested possibly using some type of grey scale color table with maybe a brighter color for higher values to really make the areas of moisture pooling stand out against the surrounding image. Right now, it can be easy for the larger difference values to become "washed out" and blend in with the surrounding imagery, especially during the warm season. So, some adjustment of the color table may be necessary as well to pull out the important parts of the difference. Some other complaints about the RGBs was that they would like to see the Day Convection RGB available on the CONUS sector instead of just the Full Disk, because of the problems in the recipe of differencing bands of different resolution, as many think it might be pretty useful once at its proper resolution. Also, many of the forecasters suggest trying to find a way to limit some of the effects of sun angle on the imagery during sunrise and sunset and also the limb effects. These differences often render the imagery unusable in those areas and if there could be a way to mitigate that, then the RGBs might get quite a bit more use particularly across areas where the limb effects play an important role.

> "Yes. Split Window and Split Cloud Top Phase. Both are a little hard to use due to color tables and the info they are trying to convey. I included some examples in a blog entry."
> *Forecaster, End-of-Day Survey*

> "I did use the Split Window product today, trying to get a feel for the low-level moisture gradient. I didn't find it particularly useful, however. It may be that particular color curve, but the strong reds/pinks are really an assault on the eyes in some regimes."
> *Forecaster, End-of-Day Survey*

> "Briefly used the Air Mass RGB but due to parallax issues, the data seemed to be inconclusive in determining the air mass differences there."
> *Forecaster, End-of-Day Survey*

> "I did look at the Air Mass RGB. Really tried to see where the PVA might have been located around the shortwave trough. Probably need more training to gain more experience with it."
> *Forecaster, End-of-Day Survey*

> "...Also, looked at the Air Mass RGB but didn't really see much added value probably due to limb effects..."
> *Forecaster, End-of-Day Survey*

3.3 Probability of Severe (ProbSevere) Model

Cooperative Institute for Meteorological Satellite Studies (CIMSS)

The NOAA/CIMSS ProbSevere statistical model, planned for operational implementation by NCO as an update to MRMS in the coming year, was evaluated in the HWT for the fourth consecutive year, with updates made since last year's experiment. The ProbSevere Model utilizes the GOES-16 data within its statistical model for the first time for this year's experiment. ProbSevere is currently undergoing tuning and assessment with the in-orbit GLM data for future demonstrations. The statistical model produces a probability that a storm will first produce any severe weather in the next 60 minutes (Cintineo et al. 2014). The data fusion product merges RAP model-based instability and shear parameters, satellite vertical growth and glaciation rates, radar-derived maximum expected size of hail (MESH), and Earth Networks (ENI) total lightning information. Additional RAP and Multi-Radar Multi-Sensor (MRMS) fields such as azimuthal shear were also used in the model this year to provide guidance on specific severe hazards of tornado, wind, and hail. ProbSevere tracks a developing storm incorporating data from both satellite and radar imagery using an object-oriented approach. As the storm matures, the Numerical Weather Prediction (NWP) information, lightning data, and satellite growth trends are applied to the overlapping radar objects. The product updates approximately every two minutes and is displayed as contours with different colors and thicknesses corresponding to different probability value bins that are overlaid on radar imagery. Data readout is available by moussing over the probability contour, revealing the probability of severe for each hazard (hail, wind, and tornado), along with the model predictor input values. This year, there was also a separate product the forecaster could load up to display a contour for each separate hazard if they so chose to. The product was evaluated on its ability to increase forecaster confidence and skillfully extend lead time to severe hazards for NWS warnings during potential severe weather situations. Additionally, feedback regarding the product display and readout was solicited.

Use of ProbSevere in the HWT

Forecasters primarily loaded the ProbSevere guidance as an overlay on either base radar data, or MRMS products (e.g., Composite Reflectivity, MESH, and Reflectivity at Lowest Altitude) toward the beginning of each shift or when storms began to initiate. Early on in the shift, ProbSevere alerted forecasters to the first storms of the day that were becoming potentially severe and warranted closer monitoring and analysis. The consistent view among all forecasters was that ProbSevere was very useful for situational awareness and for alerting forecasters of storms requiring more radar interrogation. This was especially important in busy warning environments, where there were many ongoing storms. ProbSevere was often used as a guide to quickly rank storms in terms of importance to interrogate based on higher probabilities (Fig. 5). Forecasters would also load up a four-panel layout of contours for the full ProbSevere along with the contours for each individual hazard (ProbWind, ProbHail, and ProbTor) to examine which hazard was becoming the dominate threat (Fig. 5). In other situations where warning issuance was more marginal or uncertain based on the environment or base radar data, ProbSevere would sometimes provide more confidence to issue (or not issue) the warning based on the probabilities. It is important to know that forecasters did not use ProbSevere alone to issue warnings, but instead used it as another piece of information to increase the confidence in the decision making of the forecaster. Forecasters would also often take note of the parameters within the readout of the ProbSevere contour in AWIPS-II to monitor changing input parameters to help better interpret the trends in the probability values. This gave confidence to the forecaster in how the storm attributes were evolving and how the local environment was changing, and provided much more insights into the ProbSevere algorithm performance. With all of the great uses for ProbSevere, there were also several limitations and suggestions from the forecasters to improve both the model and product display. Both uses and limitations will be discussed in this section.

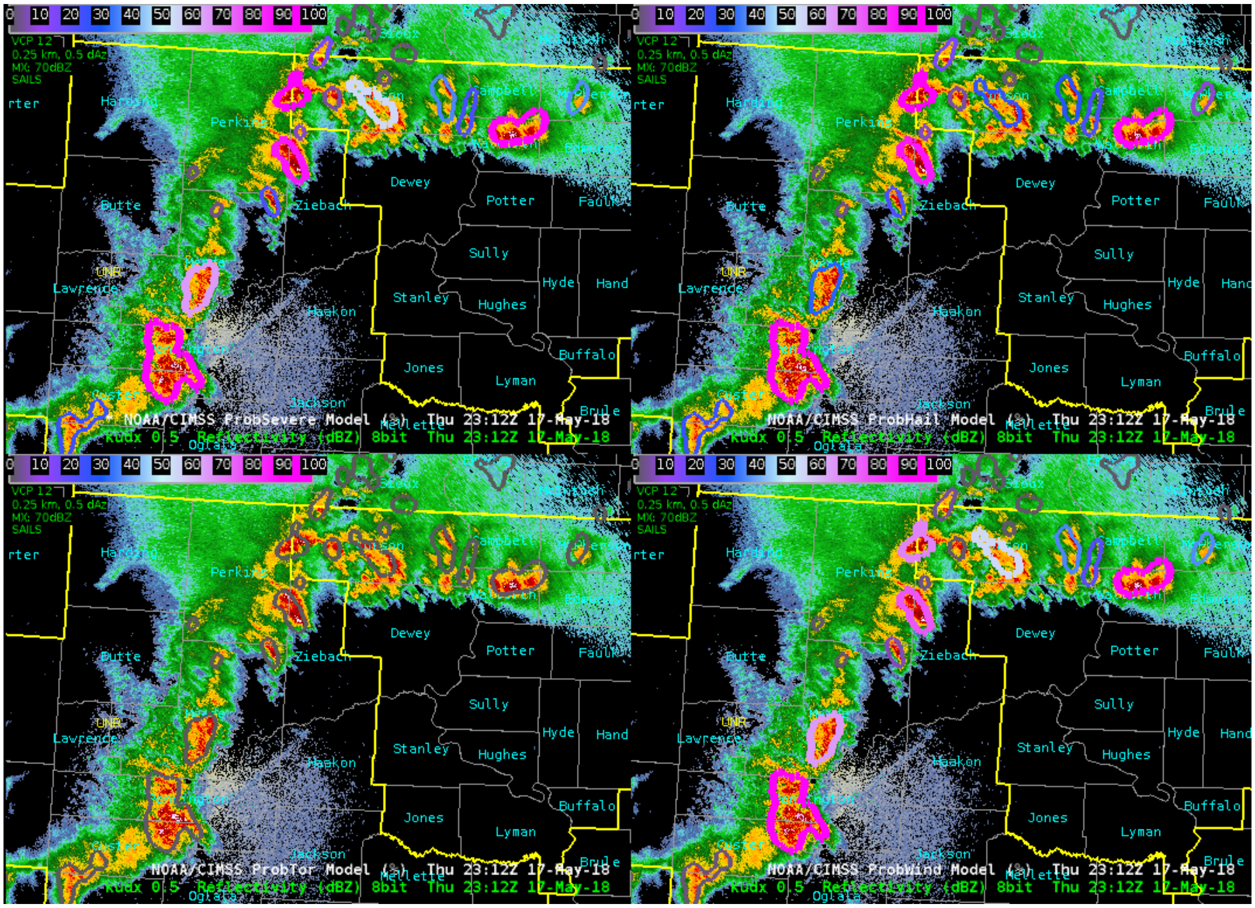

Figure 5: 2312 UTC 17 May 2018 4 panel of 0.5 degree radar reflectivity and ProbSevere All Hazards (upper left), ProbHail (upper right), ProbTor (lower left), and ProbWind (lower right) contours for a complex of storms in western South Dakota. ProbSevere helped forecasters rank storms to interrogate in this busy environment.

> “Great situational awareness tool, it gets used in the warning decision process, along with other storm scale radar products.”
> *Forecaster, End-of-Day Survey*

> “It was at least a good product for situational awareness purposes, but also having a readout also assisted in showing the trend in the storm evolution.”
> *Forecaster, End-of-Day Survey*

> “ProbHail was very useful as a SA tool, especially at the onset of the event. They keyed in on building supercells, which were quite a few in our area today with many cell mergers and a few discrete cells.”
> *Forecaster, End-of-Day Survey*

> “Provided SA for the strongest and possible more severe storm in the BYZ CWA”
> *Forecaster, End-of-Day Survey*

With the addition of inputs into the algorithm, the ProbSevere model was able to be tested in the HWT experiment this year to provide probability guidance for each individual hazard type (wind, hail, and tornado) separately (Fig. 5). Forecasters were able to load up the full ProbSevere "All Hazards" version which presented them with a full readout of all the inputs and probabilities of the model, but they were also able to load up a separate contour for each hazard type and see the readout for the inputs only associated with that particular hazard. Overall, feedback was positive on the additions of the separate contours for each hazard and forecasters liked being able to see which parameters were affecting each hazard and how those changes affected the probability guidance. Having information on the separate hazards was helpful to the forecaster in better identifying the severe risk and for improving the wording of hazards within the warning text. In fact, 94% of forecasters answered that they thought having all of the options available to them for each hazard was important and provided value to operations. The full readout was good for one-panel analysis, while the separate hazards was better to use in a four-panel display, so that the readout wouldn't take up the entire panel or go off the screen.

> "Yes, having all the options is important. The Full version is more suitable for a single-pane situational awareness display (partly because of the large pop-up data table), and the individual versions are more useful in a four-panel display. Different people like visualize things in different ways, so having multiple options is a good idea."
> *Forecaster, End-of-Day Survey*

> "I like having all three available. The individual components can be plotted in tandem (on a 4 panel) to give a quick holistic view of the area and triage accordingly. The highest probability doesn't dominate the algorithm in this case."
> *Forecaster, End-of-Day Survey*

> "On days where there are multiple threats, given that you keep situational awareness on each potential severe element, I see how this can give confidence for each individual threat."
> *Forecaster, End-of-Day Survey*

At the very least, ProbSevere enhanced forecaster confidence when issuing severe thunderstorm warnings, but played a bit of a lesser role for tornado warnings, though there were not many tornado days during this year's experiment. Forecasters were asked to assess the role that ProbSevere played in there warning decision making as it relates to confidence in the warning decision and lead time provided. Approximately 82% of forecasters responded that they felt ProbSevere increased their confidence in issuing severe thunderstorm warnings compared to approximately 61% responding that it increased their confidence in issuing tornado warnings. Similarly, 57% of forecasters responded that they felt ProbSevere helped increase their lead time for issuing severe thunderstorm warnings, compared to only 16% responding that they felt ProbTor helped increase their lead time for tornado warnings. Again, there were very few tornadoes during this experiment which could have prevented a more robust evaluation of the ProbTor probabilities. Forecasters noted that the most important value out of the ProbSevere model were the trends in the probabilities. A quick jump in the probabilities was a key indicator to forecasters that the storm was intensifying rapidly and required close examination for a possible warning. Forecasters recommend that it is still best to wait a couple of scans to see the trend in ProbSevere and to examine the base data further before issuing any warnings. When asked at the end of each shift whether they would use ProbSevere back at their office in operations, 66 out of 67 responses were that yes, they would use ProbSevere during operations back at their office.

> "Similar to above, the fact that ProbHail values were increasing due to environment and lightning, faster than what was apparent in radar imagery, provided enough confidence to pull the trigger earlier than I otherwise would have. This was especially the case with the second storm, after experiencing this with the first storm, having more confidence."
> *Forecaster, End-of-Day Survey*

> "It provided early-on confidence for warnings for hail. As a line of thunderstorms formed, it showed the wind threat before I realized it."
> *Forecaster, End-of-Day Survey*

> "When used in conjunction with other products and datasets, ProbSevere helped increase confidence and lead times with warning issuance."
> *Forecaster, End-of-Day Survey*

> "Used in conjunction with GLM products I was able to issue a SVR warning ~15 minutes earlier than I normally would have using "conventional" storm interrogation tools. When I initially issued the warning ProbSevere values were still quite low (less than 30% on ProbHail) but the increasing trend from the single digits I noticed along with the lightning activity from GLM was enough to give me confidence to issue a warning. ProbSevere aided my decision making on continuing to issue warnings on the storm even as GLM data showed a decreasing trend. Probhail values remained above 50% on the storm & MESH values were above 1"."
> *Forecaster, End-of-Day Survey*

> "While the ProbSevere certainly doesn't carry the weight of base radar data, having a tool that takes into account many of the other tools that I rely upon to make warning decisions is a time saver. The ability to mouse-over a cell and get a readout from all of the inputs is invaluable. It keeps me from having to switch screens and pull up other data on the fly."
> *Forecaster, End-of-Day Survey*

The following example shows a case where ProbSevere helped a forecaster have increased confidence in issuing a severe thunderstorm warning prior to when he would have issued if just looking at base radar data. This example also illustrates how the forecaster used the readout of different hazards to identify the main threat within the storm. This example was from a case in the Dodge City, Kansas, CWA with a developing storm in a pretty favorable environment on 01 May 2018. The following is taken straight from a blog post where the forecaster explains his decision making for the storm shown in the figure below (Fig. 6).

