GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2021 Spring Experiment Final Evaluation

Project Title: GOES-R and JPSS Proving Ground Demonstration at the 2021 Spring Experiment – Experimental Warning Program (EWP)

Organization: NOAA Hazardous Weather Testbed (HWT)

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

Duration of Evaluation: 1 June 2021 – 17 June 2021

Prepared By: Kevin Thiel (OU/CIMMS and NOAA/SPC) and Kristin Calhoun (NOAA/NSSL)

Submitted Date: 26 August 2021

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 2021 Spring Experiment, which took place virtually at the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) in Norman, Oklahoma from 1 June to 17 June 2021. The Satellite Proving Ground activities were focused in the Experimental Warning Program (EWP). A total of 17 National Weather Service (NWS) forecasters representing five NWS regions participated in the EWP experiment. They evaluated four 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. Additionally, they used cloud-based instances of the second-generation Advanced Weather Interactive Processing System (AWIPS-II), through Amazon Web Services, and web-based interfaces to interact with the products. Forecaster feedback during the evaluation was collected through daily surveys, weekly surveys, daily debriefs, weekly debriefs, blog posts, and informal conversations in the HWT. 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. Most of the products evaluated in 2021 were advancements of previous product iterations from the 2019 GOES-R/JPSS Proving Ground. This included data from the Geostationary Lightning Mapper (GLM), the Probability of Severe (ProbSevere) model – Version 3, and the NOAA Unique Combined Atmospheric Processing Systems (NUCAPS). The Optical Flow Winds product, created from the GOES-R series ABI, was evaluated in the HWT for the first time.

Demonstrated Product	Category
Geostationary Lightning Mapper	GOES-R Baseline & National Weather Service
NUCAPS Temperature and Moisture Profiles	JPSS Baseline
Optical Flow Winds	GOES-R Risk Reduction
ProbSevere Version 3	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 Baseline – Products funded through the JPSS program.

Table 1: List of GOES-R and JPSS products demonstrated within the HWT/EWP 2019 Spring Experiment

Several visiting scientists attended the EWP over the three 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), NOAA's National Severe Storms Laboratory (NSSL), National Aeronautics and Space Administration's (NASA) Short-term Prediction Research and Transition Center (SPoRT), the Science and Technology Corporation (STC), NOAA's National Environmental Satellite, Data, and Information Service (NESDIS), and the NWS. The Storm Prediction Center (SPC) and HWT Satellite Liaison Kevin Thiel (OU/CIMMS and NOAA/SPC) and NOAA/NSSL Research Scientist Kristin Calhoun provided overall project management and subject matter expertise for the Satellite Proving Ground efforts in the HWT.

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 NWS operations. 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).

Due to the ongoing COVID-19 Pandemic, all 2021 GOES-R/JPSS Proving Ground activities were conducted in a virtual environment during the weeks of June 1, June 7, and June 14. Five to six NWS forecasters volunteered each week to evaluate this year's products. Before the testbed user guides, PowerPoint presentations, and online learning modules were shared with all participants for each of the products demonstrated through Google Drive. The first day of the experiment began with one hour of introductions and product summaries from developers, with the second hour devoted to setting up forecasters with their cloud-based AWIPS instances. For these activities all participants were in a single video conference. After a brief forecast discussion, forecasters were placed into operations. All subsequent days began with a discussion of the previous day's operations involving questions from developers and feedback from forecasters, followed by a forecast discussion, operations, and daily surveys. End of week surveys were then sent to participants the Friday of each week. Each day began at 1 pm CT and ended at 6pm CT, and forecasters spent approximately four hours in operations. The condensed and static schedule of this year's experiment differs from previous years to accommodate the virtual format of the 2021 experiment, while still offering sufficient opportunities in operations for analysis in pre-convective and post-convective initiation environments.

Operations began each day following the daily forecast discussion, where two to three NWS WFOs were selected to maximize the probability of severe thunderstorms and therefore the utility of the four experimental products. Forecasters were then divided into groups for each simulated WFO, with all groups in their own video conference. Organizers and product developers could then move between the different conference rooms to interact with the forecasters at each simulated WFO. If severe convective activity in a County Warning Area (CWA) ceased or was no longer expected to occur, the Satellite Liaison would transfer the group of forecasters to a CWA with more active convection expected or ongoing. During days when convection was expected to be sub-severe, forecasters were able to create mock-Decision Support Service (DSS) events to show how the experimental products could be utilized in another major NWS function. Forecasters viewed the

GLM, NUCAPS, and ProbSevere data in the cloud-based instances of AWIPS, and the Optical Flow Winds product was available in a web-based interface.

Within operations forecasters had several tasks, such as building procedures to integrate experimental products with the ones they currently use, issuing warnings and advisories, having discussions with the subject matter experts, and writing blog posts. Discussions between forecasters and developers often involved questions from both groups concerning best display practices and applications, along with feedback from forecasters of what they were observing in real-time. Forecasters also had the opportunity to create blog posts by filling out a template through Google Drive. This year saw an increase in the number of collaborative blog posts, allowing for greater depth regarding forecaster experiences with the products and their applications in various operational scenarios. The co-PIs would then use the templates to create blog posts for the HWT EWP Blog (https://inside.nssl.noaa.gov/ewp/), publishing them the next day. At the conclusion of the experiment, all blog posts were added to the Satellite Proving Ground Blog (http://goesrhwt.blogspot.com/) as well for redundancy and consistency with previous experiments. Feedback from the 2021 GOES-R/JPSS Proving Ground from the end of day surveys, end of week surveys, blog posts, and daily debrief discussions are summarized in this report. Product recommendations are listed at the end of each section as ‘recommended’, ‘strongly recommended’ and ‘highly recommended’ in an ascending order of significance from the forecasters.

3. Products Evaluated

3.1 Geostationary Lightning Mapper (GLM)

Several gridded GLM (Goodman et al. 2013; Bruning et al. 2019) products were again tested including Flash Extent Density (FED), Minimum Flash Area (MFA), and Total Optical Energy (TOE), each with one- or five-minute summations updating every minute. NWS forecasters can currently view the FED product in their home offices from GOES-16 and GOES-17. GLM training materials on these products were made available through NWS vLab and Google Drive prior to the virtual testbed. A new addition to the 2021 testbed was the GLM Flash Points product, which provided the parallax-corrected centers from the measured GLM flashes and was available every 1 minute with generally little latency. Flash meta-data (including flash area and duration) recorded by the GLM could be accessed using the mouseover option on GLM Flash Points in AWIPS.

On the first day of the testbed, forecasters were briefed on the GLM, parallax correction, best display practices, and its differences with ground-based networks. Discussion on these topics continued throughout the week as the participants became more comfortable with the GLM products and their performance in different convective scenarios. These location differences helped highlight variations in GLM performance and data characteristics. Overall, the greatest confidence and utility from forecasters was exhibited in the FED and MFA products, followed by the GLM points and TOE, and lastly the GLM flash point data. Feedback about these products regarding applications, the utility of GLM flash point meta-data, display modifications, parallax correction, and training are addressed below.

Use of GLM in the HWT

Consistently, the greatest confidence and utility was shown in the GLM FED product, which was likely due to its current availability in AWIPS across the NWS. When asked to rate their confidence using the GLM products, 69% (39/56) of respondents answered with the two highest categories 'High' or 'Very High' for FED. In comparison the second highest rate was achieved by MFA with 50% (28/56) answering the two highest categories. These results were consistent with responses from the same question in the 2019 experiment. Often FED and MFA were noted for their ability to isolate and monitor the stronger convective cores across a CWA. TOE was often displayed within the participant's procedures, though forecasters appeared uncertain as to how it added value to a forecast. This was reflected in the daily surveys with a lower confidence in the product when compared to FED and MFA.

Trends in the gridded GLM products, when paired with the ABI, ground-based lightning networks, and radar, were frequently used by forecasters to determine strengthening/weakening updrafts. The participants naturally gravitated to FED for monitoring trends, but often AWIPS displays included MFA or TOE for additional confirmation. Limitations of the GLM to monitor lightning trends were also discussed by subject matter experts, concerning the potential for missed detections of small flashes near the updraft of an optically think cloud or closer to the edge of the field of view. This was on display as we operated in North and South Dakota where detection efficiency of the GLM is reduced, and differences can arise between the GLMs on GOES-16 and GOES-17. These points are shown in the following survey responses:

> ‘FED is still what I find most useful, but I gained an appreciation for the MFA today. Being able to better isolate core growths is huge, especially in the CI phase or during multicell convection.’
> *Forecaster – End of Day Survey*

> ‘The lightning data helped us triage storms in Florida. There were multiple storm clusters all going on at once but the lightning data, specifically the FED helped us pick out the cells/clusters that warranted more attention.’
> *Forecaster – End of Day Survey*

> ‘GLM data was not as helpful for ND [North Dakota] today, it seemed the FED didn't change much across the different parts of the storm, and I know that is a known issue with that part of the country. The MFA did still seem helpful to me, and showed areas of persistent updrafts.’
> *Forecaster – End of Day Survey*

The ability of the GLM gridded products to represent the spatial footprint of lightning flashes was frequently used for DSS scenarios. Forecasters favored overlaying FED and MFA with the ground-based lightning networks for a more complete view of a thunderstorm’s lightning characteristics. When inspecting these large flash areas with MFA, or high flash rates with FED, forecasters would often ask at what magnitude these values become ‘significant’. The significance of large values can vary considerably depending upon the location within the CONUS scene and the near-storm environment. These questions hinted that local GLM experts and case studies within a forecaster’s region may increase their understanding of the gridded products and accelerate their use in local NWS offices.