> "From 2008 to 2010 UTC, ProbHail jumped from 39% to 58%, based on a slight increase in MESH from 0.40" to 0.51" and flash rates increasing from 6 to 16. From 2010 to 2012 UTC, ProbHail increased to 95% and MESH to 0.73" and flash rate to 34. This is

valuable to note: Despite MESH still well below 1", ProbSevere is at 95%, indicating high likelihood of severe (hail in this case). A key driver being favorable lightning in a favorable shear environment. The rapid trend up in probs combined with high probs helped to provide confidence in the warning decision. This despite the fact that MESH was still well under 1", and the reflectivity core was still on its way up, but not quite to where I would want it for my typical warning decision. This storm quickly produced melting 1" hail, and later 1.5" hail."
Forecaster, 01 May 2018 Blog Post, "ProbSevere with First Storm in DDC"

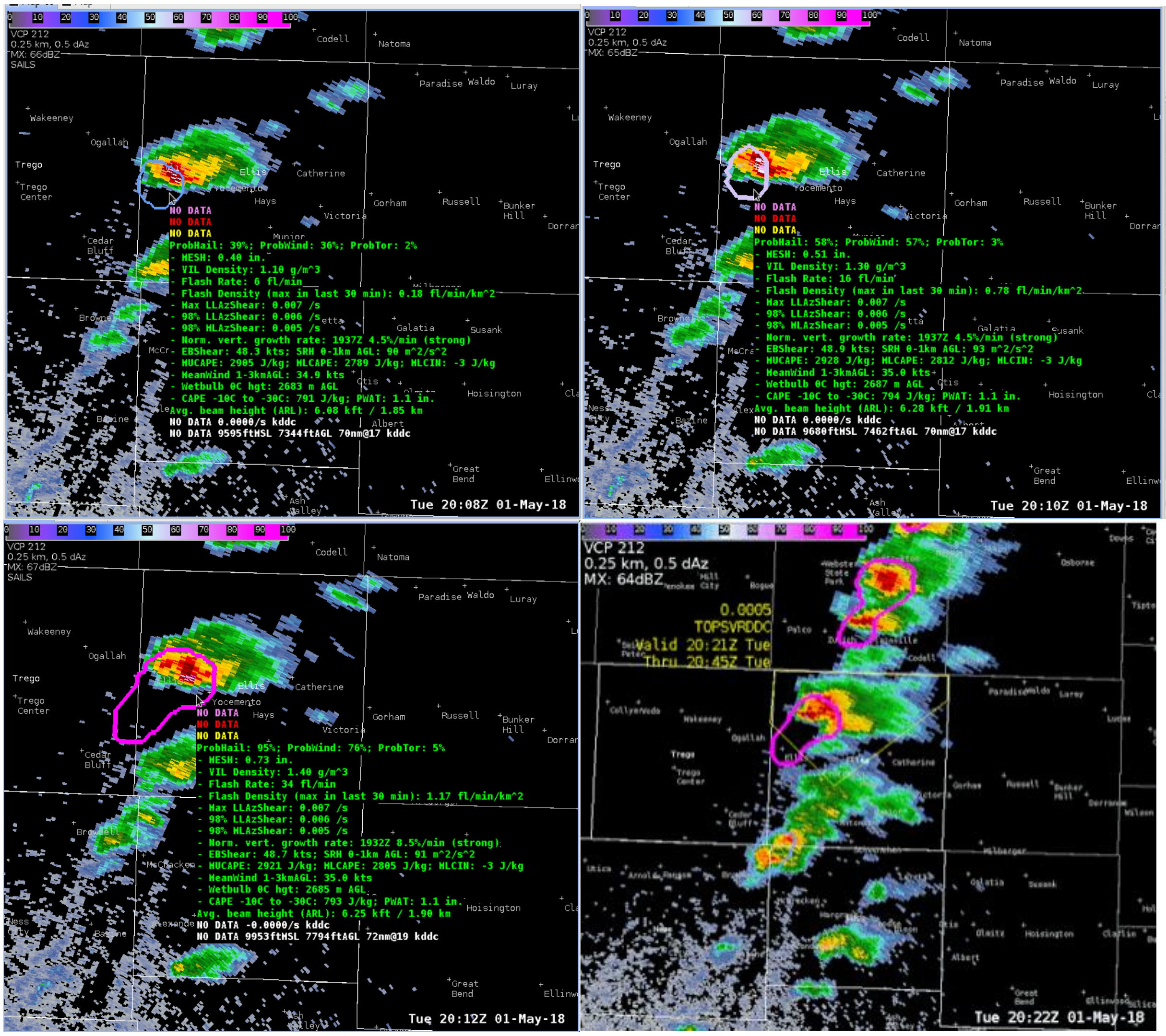

Figure 6: 01 May 2018 evolution of ProbSevere probabilities with base reflectivity on a storm in western Kansas. 2008 UTC (upper left), 2010 UTC (upper right), 2012 UTC (lower left), and 2022 (lower right) showing the issuance of the severe thunderstorm warning (yellow polygon).

Limitations and Recommendations for Improvement of ProbSevere

There were still several instances commonly pointed out by forecasters where ProbSevere was not as effective. Forecasters found that occasionally the increase in probabilities lagged slightly behind the strengthening in the base radar data both spatially and temporally. This is most likely attributed to the latency seen in the MRMS data processing and subsequent ingesting into the model causing a delay of a couple of minutes. In scenarios where storms were rapidly developing, this seemed to hamper the lead time of ProbSevere compared to that of using base radar data. Also, when storms were in close proximity to each other, the ProbSevere model would typically cluster storms together into one big contour, often times causing a false increase or reduction in severe probabilities, which rendered the probabilities less useful during those circumstances. Forecasters suggested that they would like to see the tracking improved to more closely follow the cores of storms better and offered the suggestion of utilizing the -10C reflectivity from MRMS to track instead of using composite reflectivity. It was noted that this might cause a bit more delay in the model output, but forecasters would rather see the model track the individual updrafts better to make it more useful in these situations. While on the subject of tracking, forecasters also noted numerous times when the storm tracking algorithm wouldn't include the gust front or areas of highest winds within the contour and thus produce a false low value for ProbWind when velocity data showed intense winds within the storm (Fig 7). This again emphasizes the importance of accurate storm tracking to provide the most useful probabilities from the model.

> "Doesn't seem to capture severe outflow boundaries too well, so this is an area of continued research. Merging cells tend to skew the data...so fixing this issue should be a priority."
> *Forecaster, End-of-Week Survey*

> "There needs to be a better way to discriminate the individual cells for the ProbSevere products. This would certainly improve the above mentioned Time Series Tool."
> *Forecaster, End-of-Week Survey*

> "As we discussed, the biggest potential improvement would seem to be doing a better job of identifying individual storms within a linear convective system. Possibly using reflectivity at -10 deg C would be worth the extra minute or so of latency, if it provides more consistent identification of severe cells."
> *Forecaster, End-of-Week Survey*

> "ProbSevere was a bit jumpy today owing to false "clustering" of storms. As ProbSevere identified a cluster as one storm, the values jumped significantly. However, sometimes only a scan later, ProbSevere correctly identified the activity as multiple storms nearby (i.e. multiple updraft regions). This would cause the ProbSevere to drop suddenly. From my perspective, ProbSevere is most useful for discrete or single-cell convection in strongly unstable environments."
> *Forecaster, End-of-Day Survey*

As has been the case in previous years of the experiment, ProbSevere was found to be most useful for discrete storm modes and cases where severe hail was the primary threat. This isn't all that unexpected as there is no other input for severe wind and tornadoes that performs as well as MESH does for hail guidance, and with the addition of lightning data, the hail guidance seems to have become even better calibrated to not rely solely on MESH. Forecasters did comment that with each iteration of ProbSevere that it seems to do better with the other hazards of wind and tornado than previous versions. Forecasters mentioned that the model could be trained better for other regions of the country with different environments and threats than those seen in other areas. Also, as mentioned earlier, a lot of the problems with ProbWind as well as ProbTor seemed to stem from cell tracking issues within line segments or clusters of storms. There were times where ProbWind would be low but velocity data would show a strong outflow racing ahead of the storm (Fig 7). This was also a problem when looking at ProbTor within a QLCS, where forecasters noticed that there would be circulations outside of the contour. Some forecasters offered suggestions for possibly allowing a bit of a buffer in these cases to try to pull in more of the data associated with these hazards. Forecasters also said that they would like for there to be more velocity data incorporated into MRMS and the ProbSevere model where more of the raw velocity data could be picked up to hopefully improve the performance of ProbWind.

> "It would be nice to see elevated probabilities highlighted within a QLCS and polygons separate more efficiently with storm splits. I would like to see velocity magnitude added as an MRMS product, as this would likely improve ProbWind probabilities, along with being useful on its own."
> *Forecaster, End-of-Week Survey*

> "Based on experience with the ProbWind product there needs to be more velocity or reflectivity information put into the algorithm to help prevent false alarms."
> *Forecaster, End-of-Week Survey*

> "In a linear situation, the ProbSevere picks up the entire cluster and goes with the highest value. This is misleading. So this tool really has to be used with other products to figure out where the severe risk is in that cluster. This would be an issue for the ProbTor too. The whole line would be in the polygon, when that risk is much more centered in a point. ProbTor as a separate polygon would help (which I did see is an option). The radar scans at 1 minute would update 1-2 minutes quicker than the ProbSevere. If this could match that would be great! The color scale was a tad confusing, but then again it stood out from the radar colors."
> *Forecaster, End-of-Week Survey*

> "The activity was largely sub-severe and the one storm that produced severe wind was apparent through base velocity. ProbWind did not perform as well since the gradient along the radial was weak. More velocity information going into the algorithm would help with these cases."
> *Forecaster, End-of-Day Survey*

> "ProbTor didn't handle the QLCS tors very well because the couplet is usually outside of the object"
> *Forecaster, Weekly Debrief 04 May 2018*

Suggestions for Product Display

In general, forecaster liked the ProbSevere contour display and readout options, but there were circumstances that forecasters offered some suggestions for improvement to the current display. Some forecasters noted that they would like to see an alternative color scheme that would make the jumps in probabilities stand out more and not be such a smooth gradient. Multiple forecasters suggested creating a color scale showing the probability contours changing color in ten percent increments with contrasting colors so that magnitude increases are more readily apparent. Many forecasters made their own color table to do this during the experiment and perhaps having multiple options for color tables would be helpful to forecasters. Forecasters also commented on having an easier method to quickly assess trends in probabilities such as a time series tool. Forecasters would also like to see qualitative wording added to the readout alongside the values for azimuthal shear so it is easier to quickly realize if the shear values are strong or weak and how they are affecting the probabilities. Finally, there were also numerous comments about being able to have a way to have the contour for ProbSevere and ProbTor overlaid on the same display where the changes in ProbTor could be more easily distinguished. Some suggested possibly making the ProbTor contour thicker or making it outline the ProbSevere outline with a different color scale so that the changes in ProbTor can be more easily seen without having to dig into the readout or have a separate pane devoted to just ProbSevere. Comments on the product display are shown below the figure.

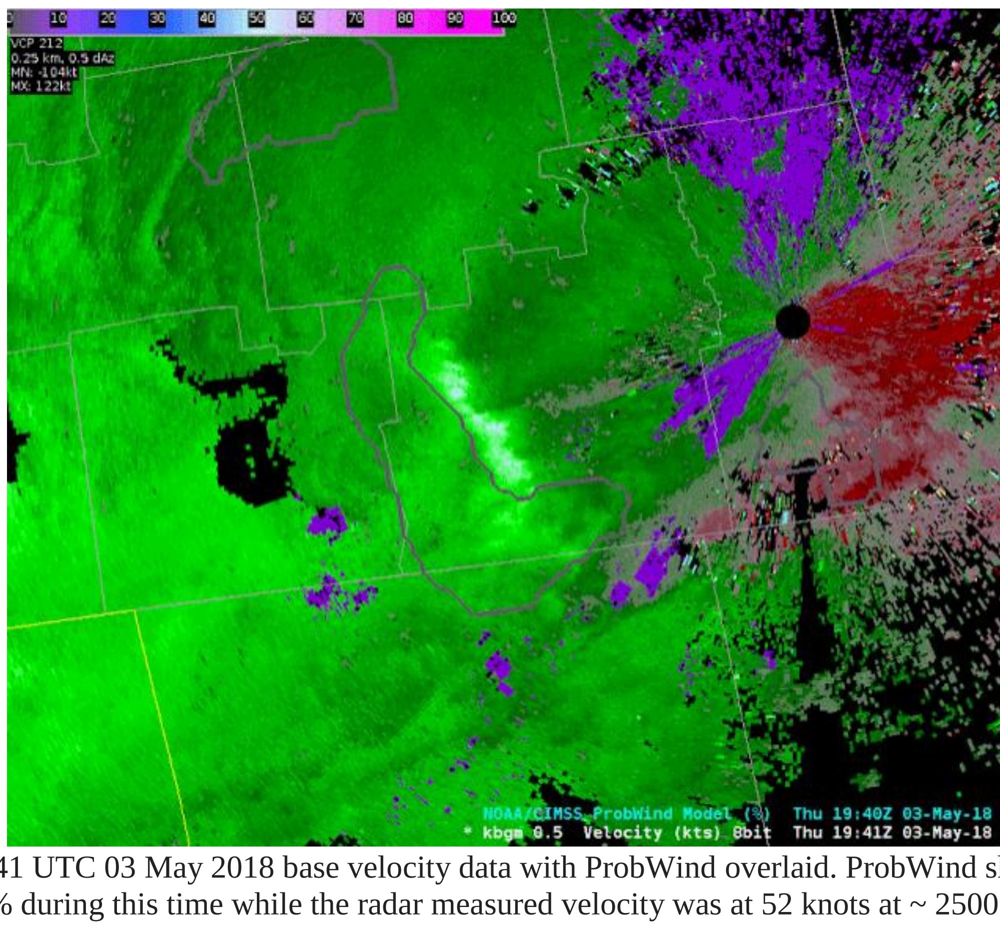

Figure 7: 1941 UTC 03 May 2018 base velocity data with ProbWind overlaid. ProbWind shows a value of 1% during this time while the radar measured velocity was at 52 knots at ~ 2500 feet AGL. The ProbWind outline is noticeable not including the strongest outflow winds in the contour.

"A ProbSevere color curve with more pronounced "breaks" in the data i.e. blue <20%, yellow 30-50%, orange 50-75%, red 75%+ or something similar."
Forecaster, End-of-Week Survey

"I think so. I would like at least the separate ProbTor. However, I would prefer to have all of this information in one display up at all times. Perhaps have the option to include a separate contour (difference color perhaps) around or within the ProbSevere all-hazards contour."
Forecaster, End-of-Day Survey

"Might be useful to display both the all hazards and tornado at the same time to be able to capture the tornado probabilities if they jump up"
Forecaster, Daily Debrief 01 May 2018

"The color scale seems not intuitive where the low end and high end fit both in the cool color range"
Forecaster, Daily Debrief 15 May 2018

"Having more discrete colors instead of a gradient or have increasing line thickness with higher probabilities"
Forecaster, Daily Debrief 15 May 2018

"ProbTor should probably have a different color scale loaded more to the lower end instead of the same as wind and hail.
Forecaster, End-of-Day Survey

3.4 Geostationary Lightning Mapper (GLM) Lightning Detection

University of Oklahoma (OU) /Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), NOAA/National Severe Storms Laboratory (NSSL) and
NASA-Short-term Prediction Research and Transition Center (SPoRT)

For the experiment, a variety of updated GLM products were created based on feedback from the initial review of GLM data in the HWT and Operations Proving Ground in 2017. Initial products included (all at 1-min intervals with 1-min updates): Flash Extent Density (FED), Event Density, Group Extent Density, Average Flash Size, Average Group Size, Total Optical Energy, Flash Centroid Density and Group Centroid Density. Due to the rapid development of these GLM products, forecasters were brought into the HWT without previous training or use of GLM data in this format. Immediate feedback early in the experiment resulted in the creation of 5-min and 2-min summary products (with 1-min updates) to help forecasters better visualize lightning trends over time. Forecasters highly utilized the 5- min Flash Extent Density (with one-min updates) as the primary GLM product. For deeper storm interrogation, storm-electrification understanding, and spatial coverage prediction forecasters also gravitated to the Average Flash Size and Total Optical Energy products at 5 min totals (with one-minute updates). Other products such as the Flash Centroid Density will likely see use in data fusion applications that incorporate flash rates and for data assimilation efforts into convection-allowing models such as the High-Resolution Rapid Refresh (HRRR).