Discussions concerning the connection between CG flashes reported from the ground-based systems and large flash areas observed in the MFA product highlighted the value added of the GLM to DSS. This is especially true for mesoscale convective systems, where horizontally expansive CG flashes may originate in the leading convective region and propagate through layered regions of charge in the stratiform precipitation region before coming to ground. The following blog posts highlight how gridded GLM products can aid in decision support services.

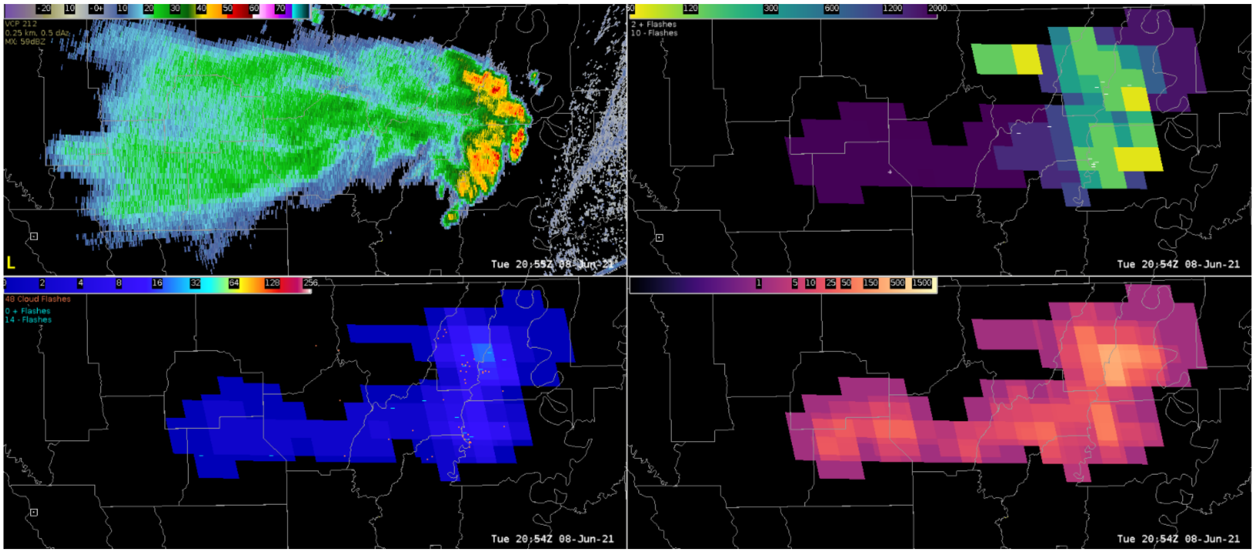

Figure 1: 'A four panel with reflectivity (top left), GLM flash area (top right), GLM flash extent (bottom left) and GLM optical energy (bottom right). The ground networks have been added to the flash area with CG strokes and then over the flash extent with polarity and cloud flashes.'

> 'The flash area (top right of the four panel) [Figure 1] shows the darker purple color extending westward, which indicates the storm mode is more of that light stratiform rain with longer flashes extending through it rather than the intense small flashes within the leading line. This can be helpful in time when you may have a DSS event and the main line has passed through, but lightning is still present in the trailing light rain. Pairing the ground networks with the GLM extent and area allows a forecaster to give DSS on the latest CG stroke within the large area.'
> 9 June 2021, Blog Post: *Analysis of a Line of Storms Moving Across Northeastern LA into Western MS*
> https://inside.nssl.noaa.gov/ewp/2021/06/09/analysis-of-a-line-of-storms-moving-across-northeastern-la-into-western-ms/
>
> 'Based on…the rapid collapse of electrical activity within the shower around 2110 UTC, a reasonably confident "all-clear" could have been given to the venue at that time…or at least until the upstream line approaches in a couple of hours, assuming it holds together'
> 9 June 2021, Blog Post: *Lightning DSS for the Mississippi Pickle Fest*
> https://inside.nssl.noaa.gov/ewp/2021/06/09/lightning-dss-for-the-mississippi-pickle-fest/

Flash points from the GLM and their associated meta-data received mixed feedback from forecasters based upon discussions and survey feedback. Their purpose was to provide a data display similar to the legacy ground-based networks, with the addition of the meta-data to still give the additional data collected by the sensor. When rating the utility of GLM products on a scale of 1 to 5 ('Not at all Useful' to 'Extremely Useful'), the GLM points had the second highest average score of 3.39, trailing only FED (3.83) and just ahead of MFA (3.37). In contrast, the meta-data associated with the GLM flash points had the lowest average score of 2.50. During training on the first day of each week, it was discussed that the GLM flash points were located by determining the optically weighted center of the flash and how this differs from the ground-based networks, even though the product output may look similar. Confidence in the GLM flash points likely came from its similarity to legacy products, as one forecaster commented in a daily survey ‘It’s comfort food; more like what we’re used to seeing’.

Because the GLM flash points were parallax-corrected, this sparked consistent discussion and feedback on the value of parallax-corrected GLM products and how they can be utilized operationally. When overlaid with the gridded GLM products (not parallax-corrected), it was noted that this allowed for a more direct adjustment for parallax. This was especially true when interrogating thunderstorms in regions like the Northern and Central Plains regions, where viewing angle is lower, parallax is higher, and due to reduced detection efficiency and overlap in the field of view GLM sensors from both GOES-16 and GOES-17 are encouraged. FED maxima and MFA minima near convective cores could be visually correlated with clustered GLM flash points. On the other hand, it was noted that the flash points could not be used by themselves due to the loss of the areal coverage of the flashes observed in the gridded GLM products. The following quotes from forecasters highlight these observations:

> ‘I can definitely see the improvement in the parallax-correction data with the Flash Points when compared with the ENTLN flashes, so it is an easy way to adjust your eyes when you are looking at the gridded data.’
> *Forecaster – End of Day Survey*

> ‘The flash points didn't tell me anything that the FED wasn't already showing me. The FED was also easier to quickly see the areal coverage of the lightning.’
> *Forecaster – End of Day Survey*

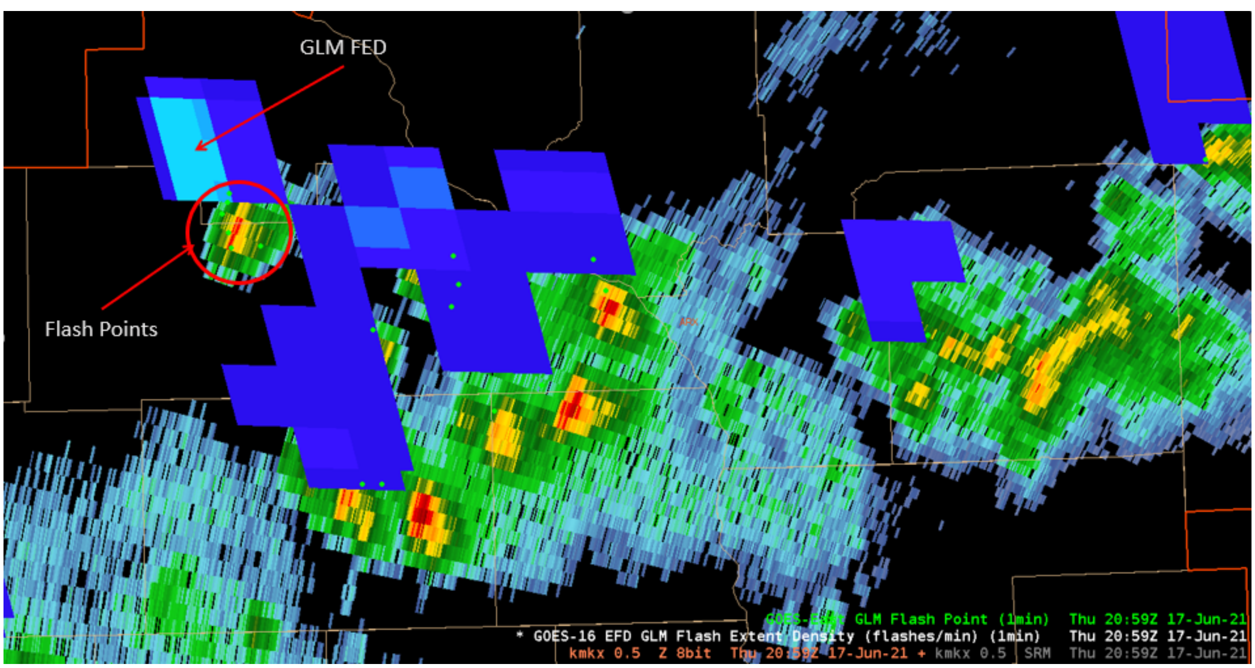

Figure 2: GLM flash points overlaid on FED and base reflectivity from KMKX. The FED maxima and its corresponding flash point cluster are highlighted.