Use of GLM in the HWT

Based on topics from the blogs and discussion, use of GLM data by the forecasters during the experiment operations in the HWT is grouped into five major categories: (1) situational awareness, (2) understanding and monitoring convective evolution, (3) comparison with ground based lightning networks comparisons, (4) lightning safety communication, and (5) training and scientific understanding.

(1) Situational Awareness
Many forecasters chose to utilize the Flash Extent Density either at 2-min or 5-min summations (updated every min) to maintain awareness of storm activity across the area of interest. Forecasters found that Flash Extent Density, particularly when summed across multiple minutes, could provide easily understood guidance on the strongest cells or call attention to quickly intensifying storms cells at a glance. Some forecasters also utilized the flash energy product for this purpose.

> "The GLM provides a good situational awareness whereas the ENI data is somewhat difficult to see given the small size of the data points."
> *Forecaster, 14 May 2018, Blog Post*

> "The main reason this cell stood out was because of a significant increase in Flash Extent Density. This increase really drew my eye to that part of the line where a cell was rapidly intensifying."
> *Forecaster, 15 May 2018, Blog post: "New Cell and Lightning Trend"*

> "The flash extent density and total energy products (two upper panels) both seem to do a pretty good job of highlighting the more active core."
> *Forecaster, 15 May 2018, Blog post: "GLM Lightning with Initiating Cores"*

If warranted by this first glance, forecasters would follow-up with a deeper dive into the GLM products and comparison against ground-based systems such as radar and CG lightning for storms nearing threshold of severe.

(2) Understanding and monitoring convective evolution
Forecasters used the lightning data in tandem with other routinely used convective monitoring tools such as radar, 1-min satellite data, and algorithms blending data such as ProbSevere. Used in this manner, the GLM data often provided more confidence in warning decisions:

> "We issued four severe warnings for northern Indiana and southern Michigan. All storms displayed good dual pol signatures, but we also noticed a sharp uptick in GLM Total Energy and Event Density on the storms we issued for. This increased our overall confidence in warning issuance."
> *Forecaster, 9 May 2018, Blog post: "GLM Upticks before warning issuances"*

> "Little to no increasing trend in GLM Flash Density makes me even more confident in

anticipating little to no development of this convection over southern NM"
Forecaster, 21 May 2018, Blog post: "ABQ - 2300Z Update"

"Storms continue to maintain (if not increase) their strength as their pivot north and northeast through the Billings CWA. Of particular interest is the rapid increase in GLM flash extent density collocated with smaller average flash area and high total energy with a quickly-developing and strengthening updraft. The increase in flash event density also aided in the decision to put out a warning for the storm (particularly given the environment), even as ProbSevere values were below 50% (but increasing). With just ProbSevere or just GLM, or just CTT data to look at, confidence would be lower than when taking in all of the datasets together."
Forecaster, 23 May 2018, Blog post: "GLM Data on Rapidly Developing Storm"

Forecasters continuously stressed the importance of seeing the GLM data across multiple products (specifically Flash Extent Density, Total Energy, and Flash Area) and in context of other data for a better understanding of both the lightning data itself and using it to better estimate convective strength. Group discussions with forecasters commonly mentioned how a single product such as Flash Extent Density could provide some information about a flash and potentially about storm behavior. However, pairing Flash Extent Density with Flash Area allowed forecasters to more easily diagnose convective specifics, for example, identifying long flashes through the stratiform region. Additionally, higher flash rates associated with the smaller flash sizes helped forecasters identify new or intensifying updraft regions within a longer convective lines. This combination of new data provided forecasters additional understanding in how storms were evolving throughout a convective event.

"As expected, an extensive trailing stratiform region developed with the maturing MCS as it moved into the central part of ICT area. One long flash was captured at 2146 UTC. Of note, flash area was in excess of 3000 km2 in the stratiform region. The total energy in the stratiform region was equivalent to that in the updraft region as the large charge reservoir was extinguished. More frequent replenishment of charges in the more turbulent storm-scale updraft compensated for the individually smaller/weaker flashes there (ref McGorman, Bruning). There was a depression in flash area along the line while the group area had more continuous flash sizes along. The larger number of groups helped make this more continuous versus the lower flash count."
Forecaster, 2 May 2018, Blog post: "Trailing Stratiform Region Flash"

"… the GLM Flash Density and Total Energy increased markedly as the merger took place. With all of these updrafts so close together, am also seeing a marked increase in Average Flash Area-especially over Grant/Cherry counties."
Forecaster, 10 May 2018, Blog post: "Supercell Merger in NE using ProbTools/GLM"

"Of particular interest to me was the increase in lightning as the merging between the cells and the line took place, followed by a decrease in lightning as the mesocyclone formed… As the more robust thunderstorm activity moved southeast of the DC area, you could see how by using the NEXRAD data along with the satellite and lightning data the tornado threat might have been on the increase toward the DC area, while a more

significant wind, hail, and heavy rain threat was favoring areas to the south."
Forecaster, 14 May 2018, Blog post: "Tornado Potential near Washington D.C"

Similar to past reviews of total lightning data, forecasters commonly noted how "*GLM data provides a supplemental dataset for monitoring storm activity in the absence of traditional radar interrogation techniques."* Forecasters commented on this during operations across and nearing the Intermountain West where radar coverage was poor or when examining oceanic areas where the aviation community is routinely required to route traffic.

> "One thing that GLM data is going to change for aviation forecasters at the Aviation Weather Center, CWSUs, etc. is Convective SIGMET (C-SIG) and CWA size…As GLM data becomes available to aviation forecasters, I think it is going to open some eyes as to just how much areal extent to lightning there can be in individual t-storms & t-storm clusters/lines/line segments."
> *Forecaster, 23 May 2018, "Aviation t-storm forecasting/warning: GLM Avg Flash Area vs CWA and SIGMET sizes"*

> "Over the Atlantic Ocean, GLM data will eliminate much of the guesswork in determining what is and what is not a thunderstorm. Not only will this help the aviation community, but it will help the Navy as well as shipping, fishing, marine, recreational boating and cruise liner communities, to name a few"
> *Forecaster, 24 May 2018 "using GLM data over the Atlantic Ocean"*

(3) Comparison with Ground-based Networks
With access to multiple lightning networks simultaneously, forecasters frequently examined the differences between these networks throughout the lifecycle of convective storms. Comparisons in timing, location, accuracy, and detection efficiency between these networks were routinely part of discussions and blog posts. Typically, forecasters noted lightning first from the GLM prior to the ground-based networks. Additionally, forecasters found the GLM better represented the spatial extent of the lightning than the ground-based networks.

> "GLM first picked up cloud flashes around 2019Z in Hancock County while the first appreciable CG strikes occurred around 2040Z. GLM shows to be beneficial in highlighting areas of convective initiation before cloud to ground sensors."
> *Forecaster, 8 May 2018, Blog post: "GLM vs Ground Network Lightning detection"*

> "… the GLM Flash Extent Density shows a much larger horizontal extent of lightning vs. the ENI pulses. The GLM average flash and group areas (two lower panels) show average flash area increasing over time as the storm grows, going from light green to medium or dark blue."
> *Forecaster, 15 May 2018, Blog post: "GLM Lightning with Initiating Cores"*

> "…a noticeable increasing lightning trend with ENTLN data wasn't really noticed until 1905-1910Z as well so GLM was able to capture the initial electrification of this storm with few minutes extra lead time. On a day like today when monitoring areas of cumulus for the first convective echoes to develop, GLM (especially GLM total energy) was

> especially useful."
> *Forecaster, 24 May 2018, Blog post: "GLM first identified developing storm in WI"*

However, forecasters often voiced confusion and concern if the flash rates and trends were different between the GLM and ground systems. As the GLM science team continues to better understand why the GLM may not perform the same for all convective environments, it is important that these caveats are communicated to the operational forecasters as well.

> "Over the course of the week I have been concerned with comparing the consistently low FED values (single digit for most storms) with the other lightning fields including ground-based networks. Visually the low counts per pixel do not grab your attention when looking at a busy scene with storms in a variety of stages of their life cycle…"
> *Forecaster, 3 May 2018, Blog post: "FED vs GED"*

> I noticed rather low values of GLM flash extent density & total energy with a severe warned storm near Wheatland in eastern WY so out of curiosity I decided to plot observed CG & cloud flashes from the ENTLN. Despite low values of Flash extent density & total energy the earth based network observed a rather active storm with numerous CG flashes & even more cloud flashes. The total energy product seemed to perform better than flash extent density in conveying the lightning activity in the storm & potential strength of the storm, but if I was only using GLM in a vacuum I likely would have greatly underestimated the potential of this storm. It is becoming more evident to me that the GLM output can differ drastically for "severe" storms depending on the environment, geographic region, & even from storm to storm during the same event.
> *Forecaster, 23 May 2018, Blog post: "Comparison of GLM & ENTLN lightning data"*

Through this comparison process, forecasters often noted the obvious displacement (parallax) of the GLM locations relative to the ground based system. This apparent shift in position away from the actual location was unexpected as forecasters become acclimated to parallax with other satellite data. However, many forecasters suggested that training and best practices remind forecasters specifically to pair the lightning data from satellite to total lightning data from the ground based systems for storm-based warning polygon locations. "The image (Fig. 8) compares cloud lightning flashes from Earth networks with GLM total energy. The GLM image gives a much more complete picture of the extent of lightning, however there is a noticeable north/south discrepancy between the two. The GLM max just southeast of the main line is shifted northward by about 10 miles relative to the ENI lightning cluster"

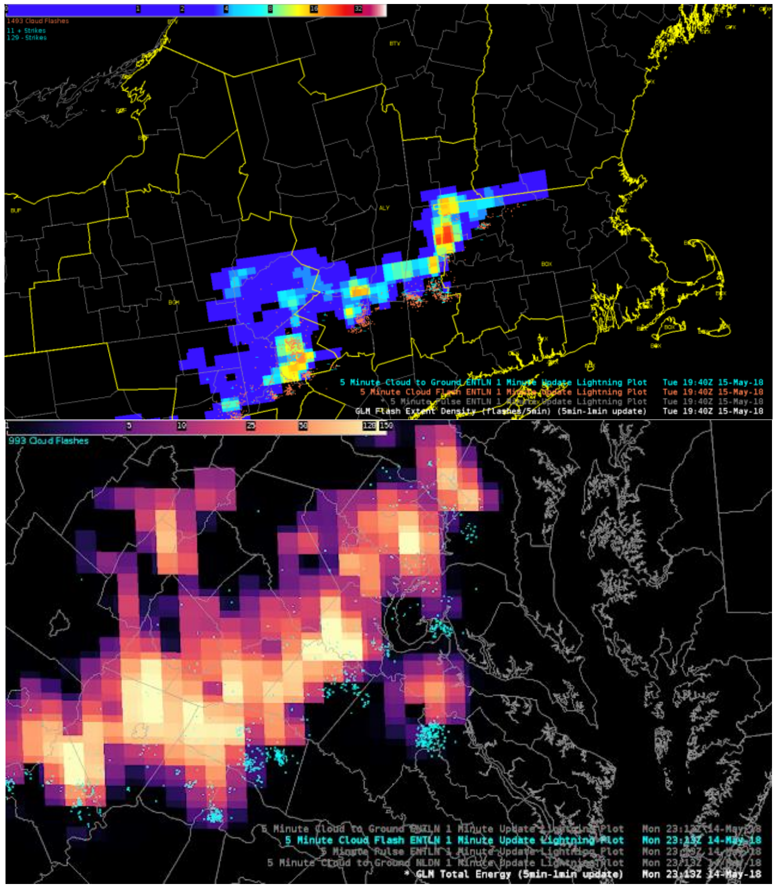

Figure 8: 14 May 2018 Forecaster screenshots from the northeastern United States depicting the parallax errors from GLM in comparison to the Earth Networks total lightning data. Top: 5-min GLM Flash Extent Density and ENTLN CG (blue) and Cloud flashes (orange). Bottom: 5-min GLM total energy (grid) and 5-min ENTLN cloud pulses (blue).

(4) Lightning Safety and Communication

Both the broadcast meteorologists and NWS forecasters discussed how the increased spatial coverage apparent within the GLM data provided an opportunity to connect and communicate with end-users and the public regarding lightning safety. Additionally, as the NWS continues to grow decision-support services, forecasters noted the impact the GLM data could have in providing guidance for outdoor venues, fire coverage, and airport weather warnings.

> "As the flash extended north into MN and another CWA, this information could be used to enhance DSS since the main convective line was still 50-60 mi away. ENI and other commercial lightning networks would not have alerted to the threat of lightning overhead at such a great distance. An event organizer or EM looking at static radar also may not be aware of the increased threat of lightning well ahead of the main storm band."
> *Forecaster, 3 May 2018, Blog post: "Give me Flash Area or give me death…"*

> "I was able to note the 1 minute average flash area spread well out ahead of the precipitation area in front on the storm and also extend in the anvil behind the storm as well. As a broadcast meteorologist, it was a great way to explain to viewers the lightning

threat away from the precipitation."
Forecaster, 14 May 2018, Blog post: "Severe Weather over Virginia/Maryland"

(5) Training, future development and general concerns
The HWT served as an introduction to each of the participants on the use and integration of GLM data into their daily operations. This included completely new and unique visualizations of lightning data, thus discussions frequently addressed training and specifically the need for more hands-on, locally-focused training opportunities. Forecasters routinely stressed the importance of integrating expert knowledge (such as that available throughout their own HWT experience) at their home offices. They found this greatly reduced confusion and frustration in integrating the products into their operational product suite. Additionally, forecasters desired more time and additional guidance to better understand each of the GLM products relative to the meteorology and also to other lightning observing systems (such as Earth Networks). Forecasters suggested that the context for use (such as DSS, warning operations, fire weather, or aviation) was important and this should "drive various need/demand for individual GLM products." Additionally, caveats regarding viewing angle, parallax, and meteorological environments or storm types should be clearly communicated within available training. At the end of their week, most forecasters saw the utility of the data, but were still left with questions regarding the individual products or relationships to meteorological phenomena. Training and future scientific research should continue to strive to answer these questions and provide context to GLM in an operational environment.

"I noted this storm (Fig. 9) in southern Arkansas was indicating a 2 minute Flash Extent Density near 45 which is near the high end of the default color table. The Total Energy registered 240 fJ at this time, which is double the top end of that respective color table. It makes me wonder what all factors contribute to high energy for a particular storm over another storm in a similar environment. What processes are at work in this instance? It's all very fascinating and could have potential benefits in the warning environment -should a correlation become evident."
Forecaster, 16 May 2018, Blog post: "GLM Total Power"

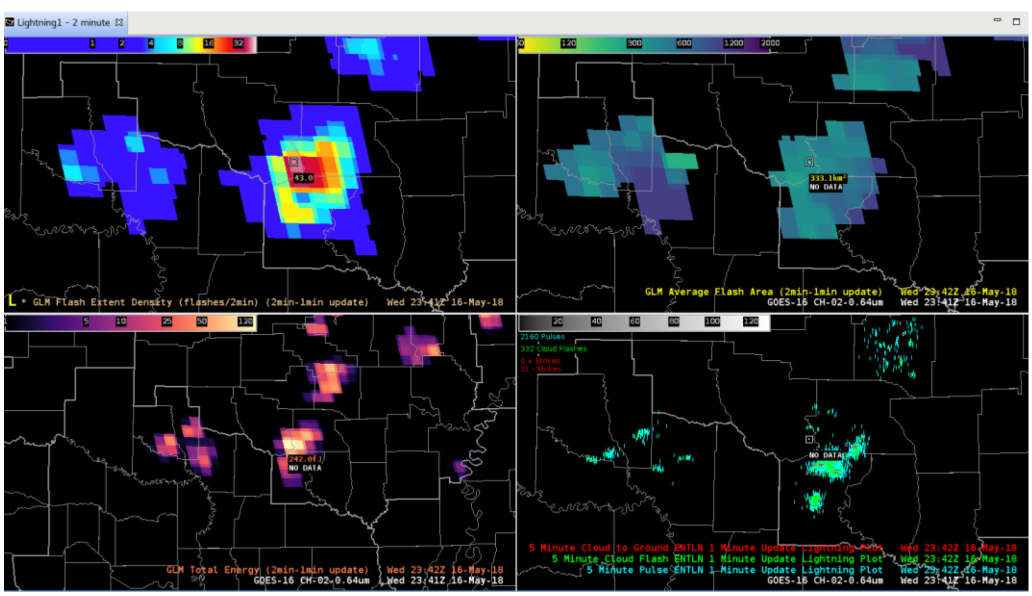

Figure 9: Forecaster screenshot from blog post on 16 May 2018. Top left: 2-min sum of Flash Extent Density. Top right: 2-min Average flash area. Bottom left: 2-min sum of Total Energy. Bottom right: ENTLN pulses (blue), flashes (green) and CG lightning (red).

> "I would like to see more training on the average flash area in how to show how the lightning strike area coincides with the event density and Total Energy. Please do not just do another module, especially for something as new as GLM! It has taken me personally hours of handling the data WITH the presence of the GLM SMEs [subject matter experts] at the EWP to have even a modest understanding of the flash area, total energy, and flash extent. Perhaps average flash area is more of an IDSS application product than use in warning operations."
> *Forecaster, 8 May 2018, Blog post: "FSD Mesoanalysis"*

Results from Daily Surveys

The daily surveys characterize many of the aspects forecasters addressed throughout the blogs and captured during discussion in the HWT. For the GLM, the daily survey included five questions:

(1) Did you find any of the following specific GLM products useful today? [rank each product for today's weather from 'Not at all Useful' to 'Extremely Useful']. Why?
(2) What was your confidence (i.e., your understanding) of each of these GLM products? [rank each from 'None' to 'Very High']. What influenced this?
(3) What update frequency did you use most often or wish you had for today's weather?
(4) What changes do suggest to the visualization/color tables (if any) based on your use today?
(5) Please note any other recommendations you have for improving GLM applications.

Forecasters ranked Flash Extent Density and Flash Energy as the most useful products, but also considered the Flash Area, Event Density, and Group area products as useful overall (Fig. 10, top panel). The centroid products were found considerably less useful by forecasters, averaging as "somewhat" or "not at all useful." Follow-up comments clearly described why FED, Flash Energy and Flash Area had the highest use. Forecasters noted they *"like to see the flash extent data and avg flash area, in conjunction with the total energy. I can conceptualize how intense or widespread a storm is by using all three*." Additionally, they noted at these three products *"helped show strengthening, weakening, and maintaining intensity*" and *"seemed to correlate to what I would estimate updraft strength."* While FED was the most highly used and ranked, some forecasters preferred Flash Energy, noting: *"I did find the total energy the most useful because I can relate to total energy better than the other parameters. It's like thinking about and looking at a 100 Watt light bulb… That's easier for operational forecasters to think about."* However, some forecasters did not utilize the energy product as easily, noting that *"measuring how bright a storm is…isn't necessarily reflective of how strong a storm is."*

Though forecasters initially ranked FED and Energy higher during the week, the usefulness of flash area generally increased as the week continued as forecasters became more familiar with the product and how to use it. For example, one forecaster noted that "the ability to differentiate between small flashes within the core of a developing storm and broad, long flashes throughout the anvil is very nice. This can not only provide some warning decision making tools, but is also a useful product for IDSS purposes. "

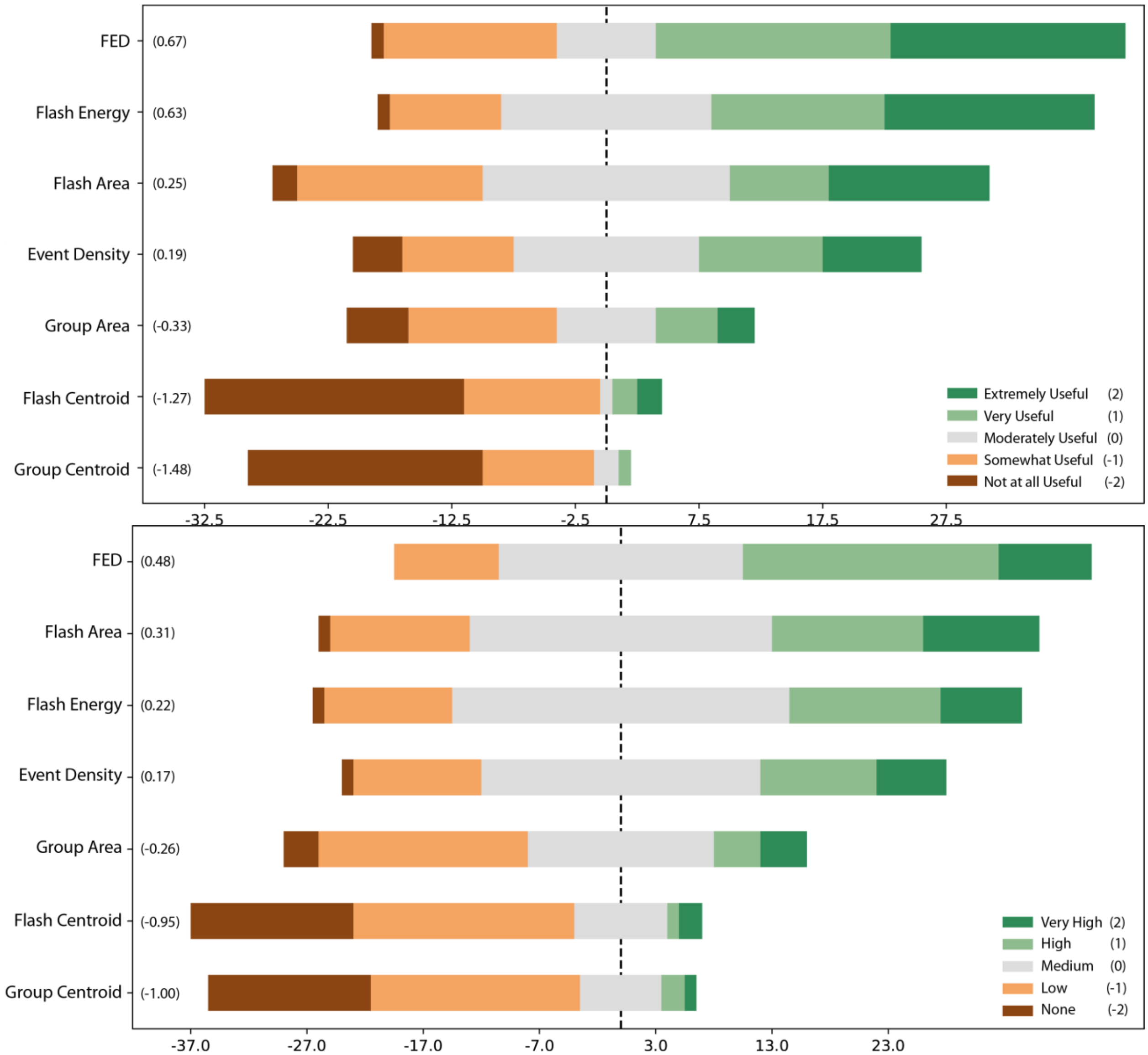

Figure 10: Results from daily surveys. Top: Forecaster opinions of product usefulness from Extremely Useful (Dark Green) to Not at Useful (Dark Brown). Bottom: Forecaster understanding of each product from Very High (Dark Green) to None (Dark Brown). Products are sorted by decreasing average in each plot. Averages were calculated according to points shown in legend and are shown to the right of each product on the y-axis.

As for the centroid products, forecasters believed these looked too much like the ground system detections and cluttered the screen: *"The centroid density looks so similar to CG strikes. I had to remind myself that's not what the product is"* Initially, some forecasters ranked all of the GLM products poorly due to lack of experience and training, noting: *"I do not have a good understanding of what the GLM products are displaying."* Rankings for all products, except the centroids, generally increased as the week continued, primarily due to increased exposure to the products.

Forecaster confidence in or understanding of each of the GLM products similarly increased as the week continued, centroid-based products withstanding. However, very few forecasters ranked their own confidence in any of the products as very high even late in the week (Fig. 3, bottom panel). Averaged across the week, forecasters had a medium confidence in understanding what a majority of the GLM products were showing them. FED and Flash Area were ranked highest, followed closely by Flash Energy and Event Density. Group area and the centroid products averaged a low understanding by the forecasters.

Forecasters found the number one thing influencing their confidence in using the products was hands-on experience working with subject-matter experts. Repeatedly in the comments forecasters noted *"getting thorough explanations from the developers was a big confidence help"* and *"it will take time to get high confidence. It's still lots of information at the cutting edge of science."* Seeing the products in context of multiple events, different environments, and *"comparing it to what was happening and seeing what warnings were issued and the severe weather reports that were received"* greatly influenced the confidence in use. However, if the GLM data did not match trends in the ground-systems or other observations forecasters during an event did lose confidence in the products. For example, "Did not see much correlation between storm intensification trends & the GLM data, but I did notice some in the ENTLN data. This lowered my confidence in the GLM for this event." This problem was compounded when working in regions of high parallax errors or in environments in the west central US (Colorado, West Texas) where charge distributions and lightning height could influence the GLM detection efficiency. Forecasters desired more specifics as to how these things influenced the GLM products, comments included: "*How parallax affects GLM data is a mystery to me. It would seem to degrade it, but by how much?"*

Finally, forecasters struggled with some of the terminology and meaning of the sub products of GLM: events and groups. This complicated the overall understanding of the GLM relative to the meteorological phenomena and other lightning detection systems. Forecasters noted that the *"Flash products are intuitive. I easily grasp the concept of a Flash and the density and areal coverage make perfect sense. The "arbitrariness" of the Event/Group products makes them slightly more difficult to grasp."* Similarly, even when forecasters found that the event/group products were understood, the "somewhat mysterious definitions of event/group reduce confidence in my ability to comprehend exactly how to use the information." The 5-min summation updating every minute was by far the most used product. This timing allowed forecasters to visualize trends, but still receive a rapidly updating product. Forecasters that used the 1-min update accumulations felt the data was too chaotic to sensibly use for situational awareness or visualizing trends. Once the 2-min summation was made available many forecasters gravitated to both that as well as the 5-min product, but continued to list the 5-min as the primary product for severe weather applications. Still, forecasters found value in the 1-min products throughout the experiment as these allowed them to see the greater detailed variation in the storm-scale cores.

As future products are being developed and changes made for the operational implementation, some forecasters also suggested that increased smoothing in the products (beyond the AWIPS 'interpolation') would be desired. Additionally, it was suggested that the GLM data is combined with ground-based systems to help with geolocation errors and detection efficiency.

Finally, forecasters commonly used the final question of the survey to address the need for increased training opportunities prior to and as the GLM data becomes part of the operational data feed. Specifically, they are looking for *"additional training on the products and their applications, points of failure, weaknesses."* This training should address "*how the fields interact with each other & what that suggests the storm is doing lightning-wise (i.e. high total energy but low flash extend density, etc.)"* and should show "*the benefits of using this data in operations and what additional information may be gleaned from it over the surface based systems."* At least one forecaster suggested that discussion with subject-matter experts should become more broadly available to those required to train others (e.g., local office focal points and science operations officers): *"While the hands-on experience is extremely valuable, I gained just as much knowledge about the products from the discussion about what I'm looking at. The brief discussion on why short flashes develop near the core of developing updrafts and longer flashes tend to occur in the anvil areas is something that I may not have grasped learning about the products remotely. It's very important to some people to explain the output and not rely on someone to intuitively understand the theory."*

Final Recommendations for Operational Implementation

Forecasters utilized the 5-min Flash Extent Density (with one-min updates) as the primary GLM product. For deeper storm interrogation, storm-electrification understanding, and spatial coverage prediction forecasters also gravitated to the Average Flash Size and Total Optical Energy products at 5 min totals (with one minute updates). Forecasters found limited-to-no usefulness and understanding of the centroid and group products. Some forecasters did prefer the Event Density Product to the Flash Extent Density due to the increased values and spatial variability, but a majority of forecasters were confused at the difference and uniqueness of one product versus the other. Forecasters also quickly gravitated to the 5-min summation and averaging products to better analyze trends than the 1-min and 2-min products, but still greatly depended on the 1-min updates to these products. **It is therefore recommended that the 5-min (with one-min updates) Flash Extent Density, Average Flash Size, and Total Optical Energy products are in the operational baseline of GLM for the NWS AWIPS.**

Latency of this product greatly introduces decreased and inattentive use. Latency of only 5-min at any given time-step greatly decreases the utility of the product in rapidly changing environments. Latency beyond 10-min resulted in forecasters choosing to ignore the product completely. **It is recommended that latency of the product is consistently less than 2 min for operational use.** *(Note: The code-base for the GLM was updated prior to the final week of the experiment and appeared to adequately address this problem).*

Usefulness of the products was consistently regarded higher than general understanding of all the GLM products, including the Flash Extent Density. **It is therefore recommended that increased training opportunities are provided to forecasters at the time of the operational implementation beyond previous required training modules and quick guides.** Locallyrelevant training utilizing local cases and expertise (such as from lightning and severe weather focal points) should be developed at an office or sub-regional level. Hands-on training should be given the highest priority. In addition to storm-growth and severe-storm-interrogation, training should address context-specific use such as Decision-Support Services and lightning safety, fire weather, aviation warnings and use over radar-sparse regions (where appropriate).

3.5 NOAA Unique Combined Atmospheric Processing System (NUCAPS) Temperature and Moisture Profiles

Joint Polar Satellite System (JPSS)

The NOAA Unique Combined Atmospheric Processing System (NUCAPS) was demonstrated in the HWT in 2018 for the fourth year in a row. The atmospheric temperature and moisture profiles are generated using an algorithm that combines both statistical and physical retrieval methods. NUCAPS combines information from both the Cross-track Infrared Sounder (CrIS) and the Advanced Technology Microwave Sounder (ATMS) instruments aboard the Suomi-NPP polar orbiting satellite to provide soundings as close to the surface as possible. These profiles are produced at NESDIS/NDE and delivered over the AWIPS Satellite Broadcast Network (SBN) for display in the National Skew-T and Hodograph Analysis and Research Program (NSHARP) application in AWIPS-II. During the experiment, swaths of NUCAPS profiles from Suomi-NPP overhead passes were created over the east coast around 1800 UTC, central US around 1930 UTC, and western US around 2100 UTC with a typical latency of one and a half to two hours before the soundings were available for viewing by forecasters in AWIPS, which is not ideal for operational use. Quality control (QC) flags associated with the NUCAPS profiles were also evaluated in AWIPS. These flags allow forecasters to quickly and easily identify which profiles within a swath passed (green) or failed (red/yellow) automated QC checks. The QC procedure just checks that a clean retrieval was obtained from both the infrared and microwave imager (green), just the microwave imager and not the infrared (yellow), or neither provided a clean retrieval (red). These QC flags do not directly determine the accuracy of the sounding and whether the sounding is an accurate representation of the atmosphere.