> 'GLM flash points really help speed up the process of identifying where the cell of interest was located [Figure 2]. In the past, I would have to make a manual, on the fly "calculation" in my head where the actual cell was located.'
> 18 June 2021 Blog Post, *GLM Flash Points*
> https://inside.nssl.noaa.gov/ewp/2021/06/18/glm-flash-points/

In terms of visualization, one consistent recommendation from the forecasters about the GLM points themselves was that five-minute composites would be more helpful to include when comparing these products with the ground-based networks in terms of lightning flash trends. Additionally, a counter of the total GLM points in the AWIPS display window was also noted as a needed addition. Meta-data from the GLM flashes were generally not seen as useful due to redundancy with other products such as MFA, though a few forecasters did note that from a science or quality assurance perspective they may add some value. To increase the value of the meta-data, color coding the points based upon flash area or duration were suggested by forecasters as potential ways to include this data in a more consolidated format.

When asked if the participants preferred the GLM data parallax corrected or non-parallax corrected (matching the ABI grids) in group discussions and surveys, responses were mixed. Some forecasters strongly preferred one over the other, while others had no preference. However, one point of emphasis stated by several forecasters on both sides was the need for consistency amongst the GLM products to reduce location uncertainty surrounding the differences in parallax. Quotes from the surveys below help to reflect the variety of thoughts regarding the topic of parallax correction:

> 'I like the parallax correction. I could go either way on correcting/not correcting, but would like it to be consistent across the suite of GLM products.'
> *Forecaster – End of Day Survey*

> 'Right now we want GLM on the native grid. We've been using it for years and forecasters have mentally corrected for the parallax error.'
> *Forecaster – End of Week Survey*

> 'Parallax corrected projects lined up very well with the radar signatures and made a lot of sense. I would like to see GLM products corrected for parallax in the future.'
> *Forecaster – End of Day Survey*

Lastly, each week of the testbed it was suggested that a storm-relative time series tool, like the one developed for the ProbSevere Version 3 product (see section 3.2), be developed for forecasters to better monitor trends in GLM flash rates. Forecasters expressed the need to quickly identify periods of rapid intensification and cessation. Some even mentioned that having the ability to quickly compare GLM flash rates to the ground-based networks would aid in identifying lightning jumps and increasing forecaster confidence in warning decisions. This idea was mentioned at least four times during our daily forecaster discussions, seven times in the daily surveys, and four times in the weekly surveys.

Recommendations for Operational Implementation

Based upon the evaluation of GLM products in the 2021 HWT Satellite Proving Ground, the following items have been recommended:

- It is recommended that the NWS continue to emphasize training forecasters on GLM-related products, often through subject matter experts within their local NWS offices. Topics may include the advantages and limitations of GLM data, intercomparisons with ground-based networks, parallax, case studies in decision support service or warning operations, and gridded product usage through an understanding of the relationships between lightning flash rate, size, and brightness.
- It is strongly recommended that the gridded GLM Flash Extent Density product stay in the baseline (Level2) product, and that the Minimum Flash Area product be added. Further exploration regarding how Total Optical Energy can be used effectively in NWS operations is also advised.
- It is highly recommended that a storm-based time series display of GLM flash rates, and potentially flash rates from other ground-based lightning networks, be developed for use in AWIPS. This may potentially utilize the storm-tracking already available with the ProbSevere v3 Model.
- It is recommended that the utility of the GLM flash point product and its associated meta-data continue to be explored. Suggested modifications to its display include the addition of five-minute composites, an on-screen total for GLM flashes similar to the ground-based networks, and that the colors of the points be made dependent on the associated flash duration or area.

3.2 Probability of Severe (ProbSevere) Model

The NOAA/CIMSS Probability of Severe (ProbSevere) Model (Cintineo et al. 2014) was evaluated for a sixth consecutive HWT, with the 2021 experiment as the first to include Version 3 (v3) of the model. ProbSevere is a statistical model that provides probabilistic guidance regarding the occurrence of severe weather within the next 60 minutes, along with its associated hazard models for severe hail (ProbHail), severe wind (ProbWind), and tornadoes (ProbTor). Version 2.0 (v2) of the ProbSevere Model (Cintineo et al. 2020a) became operational at NCEP Central Operations on 14 October 2020, with several fundamental changes made to v3 to improve model calibration (Cintineo et al. 2020b). ProbSevere v3 leverages specific ENTLN, GOES-16 GLM, GOES-16 ABI, Multi-Radar/Multi-Sensor (MRMS), and SPC Mesoanalysis data into new machine-learning models for each hazard type, producing output in the form of polygons centered around each storm. The probability of each hazard type within the storm is then assigned to the individual polygon. With day/night availability across the CONUS, ProbSevere was designed to highlight storms that are more or less likely to produce severe weather, and to increase forecaster confidence along with lead times in warning decisions.

Forecasters were able to display ProbSevere and individual hazard data in AWIPS in the form of storm-relative contours that update every two minutes, with the color of each contour correlating to their probabilities. Contours could then be sampled to provide additional information. The first data provided form sampling were the probabilities from ProbSevere v3 across all hazards, with comparison outputs of v2 to assist forecasters in calibrating themselves to the new probabilities. Often v3 outputs are lower than those in v2 to improve forecast reliability. Individual values of the meteorological parameters used to produce the probabilities were also included when sampling the contours in AWIPS. Because more variables have been introduced into v3 than can practically be displayed, the readout only shows the variables also used in v2. When displaying the ProbSevere contours, a second outer contour may appear when ProbTor exceeds a predefined threshold, as a way for forecasters to quickly identify which storms to make decisions on regrading severe thunderstorm and tornado warnings. A new addition to ProbSevere in the HWT was the ability to display probabilities of each object as a time series. By double-clicking the contour, a time series window would appear for that object to provide trends to the forecaster across all predicted hazards. Time series plots of the model predictors themselves, along with v2 and v3 output, could also be accessed through a web-based display. Forecasters were provided with a one-page summary regarding ProbSevere v3 and its associated hazard models prior to the testbed, along with a video that further discusses the model, its differences from v2, and example applications. Overall forecasters found ProbSevere v3 was an improvement from ProbSevere v2, and the new time series tool received overwhelmingly positive feedback. Use of ProbSevere v3 in the 2021 Satellite Proving Ground and feedback regarding product reliability, performance when compared to ProbSevere v2, display practices, and leveraging the time series feature are discussed in the following section.

Use of ProbSevere in the HWT

Forecasters used the ProbSevere products in a similar manner to those in previous iterations of the testbed. Contours were most often overlaid with radar data including individual WSR-88Ds and MRMS, which correlated with the storm objects as the identification algorithm is reflectivity based. ProbSevere was also used with other experimental products in the HWT including the GLM and Optical Flow Winds, to assess how well trends in ProbSevere probabilities correlated with trends in theses products. Almost all forecasters (16/17) indicated in the weekly survey that they used the ProbSevere overlay, with only two or three forecasters using the individual hazard contours for ProbHail, ProbWind, and ProbTor. When asked in the daily surveys, 58% (33/57) of forecasters responded that ProbSevere increased their confidence in deciding whether to issue a severe thunderstorm warning, with 11% (6/57) responding it did not increase confidence. The other 32% (18/57) of responses were 'NA', which was most likely influenced by the number of marginally severe scenarios during the 2021 testbed where warning decisions were rare. Of the forecasters who made warning decisions, 68% (17/25) indicated that ProbSevere helped increase warning lead time.