There were several other products presented for the NUCAPS evaluation in 2018 other than just the soundings. A new data-flow mechanism was tested this year to improve on the extremely high latency seen with the operational version of the soundings from previous year’s feedback. Soundings along with the gridded data and experimental automatically modified soundings were brought via a Direct Broadcast stream from the University of Wisconsin. The effectiveness of these soundings and the impact the reduced latency had on forecaster use of the soundings was evaluated. An experimental version of the NUCAPS profiles was available again this year, with updates, for the Suomi-NPP passes during the afternoon. This version provides a correction in the boundary layer to surface temperature and dew point using nearby surface data. The correction inputs the Real-time Mesoscale Analysis (RTMA) surface observations for the new surface temperature and dew point of the sounding and then creates a new boundary layer to replace the lowest levels of the sounding. The boundary-layer height is determined by using Equation (1) below and then creates a new boundary layer for the existing NUCAPS profile based on these data. Plan-view displays and vertical cross-sections of NUCAPS-derived thermodynamic fields were also available again this year for forecasters to view in AWIPS. Finally, NUCAPS temperature and moisture profiles generated using data from instruments aboard the European MetOp-A and MetOp-B satellites were also made available in AWIPS. Swaths of NUCAPS profiles from MetOp-B were created from passes over the east coast around 1500 UTC, central US around 1630 UTC, and western US around 1800 UTC with MetOp-A soundings created approximately one hour later. This allowed for more sampling of the atmosphere between the typical 1200 UTC and 0000 UTC Universal Rawinsonde Observation Program (RAOB) soundings. The latency of the MetOp soundings was similar to that from Suomi-NPP.

Equation (1): $z_{i+1} = [{z_i}^2 + \frac{2}{\gamma} C_H |V| (\theta_{Skin} - \theta_{Air}) \Delta t]^{\frac{1}{2}}$

z – Height of mixed layer
θ_{Skin} – Potential Temperature of surface skin (GOES-16 11/12 um)
θ_{Air} – Potential temperature of surface air (RTMA)
$|V|$ – Wind Speed (RTMA)
γ – Lapse rate of free atmosphere (NUCAPS)
C_H – Exchange coefficient (constant)

The purpose of the NUCAPS demonstration was to assess the quality and value of NUCAPS data for the severe weather forecast and warning process and to determine suggestions for improvement for readiness in operations.

Use of NUCAPS in the HWT

The key benefit of NUCAPS which has been noted by forecasters every year of the experiment continues to be the ability of the soundings to fill in the gap between the regularly launched morning and afternoon RAOB soundings with another source of observational soundings. The timing of the NUCAPS soundings is in the early to mid-afternoon depending on location across the United States, which is oftentimes just prior to convective initiation. There are latency concerns with the operational NUCAPS soundings which would cause the soundings to not be available until after convection had initiated often rendering the soundings obsolete by that point. An important focus of this year's experiment was to evaluate an alternative delivery mechanism for the Suomi-NPP soundings which would cut the latency time down to less than an hour which is what forecasters have said in previous years would be the ideal time frame to receive the soundings after observation time. Given that the availability of the NUCAPS profiles is in the early afternoon, NUCAPS was primarily used by forecasters to assess the pre-convective environment, as well as the environment ahead of developing storms or MCSs. Forecasters also noted the utility of having observationally based soundings in data-sparse regions where RAOB soundings are not observed. This provided another gap filler for areas void of any sounding data besides that from models. NUCAPS soundings were also compared to the RAP model derived soundings within AWIPS and to SPC Mesoanalysis parameters on the web to determine if the environment was evolving as expected and to compare to the model forecasts to assess the confidence in that model for the day's operations. Forecasters were also able to look at the soundings in the late morning from the European Met-Op satellites, though these soundings were found to not be as good of quality and were not seen as being very useful both with the timing (i.e., too early relative to convective initiation) and with there being a large number of soundings that were of bad quality. Forecasters didn't trust the Met-Op soundings and therefore, they were not used during much of the forecast and warning process. In addition, fields derived from NUCAPS profiles that proved useful to forecasters throughout the experiment included lapse rates, freezing level, and mid-level temperatures. Many of the surface-based parameters were not too helpful, at least from the non-corrected soundings, because of the poorer sampling in the boundary layer of the profiles. This resulted in the modified NUCAPS soundings being the most used throughout the experiment for convective purposes.

> "NUCAPS was used to analyze the pre-convective environment. The NUCAPS sounding provided a decent overview of the environment. The data, however needed some manipulation as it missed a mid-level moist layer completely."
> *Forecaster, End-of-Day Survey*

> "I had a chance to use NUCAPS to monitor the air mass in front of the squall line. I found it to be a useful tool in the forecast process. It identified steep lapse rates and some mid-level drying that contributed to the high wind threat."
> *Forecaster, End-of-Day Survey*

> "Utilized NUCAPS soundings to assess thermal/moisture profile changes outside of radiosonde launch periods and/or to compare with forecast soundings."
> *Forecaster, End-of-Day Survey*

> "NUCAPS was utilized to assess the environment during the afternoon just prior to convective initiation. A special 19z DDC sounding was utilized for comparisons. The NUCAPS matched well, just lacking relevant small details. The modified NUCAPS was not modified, so I had to adjust the surface manually. After that, thermodynamic values matched well with the radiosonde, lending confidence in the product. An idea to validate NUCAPS LCL is to compare with surface ob cloud base data."
> *Forecaster, End-of-Day Survey*

A big evaluation of the NUCAPS soundings for this experiment was the additional data pathway to provide the soundings much faster than the operational data stream. This came with the caveat that to get the data faster via direct broadcast, the fields of view used in the algorithm to make the soundings was less. Therefore, evaluation of the lower-latency soundings was compared to the operational version for any changes and relative accuracy of the soundings. For the most part, the forecasters looked at the low latency version of the soundings for their forecasts and the response was overwhelming that the faster they can get the soundings, the better. It was noted that there did seem to be more yellow and red QC dots via the direct broadcast stream than from the operational version, so there was some impact on the performance of the soundings, but overall getting the data about an hour quicker on average was of higher importance than having a few less "good" soundings to look at. Also of high importance to the forecasters from this evaluation, was having the boundary-layer-modified soundings, which were also delivered via the direct broadcast pathway and often came in shortly after the non-modified soundings (Fig. 11). Forecaster stated that having the boundary-layer modification done automatically was great and saved some time from the forecaster having to do it themselves which is important in fast-paced convective environments. Many forecasters stated that they would rather just look at the modified soundings in a convective environment because the boundary layer is so important and the operational soundings don’t represent it well in most instances. So, even if forecasters had to wait a few extra minutes to receive the modified soundings, they see great value in having those available.

> “The low-latency NUCAPS, when compared with SPC mesoanalysis, was not as much in agreement as were the modified soundings. However, they did not stray too far. Nevertheless, would prefer to use the modified soundings if they are routinely regarded as more representative of the current environment (even if we have to wait for them a bit).”
> *Forecaster, End-of-Day Survey*

> “Having products in as close to real-time as possible is very helpful when assessing a rapidly changing convective environment.”
> *Forecaster, End-of-Day Survey*

> “Low latency is important for rapidly changing conditions that may lead to convective development. In today's use, the modified NUCAPS data was much better in the low levels than the low latency version, however it was delayed in arriving.”
> *Forecaster, End-of-Day Survey*

> “NUCAPS needs to be offered as soon as possible to the forecaster. The faster we can translate that data and incorporate it into our severe forecast mindset, the better the public will be as we will have real data- not model derived top down that is not available anywhere else.”
> *Forecaster, End-of-Day Survey*

> “Having timely data is very important. Latency issues should be resolved ASAP if NUCAPS is to be useful in mesoanalysis.”
> *Forecaster, End-of-Day Survey*

> “The modified soundings were much more realistic than the non-modified version today. The low-level moisture was handled better. The non-modified version was too dry/mixed out, resulting in unrealistically low CAPE values.”
> *Forecaster, End-of-Day Survey*

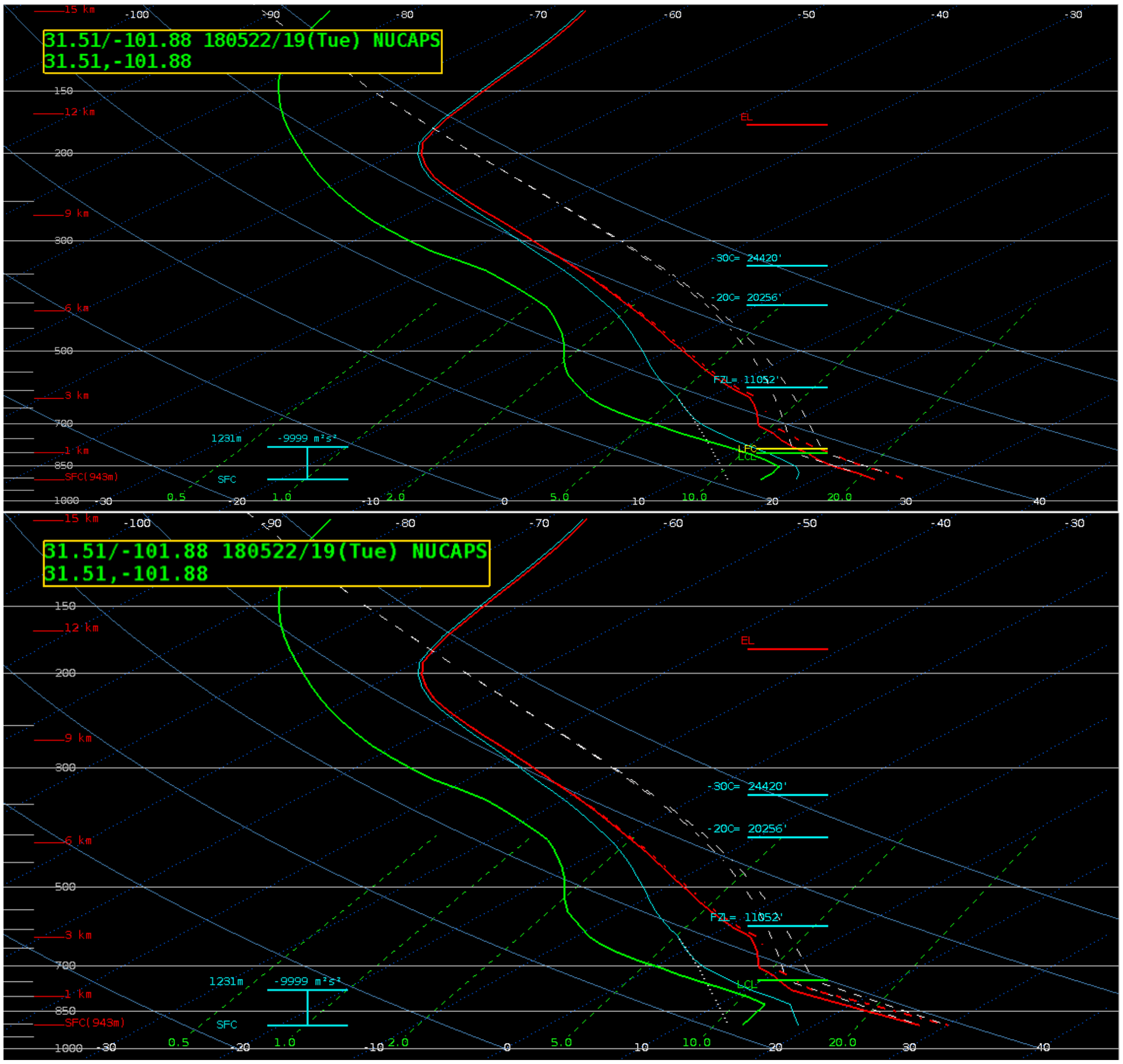

Figure 11: 22 May 2018 19 UTC Unmodified NUCAPS Sounding (top) compared to the Modified Sounding (bottom) in west Texas. Notice the modification in the boundary layer to include the warmer surface temperature and dewpoint.

Another element of the NUCAPS data that was evaluated again this year was the gridded NUCAPS products. This data was found useful at times for looking at derived parameters from the soundings in a plan view or cross section. Often times the gridded products were able to give a quick view of these parameters without having to dig through multiple soundings and piece together an analysis. The gridded data still has its flaws, which will be discussed later with other limitations, but was found to be of use in some instances. The parameter that was most used by forecasters looking at these gridded products was mid-level lapse rates. The lapse rates seemed to match up fairly well with SPC Mesoanalysis and model fields providing confidence in both the NUCAPS data and the model data to be used for the forecast that day. For the case below, a forecasters was looking at the NUCAPS 700 mb – 500 mb lapse rates and comparing them to model derived lapse rates over west Texas and he was interested in how the lapse rates had evolved from his initial analysis earlier in the day (Fig. 12). The following is directly from the forecaster via the blog. "NUCAPS 700-500 mb lapse rates were generally around 6.5-7 C/km over SE NM & W TX. Comparing these lapse rates to the HRRR, GFS, & NAM40 the NUCAPS were the least steep with the GFS & NAM40 especially forecasting 700-500 mb lapse rate values of 7-7.5 C/km. In this case NUCAPS actually compared the best to SPC mesoanalysis values which are what I usually look at in a warning operation environment. This helped add to my confidence in the overall evolution of the event & widespread severe weather still looks unlikely." In this case, the forecaster had more confidence in the NUCAPS lapse rates, as they matched up better with what he regularly used in operations.

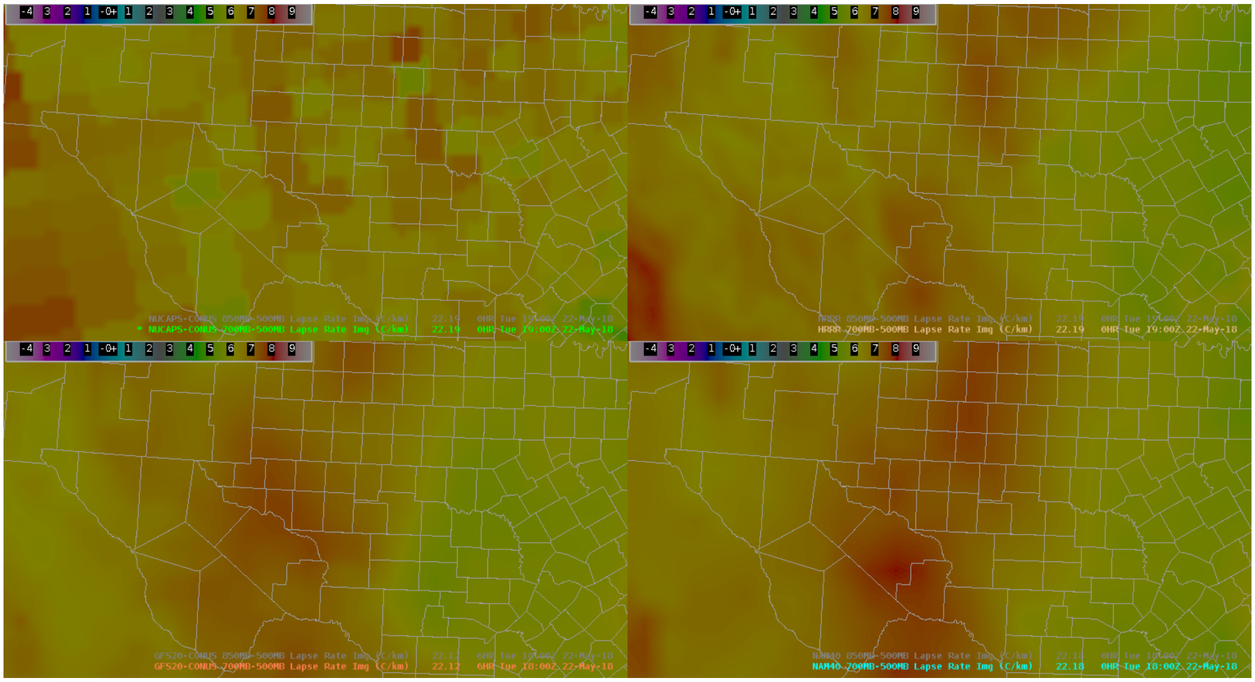

Figure 12: 22 May 2018 1900UTC 700MB – 500MB Lapse Rates from NUCAPS (upper left), HRRR (upper right), GFS (lower left), and NAM (lower right) for the southern High Plains.

Though lapse rates were the main parameter looked at by many of the forecasters as it pertained to the gridded products, some other parameters were also looked at by the forecasters. Some wanted to assess the mid- and low-level temperature gradients looking for capping inversions which might affect the profile. While mid-level temperatures seemed to do okay, the lower levels were a little less reliable due to worse sampling throughout much of the boundary layer. Forecasters also looked at dewpoint temperatures and mixing ratios to assess the amount of moisture available in the atmosphere for forecasting further convective growth and development. Overall, the forecasters enjoy having the gridded data, though there are many issues they would like to see fixed and streamlined before greater use in operations can be accomplished. Those will be discussed in the following section. The gradients of these parameters was found to be most important as often the relative values and gradients would match up while the absolute values would differ from that of other operational platforms.

> "I mainly used gridded NUCAPS to look at mid-level lapse rates. They matched up well with mesoanalysis values over MN but did not depict the EML over the Dakotas."
> *Forecaster, End-of-Day Survey*

> “Looked at Mid-Level lapse rates on gridded NUCAPS & the corrected soundings once they were available.”
> *Forecaster, End-of-Day Survey*

> “Looking at NUCAPS gridded products layered with derived cirrus imagery to help verify gradients and severe concerns”
> *Forecaster, End-of-Day Survey*

> “Used gridded NUCAPS to assess relative mid and low-mid level T and Td.”
> *Forecaster, End-of-Day Survey*

Limitations and Recommendations for Improvement

The running theme with NUCAPS from previous experiments and for this experiment is that the soundings need to be into the forecaster’s operations workspace as quickly as possible after observation time. Most forecasters agree that from time from observation to availability in AWIPS II should be an hour or less. Currently S-NPP NUCAPS soundings take about two hours to get into AWIPS operationally, and this has led to many forecasters not using the data because it is too old to use in convective operations. This latency should improve for NOAA-20 soundings with updated ground processing, but the issue should still be fixed as long as S-NPP soundings are flowing into operations. Forecasters also would like to have more soundings, especially during the early afternoon when it is important to get fast updates of the environment for anticipating convective initiation. It is important to have both S-NPP and NOAA-20 provide soundings to the forecasters around an hour apart in the early afternoon, but only if the latency can be under an hour. When asked if they would use NUCAPS in the future, with 1 being unlikely and 9 being most definitely, 11 of the 13 (85 %) forecasters responded with a 5 or above, but noted that they would respond higher if the latency issue was fixed. Testing the lower latency in this year’s experiment confirmed that idea as the forecasters used the low-latency soundings in their forecast and very rarely even looked at the operational version, which would come in an hour or so later. It was noted however that the delivery method for the lower latency sounding via direct broadcast did have more yellow and red quality control points than the operational version. So, perhaps tweaking that algorithm to more closely match the operational version should also be available in order to provide higher quality data at a faster update time. Still, the forecasters would rather see the data sooner, even with fewer usable sounding points, as the data are not as useful after convective initiation or well after the observation time.

> “Latency!!! Improve the latency. We need it ASAP. Before the convection starts. Right now it's too slow and too late.”
> *Forecaster, End-of-Week Survey*

> “Lower latency is huge for the forecaster when using this product. More lead time increases the chances that we see the profiles prior to potential initiation.”
> *Forecaster, End-of-Day Survey*

> “Timing is everything when using NUCAPS for convection, and should be available as quickly as possible. After 19Z is too late for my forecast area of eastern NC.”

Forecaster, End-of-Day Survey

"Latency still around an hour, which is ok for sounding information. Any longer is not good."
Forecaster, End-of-Day Survey

"That data arrived before the storms initiated, but not by a whole lot. The DB data included more missing sounding locations over the Panhandles, whereas the operational data stream was mostly "green". I was able to modify the data by hand to provide a pretty good look at the environment."
Forecaster, End-of-Day Survey

Forecasters also frequently commented that they liked having the boundary-layer-modification soundings above having the unmodified soundings. A lot of the forecasters would only like the modified version as the only version available, especially in a convective environment, because they have to manually modify the soundings anyway. Most of the time the modified versions would come in just after the unmodified soundings had arrived, but there were delays at times. One recommendation made by forecasters that was implemented during the experiment involved modifying the surface temperature and dewpoint to match the surface observations even if the rest of the algorithm did not run. This adjustmentwas received well and should continue to be implemented. Forecasters still would like to see the data come in faster and to cut out those times when the data would lag behind the unmodified soundings by half an hour.

"If the modified sounding data could be sent as quickly as possible, that would make it more useful to me."
Forecaster, End-of-Week Survey

"Ideally, having the modified soundings become available much sooner after (i.e. low latency) would enhance the usability of the product. Moreover, perhaps there may be ways to improve sensing through thicker cloud layers, which would help increase the number of "good QC-ed data" that becomes available. A potential severe environment is often times already muddled by relatively thick mid or high level clouds which may preclude routine usage of NUCAPS data."
Forecaster, End-of-Week Survey

"Just give out the modified soundings. I only used the green areas that were modified. Point and click is so easy to use. Then I can move on to other analysis. Wind data at a quick glance. Especially in a high shear environment."
Forecaster, End-of-Week Survey

"The modified soundings seemed very close to the data from SPC. I didn't use the NUCAPS until the modified soundings came in. I didn't want to get confused or have bad information in the other soundings. There just wasn't enough time to look through it all and edit either."
Forecaster, End-of-Day Survey

There were some other comments that forecasters had to recommend for improvements to the NUCAPS data to be more useful to operations. For the gridded data, forecasters would like for there to be a more routine way to pull up some of the valuable data. It is very clunky to go through the volume browser, so before implementing this into operations, it is recommended to have a menu structure or some menu options that would be easy to load in fast-paced environments otherwise they would rarely be examined. Also, pertaining to the gridded products, it is recommended to show all of the data and to not black out the yellow or red QCed data. The forecaster can mentally "delete" the bad data, but it is much easier to see gradients when all the data is there to begin with instead of having massive data voids, often making the data unusable. This fix was done for the last couple days of the experiment, and the feedback from the forecasters was overwhelmingly positive. There also needs to be some better parameters for the gridded data such as SBCAPE or MLCAPE instead of CAPE from certain pressure levels. The CAPE values are unusable in the current form and need to be fixed. Some other comments are that forecasters would like to see some wind data in the NUCAPS soundings, whether from GOES or some other type of sensing system just to get a quick glance of the shear and wind at different heights. They are used to having wind data in soundings and it would be a valued addition to get that integrated into the profiles. It is also recommended to have some sort of labeling system for the dots instead of just latitude and longitude, as remembering which dot is which is difficult within AWIPS when going back and forth between the soundings and the point locations.

> "Gridded product hasn't gotten much use due to a large number of clouds in the area throughout the day"
> *Forecaster, Daily Debrief 03 May, 2018*

> "Would be nice to have some wind information in the soundings"
> *Forecaster, Weekly Debrief 04 May, 2018*

> "Gridded products need to be simplified to have options of surface data instead of just pressure levels"
> *Forecaster, Daily Debrief 15 May, 2018*

3.6 All-Sky LAP Stability Indices, Total Precipitable Water, and Layered Precipitable Water Products

University of Wisconsin Cooperative Institute for Meteorological Satellite Studies (CIMSS)

Blended all-sky moisture and stability fields were once again demonstrated in the HWT this year. These fields are derived via a fusion of GOES-16 radiance observations and numerical weather prediction (NWP) forecast data. The project has three components. The first component is the GOES-R Legacy Atmospheric Profile (LAP) retrieval algorithm, a baseline GOES-R product. The LAP algorithm generates retrievals in the "clear-sky" using information from GOES-16 ABI and the Global Forecast System (GFS) NWP model forecasts as a first guess. The second component computes retrievals in some cloudy regions, mainly thin/low clouds, also using information from ABI and a GFS first guess. Finally, the GFS NWP model "fills in" the areas where no retrievals are available from the previous two algorithms due to extensive sufficient cloud cover. Combining these components together provides one blended all-sky product. Fields that are derived from this all-sky algorithm and available to forecasters during the experiment included Total Precipitable Water (TPW), Layer Precipitable Water (LPW) in the SFC-0.9, 0.9-0.7, and 0.7-0.3 atmospheric layers in sigma coordinates, Convective Available Potential Energy (CAPE), Lifted Index (LI), K-Index (KI), Total Totals (TT), and Showalter Index (SI). There was also a data type product which showed what type of retrieval (clear, cloudy, or GFS) provided the value for each pixel. The all-sky products were available every 30 minutes at the top and bottom of the hours and provided full CONUS coverage from GOES-16 ABI. The purpose of this evaluation was to gather feedback for how the algorithm could be improved to better suit forecaster needs and to discover any technical issues with the product.

Use of LAP Products in the HWT

The all-sky LAP fields were primarily utilized at the beginning of the shift each day prior to convective initiation to aid in pre-convective mesoscale analysis. The PW fields were found to be helpful in assessing moisture trends and moisture return into the region of interest. For the most part, the PW fields seemed to match up well with other analysis tools which gave the forecasters confidence in using these products in their forecasts. CAPE and LI were also useful in helping forecasters understand areas with the greatest instability. These parameter values were often looked at in a relative sense, as the absolute values were often found to be inaccurate, while the gradients and boundaries matched up well. Forecasters found it more beneficial to take note of the trends in the updating fields as the rapid increase and moisture and instability often represented regions of future convective development. Furthermore, forecasters commented that the LAP products appeared to accurately detect the location and movement of boundaries and local max/min values which served as a focus area for convective initiation. Forecasters also mentioned that the KI, SI, and TT fields are dated and rarely used in operations anymore and could be replaced with more relevant parameters.

> "I watched the trends and where a tongue of CAPE or moisture was increasing. I used these along with the mesoscale analysis."
> *Forecaster, End-of-Day Survey*

> "I used the LAP products to study the low-level environment and see if any moisture was moving toward the area and decreasing the stability. This showed up best with the CAPE and to a lesser extent with the LI."
> *Forecaster, End-of-Day Survey*

> "I could see the low level moisture advection and increasing instability quite well over the CWA throughout the afternoon. Once the nose of low level moisture and CAPE reached the boundary across the north, storms fired."
> *Forecaster, End-of-Day Survey*

> "It would be helpful to have more of the current combined parameters like SCP, STP, MLCAPE, MUCAPE, etc. rather than Total Totals, K-Index, Showalter. Those latter indices are quite dated."
> *Forecaster, End-of-Day Survey*

> “These products performed quite well, and they showed an increasing trend in PW/Instability parameters through the afternoon. The instability parameters, as in days past, appeared to be a bit low biased.”
> *Forecaster, End-of-Day Survey*

One of the more unique products from the all-sky suite of products not available with the operational LAP dataset was the layer PW fields. When asked at the end of each week if they would like to see the layered PW become operational, the answer was unanimously “yes”. Forecasters were quick to grasp the benefits of looking at PW in different layers for their convective forecasts. The layered PW provides better definition for approaching elevated mixed layers and moisture gradients, along which convection often tends to be tied. Forecasters commented that being able to identify areas of differential moisture advection is key not only to convective forecasting, but also for other phenomena, such as winter storms.

> “Yes. The ability to determine the distribution of moisture throughout the environment is fantastic. It can help identify areas of moisture pooling and dry air aloft.”
> *Forecaster, End-of-Week Survey*

> “Yes. The layer PW provides more useful information than the TPW alone. It is helpful to track the evolution of low level moisture boundaries, as well as the relationship of moisture at the various layers to each other. Drying air moving over moistening air implies increasing convective instability and severe threat, while moist air over moist air implies more of a heavy rain threat.”
> *Forecaster, End-of-Week Survey*

> “Yes, having different layers is valuable. Today I use the lowest layer to monitor the dryline. The mid-level layer would be useful for tracking an elevated mixed layer.”
> *Forecaster, End-of-Day Survey*

> “Yes, it helped locate some mid-level drying in advance of the squall line, enhancing the high wind potential.”
> *Forecaster, End-of-Day Survey*

> “I do find value in the layered precipitable water. Assessing the vertical distribution of moisture in the environment can really help one determine the vertical structure later in the day as mixing commences.”
> *Forecaster, End-of-Day Survey*

The layered PW fields proved to be beneficial in a number of cases while analyzing the pre-convective environment. One such case was in the Des Moines, Iowa, CWA on 03 May 2018. During the early afternoon mesoscale analysis, the forecasters noticed that the two lower layers show relatively high levels of moisture over the area, while the upper level reveals a nose of dryer air advancing toward the CWA (Fig. 13). The forecaster noted that this should increase convective instability throughout the afternoon and that initiation would begin on this tongue of dry mid to upper level air. As the afternoon progressed, storms fired along this gradient in dry air aloft and moved across the Des Moines CWA producing all hazards of severe weather.