While the applications of ProbSevere were overall similar, noticeable improvements from ProbSevere v2 to ProbSevere v3 were observed by the participants. When participants were asked to compare the daily performance of ProbSevere v2 and ProbSevere v3, 64% (34/53) responded that v3 was either 'Slightly better' or 'Much better' than v2. Only 9% (5/53) responded that they were the same, 4% (2/53) that v2 was 'Slightly better', and none responded that v2 was 'Much better' than v3. These results were supported by the weekly surveys where 81% (13/16) of responses preferred v3 over v2 and their WFO, 19% (3/16) unsure, and none preferring v2 over v3. In daily discussions forecasters noted that v3 appeared more consistent and less 'jumpy' with its outputs compared to v2, so trends in the probabilities could be quickly identified. Additionally, v3 was commented as more consistent and accurate in sub-severe convection, with noticeable improvements for severe thunderstorm winds (most directly through ProbWind) when compared with v2. The forecasters who were more familiar with ProbSevere expressed that the v2 probabilities listed next to the v3 probabilities was helpful for quickly calibrating themselves between both models. The following blog posts highlight the differences between v2 and v3 in various scenarios.

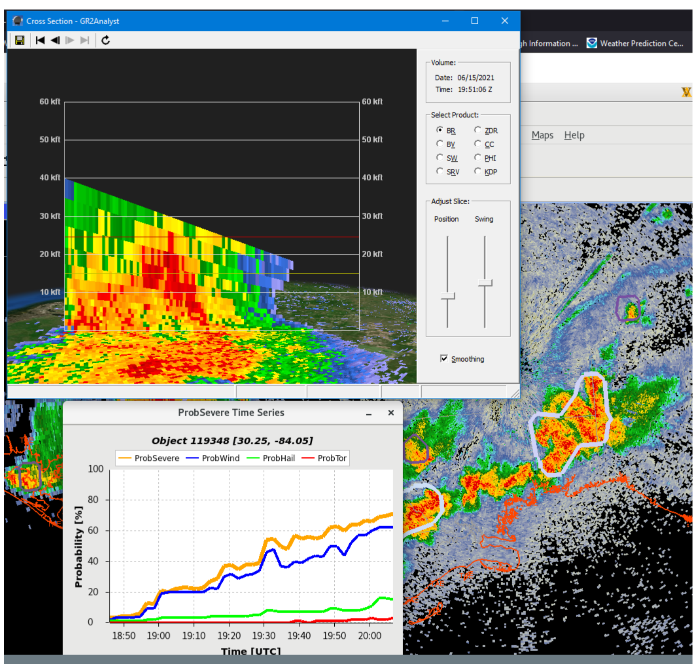

Figure 3: ProbSevere v3 contours (back panel) in the Florida Peninsula, with the ProbSevere Time Series (lower left window) and a cross section of the same storm in GR2Analyst (upper left window).

> 'In this case, the v2 showed a 65% chance for severe hail but the v3 only showed 7% for severe hail. Severe wind was 52% and 58% respectively. Looking at the cross section and knowing the PW values are in the 1.7-1.8 range shows the main threat would be more of a wind/rain rather than hail [Figure 3]. This is a big improvement. To be honest, the issue with over forecasting hail on the v2 is a big reason why I usually don't use prob severe. Seeing this change with V3, I am much more likely to be looking at it as it seems to be more refined and takes the climate, area and conditions into account before producing significant hail or wind values.'
> *17 June 2021 Blog Post: Prob Severe v3*
> https://inside.nssl.noaa.gov/ewp/2021/06/17/prob-severe-v-3/

> 'In this event, the prob severe there was a sig wx statement and severe thunderstorm warning put out by the Huntsville office. Around that time, the prob severe was increased specifically for the prob severe wind component. The version 2 had a prob severe value of 3% while the version 3 had a 53%. Version 3 better captured the significance of the storm with a 40 mph gust reported around the same time.'
> *4 June 2021, Sampling of Sub-Severe Convection Across the Southeast*
> https://inside.nssl.noaa.gov/ewp/2021/06/04/sampling-of-sub-severe-convectoin-across-the-southeast/

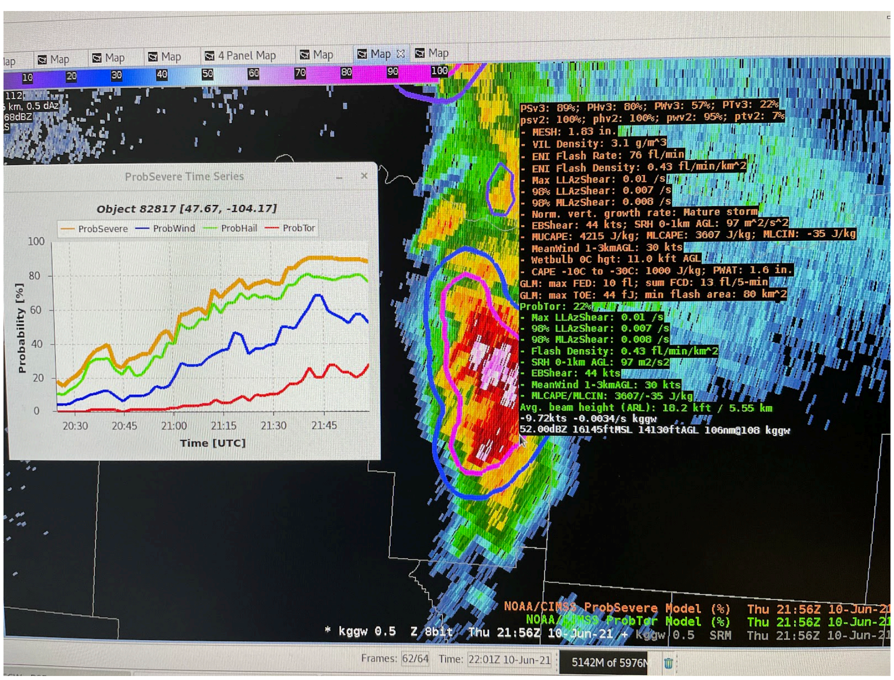

Figure 4: ProbSevere contours overlaid on a radar display from KGGW on 10 June 2021. Probabilities and model inputs are shown for ProbSevere and ProbTor in the mouseover readout, with the ProbSevere time series for the storm shown in the left window.

> ‘This cell northeast of Glasgow [Figure 4] showed some of the highest probabilities of the week for both algorithms, including a ProbTor of 22% according to the newer algorithm, and 7% according to the older one…In this instance, ProbSevere3 and 2 did a great job. Since ProbSevere3 is more conservative than 2, it is worth the time for forecasters to compare the 2 algorithms side-by-side once they are both available. This will help them calibrate their thinking as to what amounts usually produce certain amounts of wind damage, hail size, and tornadic activity. For example, just working with ProbSevere3 a few days, I know that a 22% tornado probability and 57% wind threat is very high, at least for supercells in the northern Plains in June.’
> *15 June 2021 Blog Post: Storms Near Glasgow MT on Thursday*
> https://inside.nssl.noaa.gov/ewp/2021/06/15/storms-near-glasgow-mt-on-thursday/

Feedback regarding the new ProbSevere time series tool was overwhelmingly positive, with forecasters quickly finding the information provided useful in thunderstorm interrogation and warning decisions. Participants were frequently observed in operations using the new time series tool to identify trends and monitor severe convection. In the daily surveys, 93% (39/42) of responses indicated that the ProbSevere time series tool was useful in their warning decision process that day. In daily discussions forecasters consistently cited the utility to diagnose trends using time series, along with using the web-based display to examine what meteorological variables that were driving the probabilities for each product. When asked in the weekly surveys how often the forecasters would use the ProbSevere time series tool, 38% (6/16) responded with the highest possible answer of ‘All of the time’. The next two highest categories ‘Most of the time’ and ‘Occasionally’ received the remaining votes at 31% (5/16) each, while the lowest two categories ‘Seldom’ and ‘Never’ received zero responses. Quotes regarding the use of the ProbSevere time series tool are highlighted below.

> ‘I would prioritize the time series tool. Monitoring the trends of a few high priority storms would be highly valuable.’
> *Forecaster – End of Week Survey*

> ‘Being able to see any type of data in a time series with prob severe data is very helpful in the warning environment. There can be so much going on in the office during severe weather that keep up with the trends of various storms is almost impossible. Being able to see trends in various data fields would be very helpful in staying situationally aware of what is going on with the storm activity.’
> *Forecaster – End of Week Survey*

> ‘I found the ProbSevere time series helpful today as we “triaged” storms and tried to identify storms that may become severe. While the capping inversion stayed strong and therefore prevented storms from becoming severe, it was great to see storms follow a similar intensification process identified by the Prob time series time graph.’
> *18 June 2021 Blog Post: ProbSevere Time Series*
> https://inside.nssl.noaa.gov/ewp/2021/06/18/probsevere-time-series/