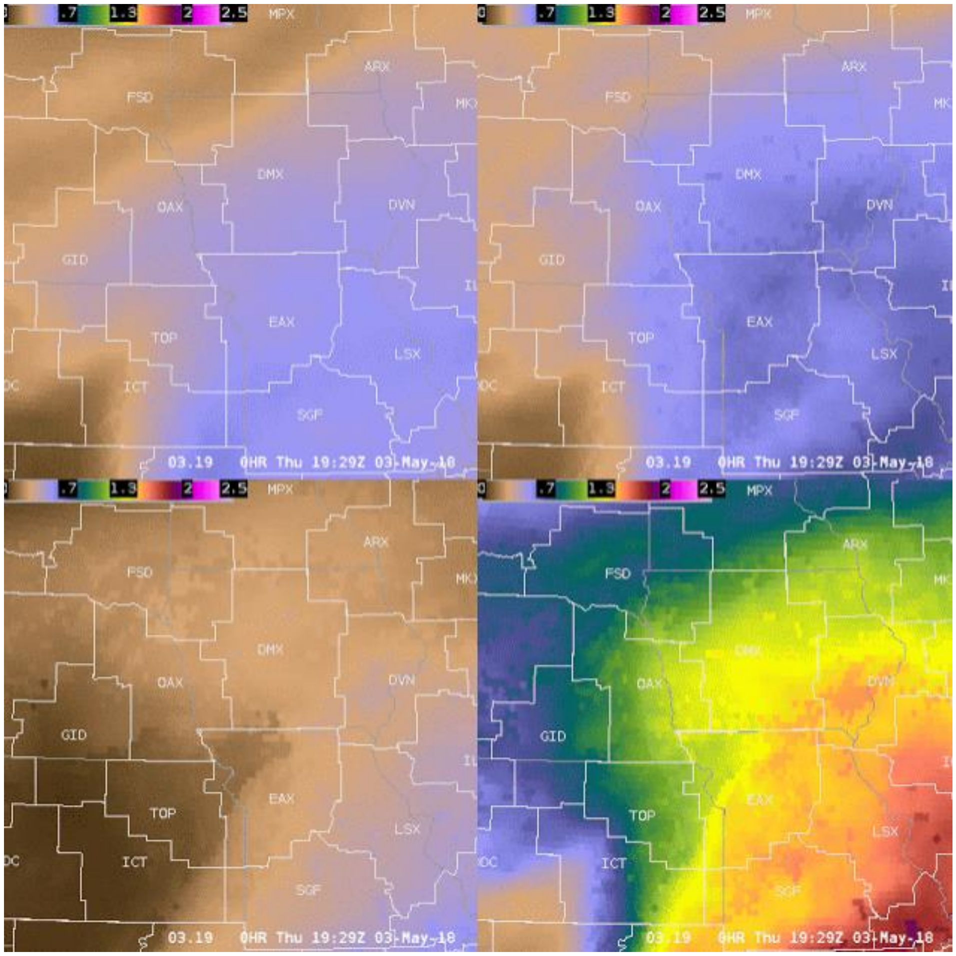

Figure 13: 03 May 2018 1929 UTC LAP layer precipitable water fields. Sfc-0.9 sigma level (upper left), 0.9-0.7 sigma level (upper right), 0.7-0.3 sigma level (lower left), and Total Precipitable Water (lower right) across the Midwest.

Limitations and Recommendations for Improvement

The LAP fields were often used in concert with and compared to other datasets, including radiosondes, model data (e.g., RAP, GFS), SPC Mesoanalysis, and NUCAPS soundings. Forecasters identified some advantages of using the LAP fields over some of these other datasets, such as the higher temporal refresh rate of the products and the relatively low latency. However, there were some drawbacks to the products and some recommendations from forecasters for future product improvements. The forecasters like that the all-sky products fill in the spatial gaps seen from the baseline clear sky products, but there were a few suggestions to improve this algorithm. First, many of the forecasters commented that they would like to see forecasts from a different NWP model than the GFS, one that updates more frequently (e.g. RAP, HRRR), used as a first guess over CONUS. It is thought that this might improve the accuracy of some of the values in the parameters, such as CAPE, as they tended to be low throughout most of the experiment when compared with other observations. This also might help with another problem noted by forecasters, which was that in areas of substantial cloud cover where the GFS forecast was solely used to fill in the gaps there tended to be a sharp delineation between those areas and the areas using the satellite retrievals (Fig. 14). Forecasters would like to see these errors corrected somehow, whether by smoothing the gradient some between those different retrievals to show a more constant gradient or by improving the model fields with use of a different model. Forecasters did like having the retrieval type field and think that that is important to keep so that they can see why the certain transitions are there and which retrieval they are getting for a certain point.

> “I really like them and would love to see all of them be operational. A suggestion would be to use a model other than the GFS that has higher resolution and/or a more frequent updates to fill in the cloudy gaps. This would be especially important when the GFS is suffering convective feedback.”
> *Forecaster, End-of-Week Survey*

> “Perhaps use a model such as the RAP or HRRR over the CONUS for an improved first guess and higher resolution (combined with higher resolution GOES-R instead of 10 km).”
> *Forecaster, End-of-Week Survey*

> “I noticed the CAPE fields were quite different from what the HRRR & mesoanalysis had for analyzed CAPE. Looking at the clear-sky vs all-sky CAPE I realized the GFS was the culprit with extraneously high values over NW & SE MN. Thus I had zero confidence in the CAPE products today.”
> *Forecaster, End-of-Day Survey*

> “The All-Sky Lap fields (particularly CAPE and LI) seemed to be a bit underdone in today's environment (when compared to NUCAPS and SPC mesoanalysis). This seems to be a trend that has been noted for several days now.”
> *Forecaster, End-of-Day Survey*

> “Use HRRR/RAP/or another hourly updating field to fill in the missing data on the all-sky products. GFS data every 6 hours is just too much latency especially when the forecast starts off wrong.”
> *Forecaster, End-of-Day Survey*

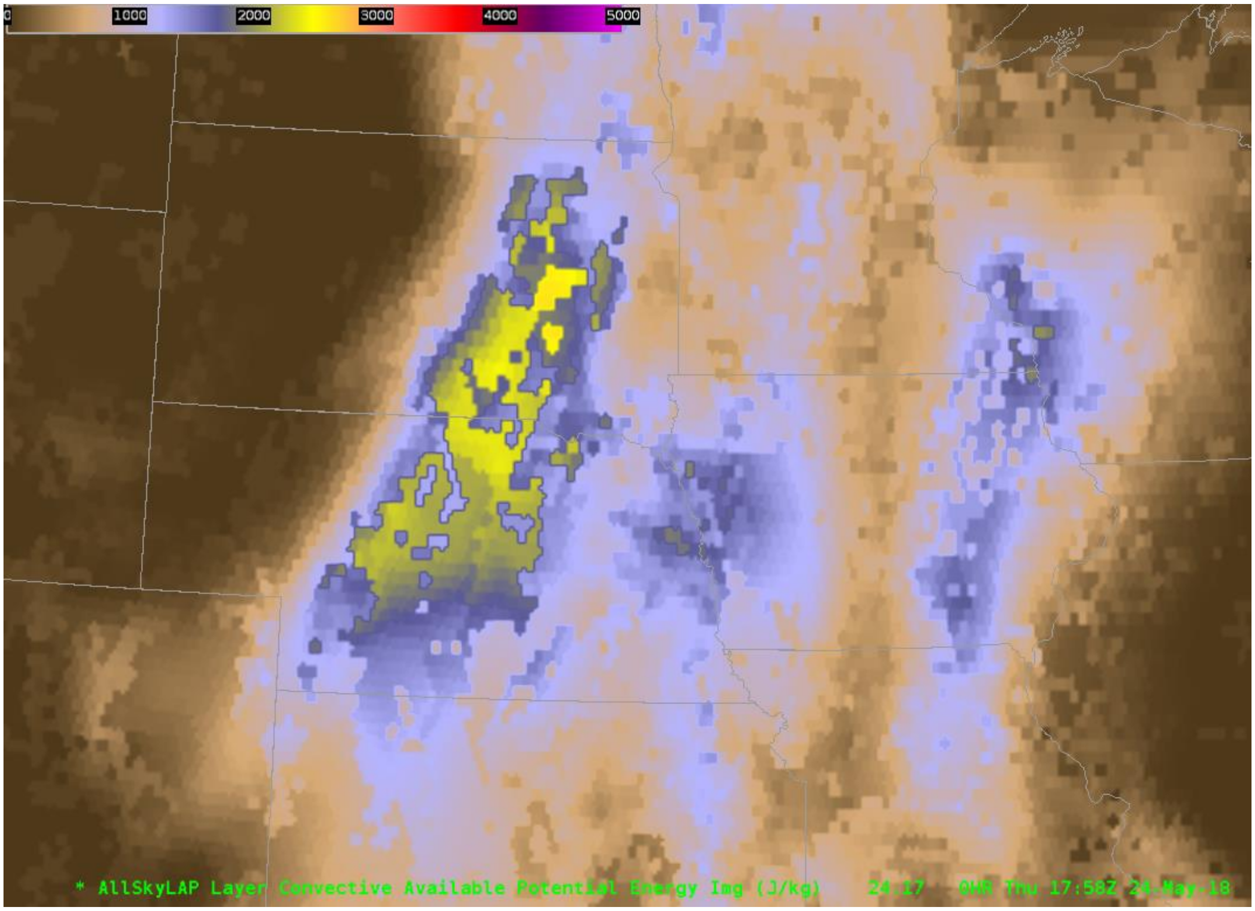

Figure 14: 24 May 2018 1758 UTC All-Sky CAPE over the Central Plains. Notice the sharp gradients and incontinuities over central Nebraska and eastern South Dakota owing to GFS only data to fill in cloudy areas and the clear and cloudy sky retrieval areas.

Forecasters also commented on improving the resolution of the products both spatially and temporally. With the new imagery from GOES-16, forecasters are now used to the new crisp resolution of that imagery and would like for that to be transferred somewhat to these products as well. When asked at the end of each week what spatial resolution they would like to see for these products, of course many asked for the full 2-kilometer resolution to pick out fine gradients, but many also said that 4-6 kilometer would be a good improvement as well, especially over the CONUS. There were many responses that 10 kilometers was still sufficient for pre-convective analysis and the forecasters unanimously didn’t want to sacrifice the timing and latency at the expense of finer resolution. Forecasters were also asked what temporal resolution they would like to see the imagery at and the highest response was 46% wanted 15 minutes, while 5 minutes and 30 minutes got ~23% of the vote each.

> “I was happy with the product at 10 km but if it's possible to go down to 2km resolution while keeping similar latency & temporal resolution I'd like to see that data.”
> *Forecaster, End-of-Week Survey*

> “I think having a 4- or 6-km resolution would be beneficial, with the option to smooth the data to a degree so it doesn't appear so blocky.”

Forecaster, End-of-Week Survey

"The 10 km resolution was sufficient for pre-convective analysis."
Forecaster, End-of-Week Survey

"If they could be produce more frequently that would be helpful."
Forecaster, End-of-Day Survey

"Increased temporal scale (i.e. shorter times between updates) would be valuable."
Forecaster, End-of-Day Survey

"Higher temporal resolution is all I can ask for."
Forecaster, End-of-Day Survey

There were also a few comments about the color tables, especially for the CAPE and the Layered PW fields. Many forecasters thought the color tables could be adjusted a bit on the lower end to make some of the jumps pop. Particularly in the LPW fields where the high end of the scale is rarely reached. Forecasters commented that more dynamic range in these fields would be valuable when looking at the products in operations. The CAPE color table could also use a little adjusting, specifically looking between ~500 – 1000 J/kg, as a lot of the gradient in that range gets washed out. Those might seem like low values, but can be high in some environments and regions and might be all it takes for severe storms.

"It may be more of an AWIPS thing, but I thought the scaling on the color tables could be tweaked so there's more dynamic range."
Forecaster, End-of-Week Survey

"The color table were a little muddy. I would want a more obvious color table, or jump from 500-1000 or so."
Forecaster, End-of-Day Survey

"I'd like to see more dynamic range with the color table for the individual LPW layers."
Forecaster, End-of-Day Survey

"Color table for LPW could be more sensitive."
Forecaster, End-of-Day Survey

Overall, the reception of these products was positive among the forecasters each week. A majority would like to see these products pushed to the field as they see them providing value over the baseline product currently in operations.

"I find the LAP products to be extremely useful and would love to see them pushed to the field."
Forecaster, End-of-Week Survey

> “They definitely provide a more complete picture vs. the clear-sky only, which has a lot of gaps. In the CAPE imagery, you could see clearly some of the areas that were modified from the basic GFS.”
> *Forecaster, End-of-Day Survey*

> “Yes definitely. The cloudy sky retrievals seem to provide comparable values to those from the clear sky. So this is value added.”
> *Forecaster, End-of-Day Survey*

3.7 Convective Initiation and Severe Convective initiation Probability (CI)

University of Alabama in Huntsville (UAH) and
NASA Short-term Prediction Research and Transition Center (SPoRT)

The GOES-R CI and Severe CI products were also back for the 2018 Spring Experiment. The CI product provides the probability of any convective initiation (35 dBZ echo) in the next 0-2 hours and Severe CI gives 0-2 hour probabilities that a given cloud object will develop into severe convection. It does not distinguish among the severe hazards, only that there will be severe convection. The probabilistic products are produced using a logistic regression framework and a training database of approximately 250,000 objects. Due to calibration issues, this training database is still constructed off of previous GOES 13/15 imagery, while the training dataset for GOES-16 imagery gets built up during this convective season. The Severe CI product uses a separate training database and identifies cells associated with severe threats. The CI products use GOES cloud properties and fields from the Rapid Refresh model to generate the forecasts (Mecikalski et al. 2015). The products update every 5 minutes and were available for both the East and West GOES positions. Continued feedback on GOES-R CI and Severe CI performance and best practices in operations was desired.

Use of CI in the HWT

The forecasters viewed the CI and Severe CI product throughout the entire forecast shift each day as an overlay on satellite imagery. Early on in the shift, prior to convective initiation, forecasters monitored the CI product for areas in the cumulus field that would be the first to initiate. Forecasters found that rather than focusing on the probabilities of individual cloud elements, more benefit was gained when viewing general trends in the broader group of probabilities. Also, forecasters seemed to focus on relative values and trends instead of looking for a large grouping of high probabilities within the cumulus field. That information alerted the forecaster that the general area is close to supporting deep convection and to monitor closely for initiation. Just as important as knowing where convection might develop was knowing where convection would not develop, and low probabilities in the CI field enhanced forecaster confidence in some cases that areas would remain free of deep convection.

> “I used the CI in multiple ways. Before during and after convection. The CI gave a great sense of which areas I needed to focus on. And which areas surprisingly had rogue convection.”
> *Forecaster, End-of-Week Survey*

> "I found them useful to identify where convection first initiated over the mountains in SW TX, with higher values over the mountains & much lower values over the plains. However I really did not notice CI prob values get higher than 50% before storms initiated which lowered my confidence in the product somewhat. Ideally I would like to see an increasing trend in an area from low-mid-high values before seeing the first echoes show up."
> *Forecaster, End-of-Week Survey*

> "Yes. It helped pinpoint where to watch for development."
> *Forecaster, End-of-Day Survey*

> "Somewhat, CI worked as a pretty good null case in this event with the capped area over NW MN never having values greater than 20-30%"
> *Forecaster, End-of-Day Survey*

At the end of each week, forecasters were asked how much of an impact the CI and/or the Severe CI product had on their operations for the week. All but one forecaster answered that the products had a small, or very small, impact. Overall feedback on the products in operations this year was low and there were several problems noted by the forecasters and significant improvement need to be made for these products to move into operations. Forecasters noted that the times that the product did work well, was during times when convective initiation was expected. It was often shown that the forecasters could see convective initiation in the one-minute satellite imagery before probabilities on the CI product would increase to a value over 50%. All the forecasters really like the idea of this tool, especially in regions where radar coverage is sparse and more reliance is placed on satellite data to show storm initiation. However, significant improvements need to be made to show the value of the tool in an operational environment.