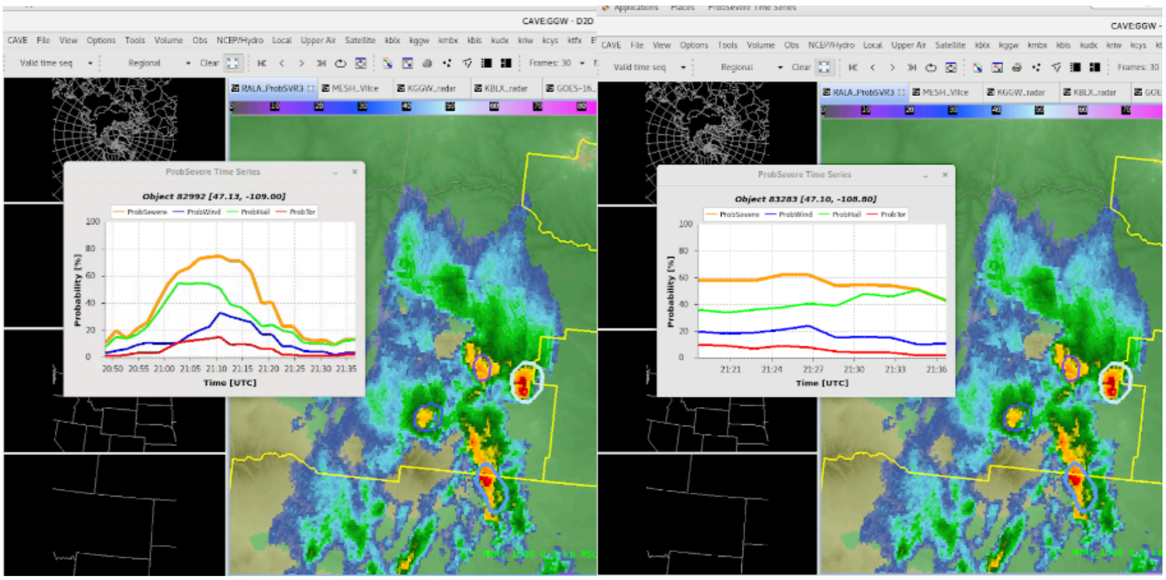

Figure 5: 'MRMS Reflectivity At Lowest Altitude and ProbSevere3 time-series for the left-splitting (left) and right-splitting (right) storms. Within a couple of minutes of the split, ProbSevere3 correctly predicted that the right storm would become dominant and generally maintain its intensity. Meanwhile, the left-splitter would quickly weaken.'

> 'Suppose you were thrown into radar duties without time for a full-on environmental analysis. In an environment conducive to splitting cells, ProbSevere3 can quickly provide guidance allowing the warning forecaster to anticipate which cell will become dominant.'
> *15 June 2021 Blog Post: ProbSevere v3 and Splitting Cells Over Montana on 10 Jun 2021.*
> https://inside.nssl.noaa.gov/ewp/2021/06/15/probsevere-v3-and-splitting-cells-over-montana-on-10-jun-2021/

A majority of the suggested improvements to ProbSevere included expanded utility of the time series product, along with display elements such as the contours. Forecasters were asked to rate five research and development efforts in terms of importance. These efforts included improving object identification, projection/extrapolation of threat probabilities, ProbTor, transparency of ProbSevere, and the ProbSevere time series tool with model predictors. Four of the five average ratings (1 is most important, 5 is least important) were between 2.56 and 2.94, overall generating inconclusive results about which was deemed most important. The lowest average score went to improving the transparency of ProbSevere (3.69) potentially indicating that forecasters were able to understand ProbSevere output from a combination of previous experience, training prior to the experiment, and use throughout the week. Additional comments regarding improvements to ProbSevere by the forecasters are included below.

> 'Improve visualization to show brighter colors at lower values since v3 doesn't seem to max out as high as v2. Ensure very little to no latency to MRMS data and ProbSevere data arriving at WFO. Maybe have option to not contour really low values (<20%) to better distinguish which storms actually need to be monitored.'
> *Forecaster – End of Week Survey*

‘Being able to easily visualize which storm the time series is for, either by bolding the contour or making it flash.’
Forecaster – End of Week Survey

‘Prob Severe Table Ideas: If the tables could open in a floating tab in Awips that would be very helpful. This way you can manually dock and move the tab around in awips. This way you can quickly view the tab and keep it open until you want to just close it. You can also rename the tab to whatever will help you keep track of the storm that it belongs to. This would get around needing to color code multiple tables etc… A separate tab will also allow room to show additional information (beyond just prob severe).’
4 June 2021 Blog Post: Northern VA/Maryland Convective Episode on 6/3/2021
https://inside.nssl.noaa.gov/ewp/2021/06/04/northern-va-maryland-convective-episode-on-6-3-2021/

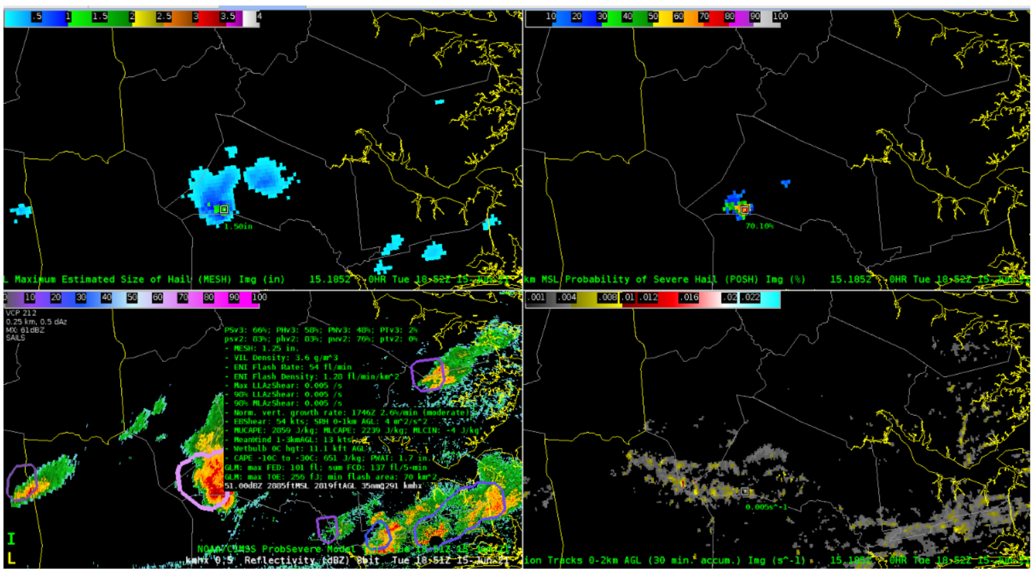

Figure 6: A four-panel display of MRMS MESH (upper left), Probability of Severe Hail (upper right), -10°C Isothermal Reflectivity overlaid with ProbSevere and sampled values (lower left), and 0.2 km Rotation Tracks (lower right).

‘The ProbSevere readout is useful with all the details shown, but when in 4-panel mode [Figure 6], it takes up a lot of space, sometimes with some data being omitted by the frame. Would it be possible to have the option of a simplified readout displaying only the top 2 line probabilities and then have the option of displaying the additional data? (Perhaps akin to the double left click on a ProbSev polygon for the time series chart)’
17 June 2021 Blog Post: ProbSevere Data Readout Sampling
https://inside.nssl.noaa.gov/ewp/2021/06/17/probsevere-data-readout-sampling/

Recommendations for Operational Implementation

Based upon the evaluation of ProbSevere v3 products in the 2021 HWT Satellite Proving Ground, the following items have been recommended:

- It is highly recommended that development and implementation of the AWIPS ProbSevere time series tool into Version 3 continue, so forecasters can readily diagnose convective trends when making warning decisions. Features such as displaying multiple time series, providing model predictors as a time series, and allowing the data to be displayed in AWIPS as a floating tab are all potential areas of investigation.
- It is strongly recommended that ProbSevere Version 3 be implemented into AWIPS, for increased performance in both severe and sub-severe environments when compared with Version 2.
- It is recommended that future training of forecasters with ProbSevere Version 3 involve recalibration for new probabilities, the variables included for each hazard model, performance metrics for each WFO's CWA, and case studies from a variety of convective intensities and modes.

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

The NOAA Unique Combined Atmospheric Processing System (NUCAPS) was evaluated in its sixth consecutive GOES-R/JPSS HWT Satellite Proving Ground (Esmaili et al. 2020; Gambacorta 2013). Temperature and moisture profiles from NUCAPS were used to provide four products for the 2021 experiment: NUCAPS Profiles, Modified NUCAPS Profiles, Gridded NUCAPS (Berndt et al. 2020), and NUCAPS-Forecast. Profiles in the 2021 HWT were generated using select channels from a mix of infrared and microwave sounders in low-Earth orbit onboard six satellites (Table 2). Overpass times varied between the satellites, with certain clusters passing during the morning and afternoon hours locally. Morning data consisted of the MetOp Series Satellites passing over the CONUS between 13 and 18 UTC. Afternoon data were from Aqua, NOAA-20, and SNPP passing over the CONUS between 16 and 21 UTC. To reduce data latency in the HWT from over 60 minutes to about 30 minutes, the initial profiles used to create NUCAPS Profiles and Gridded NUCAPS were received via a direct broadcast link. All satellites were available throughout the experiment with exception of the SNPP, due to an instrument failure just before the start of the testbed.

Satellite	**Infrared Sounder**	**Microwave Sounder(s)**	**Local Overpass Time (Approx.)**
NOAA-20 *SNPP*	Cross-track Infrared Sounder (CrIS)	Advanced Technology Microwave Sounder (ATMS)	1:30 PM
MetOp-A/B/C	Infrared Atmospheric Sounding Interferometer (IASI)	Advanced Microwave Sounding Unit (AMSU)/ Microwave Humidity Sensor (MHS)	9:30 AM
Aqua	Atmospheric Infrared Sounder (AIRS)	Advanced Microwave Sounding Unit (AMSU)	1:30 PM

Table 2: Information regarding the infrared and microwave sounders on all the satellites integrated into NUCAPS for the 2021 HWT.

NUCAPS Profiles were viewed in the National Skew-T and Hodograph Analysis and Research Program (NSHARP) application in AWIPS, along with the addition of the Skew-T and Hodograph Analysis and Research Program-python (SHARPpy) application (Blumberg et al. 2017) for the 2021 experiment. Sounding points in each application were colored red, yellow, and green based upon the quality of the retrievals between the microwave and infrared sounders. Green (red) represented where both the microwave and infrared data were successfully (unsuccessfully) retrieved. Yellow points then represented where the microwave retrieval data was successfully collected while the infrared retrieval failed. These QC flags do not directly determine the accuracy of the sounding and whether the sounding is an accurate representation of the atmosphere, therefore points colored yellow may still contain useful information to a forecaster. Modified NUCAPS Profiles, available only in NSHARP, provided an adjustment to the boundary layer by using Real-Time Mesoscale Analysis (RTMA) data over the previous seven hours. NUCAPS Profiles often struggle to generate accurate retrievals within the boundary layer, which may have substantial impacts on convective parameters often used by forecasters, therefore an adjustment to the profiles is intended to increase the accuracy of the measurements.

Along with viewing the individual soundings from each overpass, forecasters had the ability to view NUCAPS data in a planar view using the Gridded NUCAPS product. Specific meteorological fields from each sounding could then be interrogated from an overpass on a constant pressure surface or layer. Gridded NUCAPS has been shown to identify regions of cold air aloft and assess pre-convective environments. NUCAPS-Forecast then expands the temporal utility of the Gridded NUCAPS product by advecting soundings from NUCAPS forward in time. Soundings are advected forward using the NOAA Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model from one to six hours assuming an adiabatic environment, followed by remapping temperature and moisture profiles and calculating convective indices using SHARPpy. The final products are then displayed in AWIPS at a resolution 0.5 degrees or approximately 50 km.

All four NUCAPS products were evaluated in the 2021 experiment to determine their utility in pre-convective environments including product accuracy, display quality, forecaster understanding, and readiness in operations. Prior to the testbed forecasters were provided with several quick guides for each product, a document with the timing of overpasses from each satellite, a video on NUCAPS soundings for the 2021 HWT, and a NASA SPoRT training module with case studies featuring NUCAPS. Overall forecasters found value in using NUCAPS in pre-convective environments, especially NUCAPS Profiles. Though limitations in using NUCAPS to assess temporal trends still exist. Use of the NUCAPS products in the 2021 experiment along with feedback from forecasters regarding product applications, limitations, and display characteristics are described in the following section.

Use of NUCAPS in the HWT

As was the goal of NUCAPS, its products were most often used to interrogate pre-convective environments for their ability to initiate and maintain thunderstorms. Forecasters often used individual soundings from NUCAPS Profiles and Modified NUCAPS to understand the strength of low-level capping inversions and lapse rates for the ability to initiate convection. Amongst all NUCAPS products, the NUCAPS Profiles were most often used by forecasters and in surveys viewed as the most useful. In the daily surveys, 70% of all responses (38/54) stated they used the NUCAPS Profiles that day, followed by Modified NUCAPS at 51%, Gridded NUCAPS at 36%, and NUCAPS-Forecast at 20%. When asked how often each NUCAPS product provided a positive contribution to the forecast, NUCAPS Profiles were again at the top with 61% (28/46) answering ‘About half the time’ or ‘Usually’. Reported applications of NUCAPS Profiles and Modified NUCAPS from forecasters are provided below.

> ‘I used NUCAPS profiles to compare them with how much the cap had eroded in the area from the morning BIS sounding. This verified that winds would be reaching the ground.’
> *Forecaster – End of Day Survey*

> ‘We used the NUCAPS soundings to get a feel for the environment around the convection that was ongoing in JAN's [Jackson, MS] area. The modified soundings

seemed to have a better handle on the low level lapse rates, but we understand there are still some imperfections when it comes to the near sfc mixed layer. Regardless, it was a nice confidence booster that lapse rates were going to be marginal and large hail was not going to be in the cards today.'
Forecaster – End of Day Survey

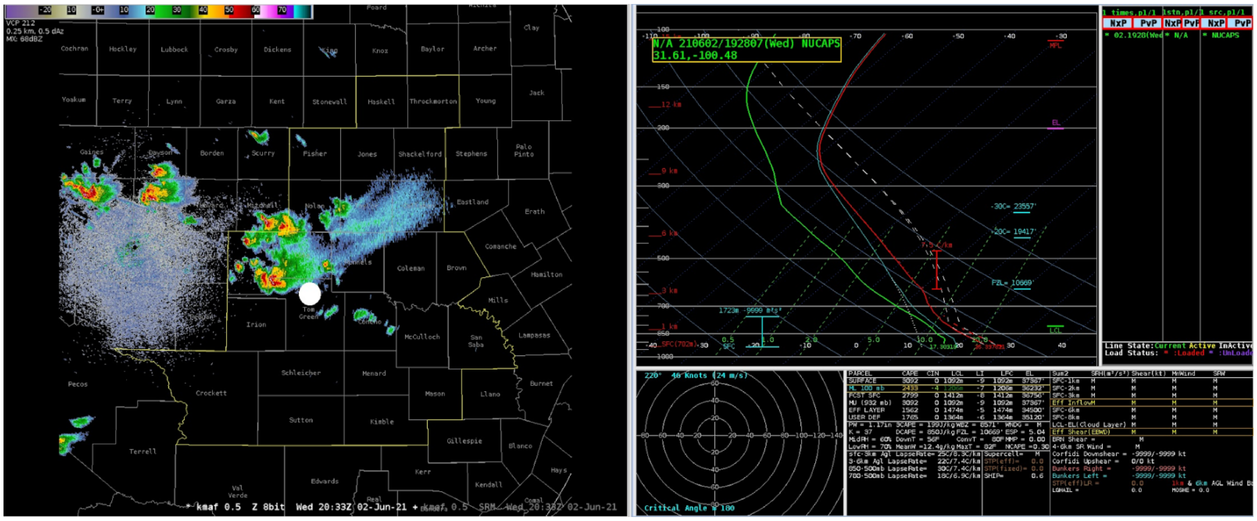

Figure 7: A radar image (left) from KMAF at 2033 Z paired with a NUCAPS Profile (right) from 1927 Z. The location of the sounding is shown with the white dot on the radar image.

'The San Angelo CWA was expecting severe convection in the afternoon. There was a NOAA-20 pass over their CWA at 1927Z. A NUCAPS Sounding in the clear air ahead of ongoing convection was chosen. The approximate point of this sounding is shown by the white circle in the left image [Figure 7]. The sounding in the right image showed an environment very favorable for severe convection, including hail. Since ongoing severe storms were heading in this direction (left image), the storms could be expected to maintain their intensity or possibly strengthen.'
4 June 2021 Blog Post: NOAA-20 Pass Ahead of Severe Convection
https://inside.nssl.noaa.gov/ewp/2021/06/04/noaa-20-pass-ahead-of-severe-convection/