> "When CI was most valuable, it was predicting convection in areas where we would normally expect it anyway. In predicting new convection, especially with no obvious forcing mechanism (e.g., air mass, post-frontal instability), it didn't do well at all."
> *Forecaster, End-of-Week Survey*

> "In this case, convection did not initiate under the thin cirrus. Referring to IR window imagery, the cirrus were thick enough to block the signal below. However in visible imagery, the cu field and initiation could be seen easily. See blog post"
> *Forecaster, End-of-Day Survey*

Limitations and Recommendations for Improvement

Forecasters noted several deficiencies with the CI and Severe CI product throughout the course of this year's experiment. The number one unanimous problem throughout the four weeks was the inability of the algorithm to show any general trends in the values. The values were consistently on the low end throughout most of the cases and seemed to really jump around between really low and really high values from scan to scan for the same cloud elements. This "jumpy" behavior really decreased confidence in forecasters for using this product. Forecasters would like to see a steadier rise in the probabilities to trust that initiation is imminent and to be monitored more closely. The algorithm also didn’t really pick up on impending initiation very well on most days, even in environments that were prime for storms and it was known storms were about the initiate. The algorithm was particularly bad in environments not as conducive to storm formation, where more “air mass” type pulse storms ruled the day. Overall, forecasters were really underwhelmed by the performance of this product and have said that some pretty major improvements need to take place before this could be operationally relevant.

> “Probs were not consistent from time to time - a lot of apparently random jumpiness. Did not provide lead time for anticipating storm development.
> *Forecaster, End-of-Week Survey*

> “I was underwhelmed by the CI product. It was very noisy and had a lot of spurious high probabilities in areas that weren't conducive for convective development. I didn't have much confidence in the product.”
> *Forecaster, End-of-Week Survey*

> “The CI and Severe CI probabilities, at least for today, did not provide much in the way of substantial value as the data was very jumpy (i.e. going from very low to very high values). These products were disregarded fairly quickly.”
> *Forecaster, End-of-Day Survey*

> “The severe CI was very jumpy and not deemed reliable. There were times the probabilities would go from less than 30% to more than 90%, only to go right back to less than 30%.”
> *Forecaster, End-of-Day Survey*

Another major flaw in the algorithm identified throughout the experiment was the Severe CI probability being higher than the CI probability for the same cloud element (Fig. 15). Forecasters were really taken aback by this, which hindered their confidence in the product even more. Intuitively, the probabilities for severe convection should be lower than the probabilities for convection since you have to have convection before you can have severe convection. Forecasters recommend that a check be put in so that severe probabilities never exceed general convective probabilities for the same cloud element. Besides those times when Severe CI would be higher than CI, generally the Severe CI probabilities were thought to be too low most of time. The severe probabilities would generally be lower than 20% even on days when the environment was ideal for severe convection and multiple severe storms initiated. Some forecaster didn’t think that a severe CI product was necessary, as it didn’t provide significant benefit and oftentimes forecasters are aware of the environment and anticipating that storms will become severe.

> “I think both of these products will require significant improvement before their usefulness is compared. For instance, the severe probability should never be higher than the regular CI, but that was the case numerous times.”
> *Forecaster, End-of-Week Survey*

"I didn't see much use in the severe CI products this week but that may be a result of not having the best environments to look at across the country. I do think a general CI probability product is useful."
Forecaster, End-of-Week Survey

"My problem with the CI versus the severe CI is the disconnect in probabilities. The severe CI should never display a higher probability than the regular CI, but I saw this on numerous occasions. This limits my confidence in using this product in operations."
Forecaster, End-of-Day Survey

"No. Did not trust the severe CI product due to its jumpy probability spikes and decreases. Don't understand why the severe CI can give higher probabilities than the non-severe CI for the same cloud elements."
Forecaster, End-of-Day Survey

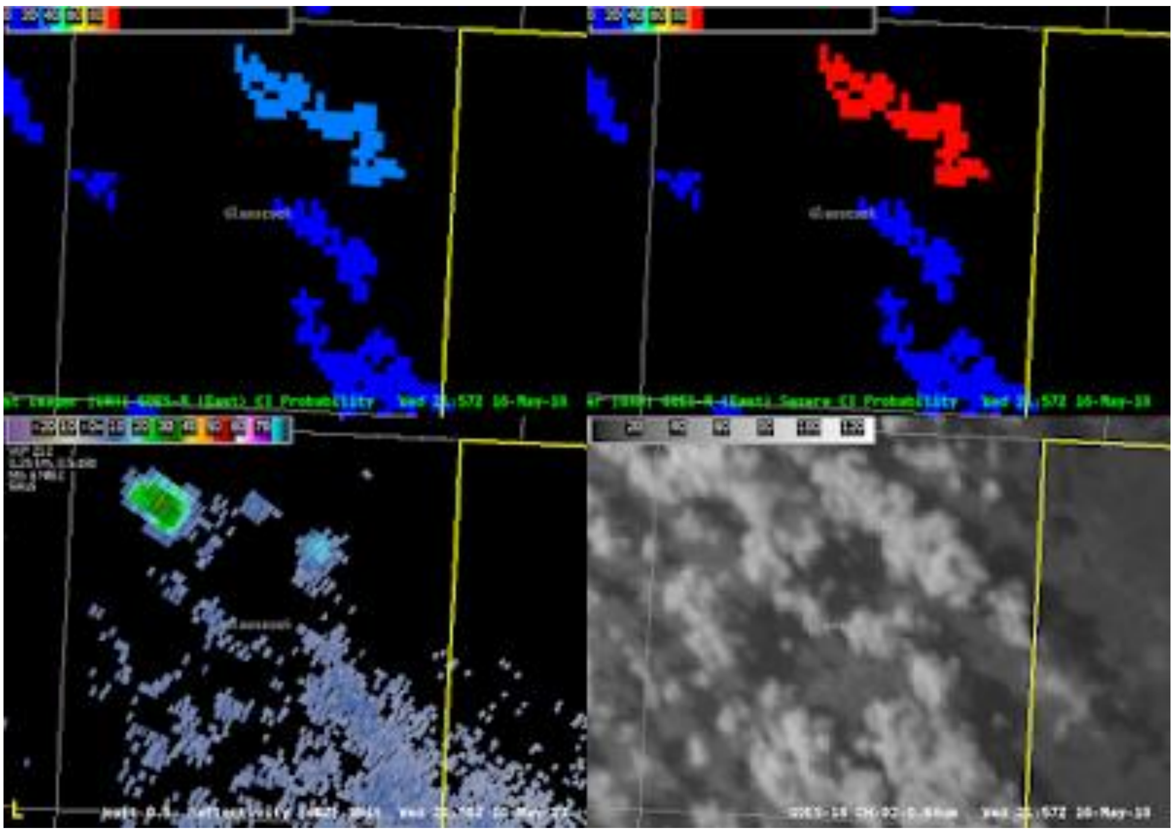

Figure 15: 16 May 2018 2157 UTC CI Probability (upper right), Severe CI Probability (upper right), visible satellite (lower right), and base reflectivity (lower left) showing the really high values of severe CI (>90%) while regular CI shows values much lower (~40%) for the same cloud element.

Another common problem noted throughout the experiment was that the product still needs continued improvement to work well under thin cirrus clouds. It seemed to pick up more cloud object under thin cirrus than in years past, but the algorithm still struggled to identify any real initiation under the upper level clouds. Sometimes initiation was still visible in the satellite imagery but was not picked up by the CI algorithm. This needs continued work to continue to improve the detection in these more complicated scenarios.

> "Yes, I did notice some data under thin cirrus. It was early in the event & the probabilities were low. They also looked more blobby than they do in clear sky."
> *Forecaster, End-of-Day Survey*

> "Yes, and the algorithm struggled at times to track those objects. I think it still needs some work on detecting close to or under cirrus."
> *Forecaster, End-of-Day Survey*

> "Yes, and it had problems tracking or even showing cells that eventually developed quickly."
> *Forecaster, End-of-Day Survey*

Comments on Product Display

Forecasters also provided some comments for improvement of the product display. The common complaint was associated with the color table. The dark blue color used for the low probabilities stood out more versus the higher probabilities shown in lighter blue up to 50%. Most forecasters commented that the lower values, <30%, could be taken out or made more transparent as they didn't really add much value. Also, forecasters would like for there to be some kind of indication that the algorithm indicates initiation has taken place instead of just seeing the value go away. Also, with the addition of GOES-16 mesoscale sectors sampling at one-minute update times, forecasters would like to see this tool follow suit and update every minute when loaded on a mesoscale sector. Many times forecasters commented that they just used the one minute data to look for initiation as it was more useful than the CI tool updating every five minutes.

> "Only showing higher probabilities would be useful (above ~30%). I liked having the data at the resolution provided because it was easy to overlay on visible satellite & really identify the areas of congested cumulus that may be next to initiate."
> *Forecaster, End-of-Week Survey*

> "The color table really needs some attention. Didn't really find the blue color useful as it has such a large range between 10-30%. Perhaps make those probabilities transparent color and then with rising initiation keep the steps you have up to 90%. Also, noticed that the algorithm seems to have too many areas that are just 10%. Not really seeing much value in that other than showing what is a cumulus field."
> *Forecaster, End-of-Day Survey*

Other Comments

> "I would use the CI products on social media. Maybe on TV if it had a better look. I would also use this as a decision making process. When my production crew wants to leave for the night but I'm there babysitting the radar. This extra tidbit will help me

decided if they can leave at midnight or if we need to overlap with the AM crew to break in for a warning."
Forecaster, End-of-Week Survey

"This is a product that would seem to lend itself better to broadcasters, so I would like to see it succeed. But if it doesn't really work, it won't be used."
Forecaster, End-of-Week Survey

4. Summary and Conclusions

The GOES-R and JPSS Proving Ground conducted four weeks of satellite product evaluations during the 2018 Spring Experiment in the Hazardous Weather Testbed. Fourteen NWS forecasters and three broadcast meteorologists evaluated many GOES-R and JPSS products and capabilities, and interacted directly with algorithm developers during the experiment. With GOES-R and JPSS being the sole focus of the demonstration, participants agreed that they had ample opportunity to subjectively evaluate, identify strengths and weaknesses, and suggest potential improvements for all of the products. An abundance of feedback was captured from participants via multiple methods, including daily and weekly surveys, daily and weekly debriefs, real-time blog posts, informal conversations in the HWT, and the "Tales from the Testbed" webinars. This feedback included suggestions for improving the algorithms, ideas for making the displays more effective, best practices for product use, and highlighting specific forecast situations in which the tools worked well and not so well.

Training, in the form of Articulate PowerPoint presentations for each product, was generally well received by participants. They were able to complete the training before arriving in Norman, and felt that it provided them with a basic understanding of each of the products. The only product(s) that forecasters felt less prepared to use were the RGBs, and more training of RGB applications was requested by every forecaster. Based on past feedback, more time was spent at the start of each week as a group going through each of the products in AWIPS. This included a brief refresher about each product, a tutorial on where to load the products in AWIPS, recommendations for pre-built procedures, and caveats. Starting the week with this walkthrough was applauded by participants, and contributed to a smooth start to experimental operations. Similar to last year, an information sheet listing each product under evaluation, its location in AWIPS-II, and contents of notable procedures was created for reference during experimental operations. The pre-built procedures were well appreciated (especially by the broadcast meteorologists) as they facilitated a quick start to operations.

For the fifth year, broadcast meteorologists participated in the Proving Ground Experiment equally with the NWS forecasters. Once again, the inclusion of broadcast meteorologists in the HWT activities went smoothly and proved to be fruitful for all participants. The broadcasters received a unique glimpse into the life of a NWS forecaster during simulated severe weather operations, noting the massive amount of data a forecaster must sift through and the substantial responsibility and stress one feels in such situations. Similarly, the interaction allowed NWS forecasters to gain insight from the broadcast meteorologists on some of their responsibilities, helping to unify the two groups. Broadcasters found at least some utility in all of the products demonstrated, and especially look forward to the high temporal resolution satellite imagery and using the GLM data to communicate lightning threats to the public. AWIPS familiarization at a nearby WFO prior to their arrival in Norman was vital for their successful participation in HWT activities during the week.

Overall, participants enjoyed their experience in the HWT, and felt that the experiment was very well organized. With the emphasis being on baseline satellite products and future capabilities, this activity helps to reinvigorate the use of satellite data in severe warning operations, fostering excitement and increased preparedness for the use of the GOES-R series satellite technology. Participants found at least some utility in all of the satellite products demonstrated, and look forward to using the data more in operations.

More detailed feedback and case examples from the HWT 2017 GOES-R/JPSS Summer Experiment can be found on the GOES-R Proving Ground HWT blog.

Archived weekly "Tales from the Testbed" webinars

Acknowledgements

Multiple NWS and broadcast meteorologists participated in this experiment and provided detailed feedback that has gone into the recommendations in this report, it would not be possible to complete research-to-operations without their willingness to openly play with experimental data. Tiffany Meyer also played a crucial role in the experiment as the one to install AWIPS 2 to handle the experimental data and applied many changes and modifications to the products before and during the experiment. Eric Bruning developed the code base that created the GLM products viewed in the HWT. In addition to Eric and Tiffany, Adrian Campbell also made local changes to allow for 2 minute and 5 minute accumulating GLM data requested by forecasters. The NWS Operations Proving Ground, Scott Rudlosky, NASA-SPORT, Joe Zajic, and Lee Byerle developed the initial AWIPS2 implementation and coordinated data transfer necessary for real-time use of the GLM data in the HWT. The following people also participated as focal points and subject matter experts throughout the week for the products demonstrated in the HWT: Eric Bruning (Texas Tech University), Chris Schultz (NASA-Marshall Space Flight Center), Samantha Edgington (Lockheed-Martin), Clem Tiller (Lockheed-Martin), Geoffrey Stano (NASA-SPORT), Justin Sieglaff (University of Wisconsin - CIMSS), Jack Dostalek (Colorado State University – CIRA), Wayne MacKenzie (NESDIS), Kris White (NWS WFO Huntsville, AL), Nadia Smith (Science and Technology Corporation), Ashley Wheeler (Science and Technology Corporation), Emily Berndt (NASA-SPORT), Chris Vagasky (Vaisala), and Herb Winston (Vaisala).

5. References

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29**, 639–653, doi:10.1175/WAF-D-13-00113.1.

Goodman, S. J., and Coauthors, 2012: The GOES-R Proving Ground: Accelerating user readiness for the next-generation geostationary environmental satellite system. *Bull. Amer. Meteor*. Soc., **93**, 1029–1040, doi:10.1175/BAMS-D-11-00175.1.

Goodman, S. J., R. J. Blakeslee, W. J. Koshak, D. Mach, J. Bailey, D. Buechler, L. Carey, C. Schultz, M. Bateman, E. McCaul Jr., and G. Stano, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res*., **125–126**, 34-49. doi:10.1016/j.atmosres.2013.01.006

Setvák, M., K. Bedka, D. T. Lindsey, A. Sokol, Z. Charvát, J. Šťástka, and P. K. Wang, 2013: A-Train observations of deep convective storm tops. *Atmos. Res.,* **123,** 229–248, doi:10.1016/j.atmosres.2012.06.020.

Mecikalski, J. R., J. K. Williams, C. P. Jewett, D. Ahijevych, A. LeRoy, and J. R. Walker, 2015: Probabilistic 0–1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data, *J. Appl. Meteor. Climatol.*, **54**, 1039–1059. doi:10.1175/JAMC-D-14-0129.1