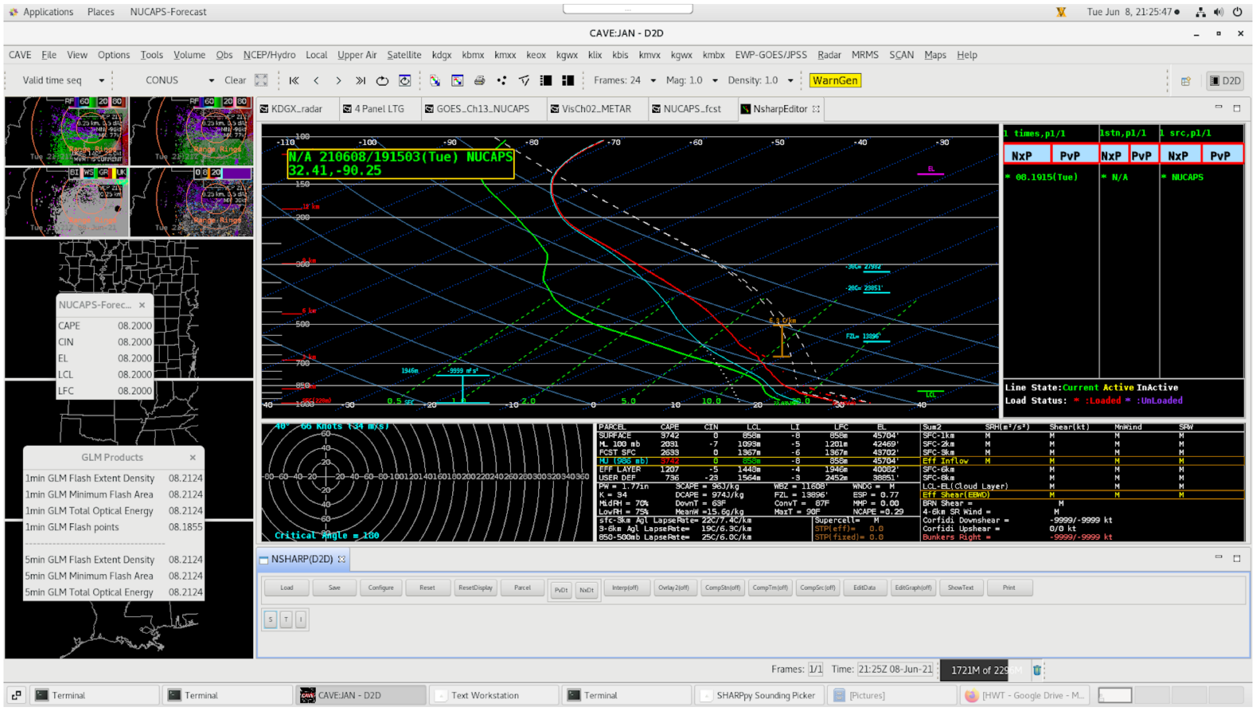

Figure 8: A Modified NUCAPS sounding from 1915 Z near Jackson, MS, displayed using NSHARP in AWIPS.

> 'A modified NUCAPS sounding from near Jackson, MS (which became available later), indicated plenty of instability/CAPE (2000-3500 J kg-1), suggesting that the thundershower would be maintained as it advected toward the Pickle Fest location. This would be a good time for a "heads-up" to the event venue or EM. The unmodified NUCAPS sounding (not shown) still suggested sufficient instability aloft for the storm to maintain itself.'
> *9 June 2021 Blog Post: Lightning DSS for the Mississippi Pickle Fest*
> https://inside.nssl.noaa.gov/ewp/2021/06/09/lightning-dss-for-the-mississippi-pickle-fest/

During each week NUCAPS developers also spent time showing forecasters how to use the SHARPpy application for displaying and comparing soundings from NUCAPS Profiles and rawinsondes. Often forecasters found SHARPpy easier to use, especially when comparing multiple NUCAPS soundings or in comparison with models and/or observations. Consistently during group discussions forecasters expressed that they would like for Modified NUCAPS to also be made available in SHARPpy. Participants did note that sometimes values from the same sounding when displayed in NSHARP and SHARPpy contained different surface temperature and dewpoints, however this was cited back to a display difference within AWIPS. Few forecasters used the pop-up Skew-T tool in AWIPS, however those that did found it helpful to observe how upper-level features were modifying the atmosphere in mesoscale environments.

Evaluating parameters such as Convective Available Potential Energy (CAPE), Convective Inhibition (CIN), Precipitable Water (PWAT), and Lifted Index (LI) values from profiles along with Gridded NUCAPS – and sometimes NUCAPS-Forecast – were also leveraged by the participants. When using the information provided by the gridded products, forecasters most often paired them with the SPC Mesoanalysis and numerical model data to understand the accuracy of the NUCAPS observations and how to apply them operationally. Comments regarding Gridded NUCAPS were often centered around how these values compared to the SPC Mesoanalysis data. Additionally, the ability to identify horizontal gradients in Gridded NUCAPS fields sparked discussion about the best display practices within AWIPS. Some forecasters found the values, when smoothed, less representative of the environment, while viewing the data as uniformly colored grid cells of resolutions greater than 50 km made identifying gradients in certain fields difficult. Uses of Gridded NUCAPS and its display as noted by forecasters are highlighted below.

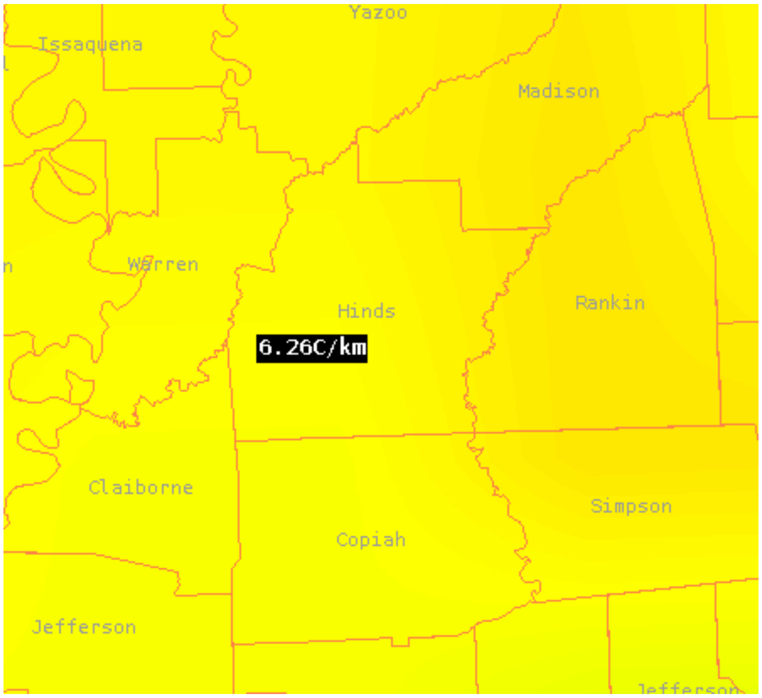

Figure 9: 700-500 mb lapse rates (°C per km.) over the southeastern United States from the Gridded NUCAPS display in AWIPS. Image is valid for 1600 Z on 2 June 2021.

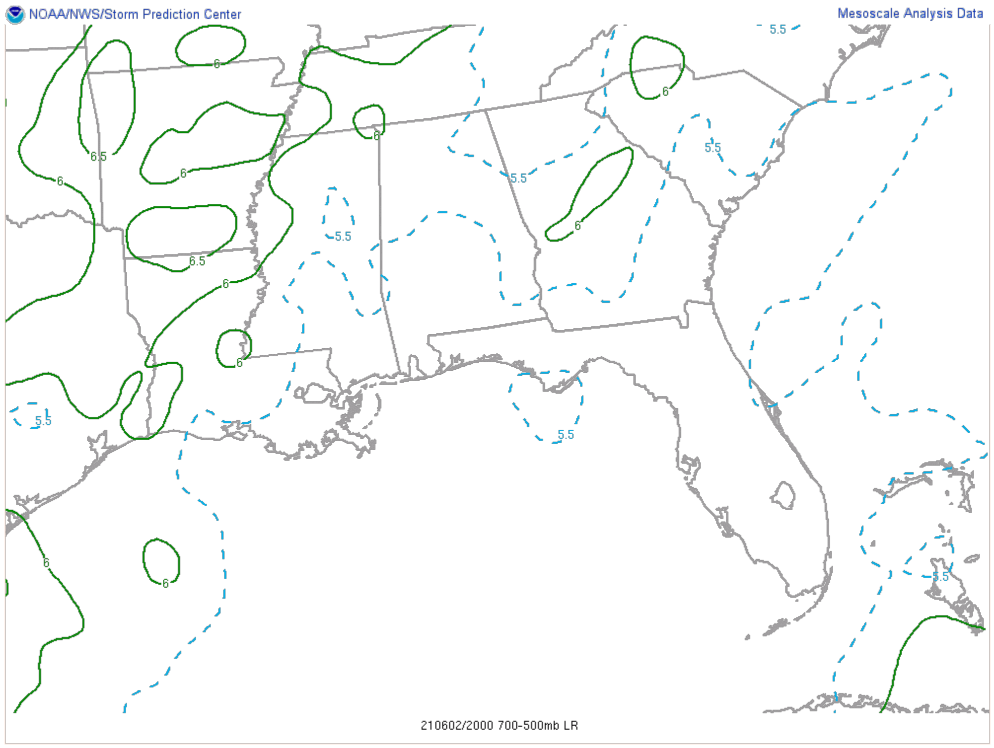

Figure 10: 700-500 mb lapse rates (°C per km.) over the southeastern United States from the SPC Mesoanalysis page. Image is valid for 2000 Z on 2 June 2021.

'The NUCAPS mid-level lapse rates [Figure 9] were fairly representative when compared to the SPC mesoanalysis page [Figure 10]. This was further evidence that large hail was probably not going to be in the cards for the Jackson area today, but marginally severe wind gusts would be something to watch.'
4 June 2021 Blog Post: Sampling of Sub-Severe Convection Across the Southeast
https://inside.nssl.noaa.gov/ewp/2021/06/04/sampling-of-sub-severe-convectoin-across-the-southeast/

'Gridded NUCAPS and individual NUCAPS soundings at 1840z showed steeper 700-500mb lapse rates than what was shown on the SPC mesoanalysis and some of the morning soundings, in areas away from convection. It's hard to say which one was right, but the hail cores observed do seem more consistent with 700-500 mb lapse rates of near 7 C/km or greater. (Note that it would be useful to have contours to go with the images on the gridded NUCAPS plots.)'
17 June 2021 Blog Post: Storms in Northern and Central Florida
https://inside.nssl.noaa.gov/ewp/2021/06/17/storms-in-northern-and-central-florida/

'Using SPC data analysis (CAPE, MLCAPE, PWATs), I did see some good correlation with the larger CAPES (modified soundings) in areas that should have been based on satellite trends also. I did look at PWATs/Lapse Rates/K-Index on the gridded NUCAPS Awips display. The actual values look reasonable based on satellite/convection but the gradient of these indices are more useful operationally. It also gives you a good

> perspective on the greatest threat areas prior to initiation or areas that will continue to develop if already ongoing.'
> *Forecaster – Daily Survey*
>
> 'We looked at TPW in the gridded NUCAPs data. Found some swaths to match nearly identically to the 18 UTC special soundings while other swaths were unrealistically high and therefore unusable. NUCAP sounding did not capture the mid-level dry air, capping inversion or EML that the 18 UTC special soundings identified and also was overall too moist through the column.'
> *Forecaster – Daily Survey*

Amongst all four products in the 2021 experiment, NUCAPS appeared to be the most negatively impacted by a condensed schedule due to the number of available products and its primary utility in pre-convective environments. Additionally, simulated CWAs this year were often placed in uncapped environments where convection quickly initiated or was already underway by the time operations began. Within the weekly surveys, a few forecasters indicated they did not have enough time to view all the NUCAPS products before convection started. NUCAPS-Forecast received limited attention each day as previously mentioned in the daily surveys, which likely impacted how useful forecasters found the product. In the daily surveys, NUCAPS-Forecast was reported as having the least contribution to the day's forecast, with 72% (26/36) responding with the two lowest categories of 'Never' or 'Seldom'.

Data availability during the operation periods appeared to be one of the biggest deficiencies in NUCAPS data throughout the testbed for a few reasons. Forecasters overwhelmingly responded that they had difficulty using multiple overpasses within NUCAPS to assess temporal trends. When participants were asked how often they used each NUCAPS product to assess temporal trends in the daily surveys, over 50% of responses for each category were for the lowest value 'Never'. Additionally, 79% (34/43) of responses indicated that soundings from multiple soundings were not helpful that day for several reasons. Forecasters explained issues like those mentioned previously regarding the limited time in operations and working in mostly uncapped environments where convection quickly initiated. Additionally, the loss of SNPP data and the longer latency of the experimental Aqua products further restricted the ability for forecasters to diagnose environmental trends prior to convection initiation with the set of multi-satellite early afternoon overpasses across the United States.

When forecasters were asked to suggest potential improvements for each product in the weekly surveys, the most common response across all four products was greater data availability. Most reasons for this response parallel those above concerning issues with diagnosing trends. However, one additional point to consider was forecasters not always being aware of when each satellite overpass occurred within NUCAPS. This was suggested by the high frequency that forecasters would ask developers when an overpass would occur in the daily discussions and in operations. The following quotes summarize the following issues raised by forecasters from the previous three paragraphs.

‘I was not able to properly use NUCAPS at all due to the time spent with Day 1 virtual "setup and familiarization" and the "short" 4-hour weather session as opposed to the more typical 8-hour HWT days.’
Forecaster – End of Day Survey

‘Due to storms already ongoing, we were focused on assessing prob severe and the GLM.’
Forecaster – End of Day Survey

Ultimately the thing that would make it more useful is having more frequent data from it, so more satellite overpasses would be most important.
Forecaster – End of Week Survey

Recommendations for Operational Implementation

Based upon the evaluation of NUCAPS products in the 2021 HWT Satellite Proving Ground, the following items have been recommended:

- It is strongly recommended that the use of NUCAPS in SHARPpy continue to be explored, along with training for increased use in the HWT. Implementation of Modified NUCAPS soundings within SHARPpy is also recommended.
- It is recommended that the creation of a merged-Gridded NUCAPS product, which combines data from overpasses in close temporal proximity, be explored to potentially increase product utility.
- It is recommended that future forecaster training efforts regarding NUCAPS focus on best display practices, especially with the Gridded NUCAPS product, so users can fully leverage the available temperature and moisture profiles.
- It is strongly recommended that relevant overpass times be clearly stated for each NWS CWA, such that forecasters know when data are being collected from each satellite and made available through NUCAPS. Look-up tables, web displays, or other forms of relevant media may be used to disseminate this information.

3.4 Optical Flow Winds

One product evaluated in the HWT for the first time was the Optical Flow Winds product from NOAA-NSSL and UW-CIMSS (Rabin 2021), which is created using 1-minute ABI imagery from the GOES-16/17 mesoscale scenes using an optical flow technique (Brox et al. 2004). The product provides wind vectors at a higher spatial (2 km) and temporal (approximately 3 minutes) resolution when compared to the GOES Atmospheric Motion Vectors (AMVs), with the intended purposes of monitoring divergence signatures near the overshooting tops of thunderstorms, vertical shear between cloud layers, and mesoscale boundaries. For the 2021 satellite testbed, the five most recent Optical Flow Wind outputs were displayed in a web-based tool for each mesoscale scene. Users could select color-coated wind barbs with a resolution of 10 km from five product levels (1000-800, 800-600, 600-400, 400-200, and 200-100 mb) along with an 'All' product which provided a color-enhanced image of wind speed across all pressure levels at a resolution of 2 km. For the layered wind vectors, enhanced imagery from the ABI Clean-Infrared band (10.3µm) was used as a background image for the products to depict fine changes in brightness temperature at a resolution of 0.1 K. Additional tools in the web interface included a 'Zoom' feature, animation controls, and a 'Show' button which opens single images in a new browser tab to save individual images.

Prior to the testbed, forecasters were given a product guide outlining how the product was made, features of the web-based tool, and case studies showing how the product responds in cases of severe convection. When deciding locations each day for warning and DSS operations, regions of interest coinciding with GOES-16/17 mesoscale scenes were given an elevated priority, so forecasters had more opportunities to use the Optical Flow Winds product. Overall forecasters liked the Optical Flow Winds product and found increased utility when compared to the GOES AMVs, but felt that several modifications to the product, along with increased training, would allow them to further exploit Optical Flow Winds in pre-convective and convective environments. Feedback from forecasters regarding applications, display modifications, the potential for derived products, and limitations to product use are described in the following section.

Use of Optical Flow Winds in the HWT

During the first two weeks of the testbed a few display modifications were made to the Optical Flow Winds web tool, as several forecasters had difficulty referencing locations within the GOES mesoscale scene. Updates included removing latitude and longitude lines, adding more visible colors for state and CWA boundaries, and including surface observations and placenames. Discussions with forecasters indicated that these updates were well received and improved their ability to relate features observed from the product with those in AWIPS.

During operational periods, forecasters used the Optical Flow Winds most often to monitor cloud top divergence signatures in conjunction with ABI, GLM, and ProbSevere products. Intensifying wind speeds, coinciding with divergence signatures from the wind barbs in the Optical Flow Winds, were correlated with cooling cloud tops, increasing flash rates, and an intensifying updraft and an increased likelihood of severe weather. Forecasters also used the Optical Flow Winds to diagnose features in environments prior to convection, such as utilizing low level clouds to identify the location of mesoscale boundaries. Changes in vertical wind shear across the GOES mesoscale scene could also be ascertained and, when paired with ongoing convection, provided evidence for convection moving into kinematically favorable environments. Quotes from forecasters in the daily surveys and blog posts highlight these uses:

> 'There was a very complex flow pattern over the area, with a MCV and several boundaries in the area. The OFW were helpful in determining the depth of each of these features.'
> *Forecaster – End of Day Survey*

> 'Divergence was noted once storms became organized. This was helpful to see the winds reinforcing that the storm was maturing. They matched positively with radar and lightning trends, and showed the storm maturing and becoming stronger before the radar scanned it.'
> *Forecaster – End of Day Survey*

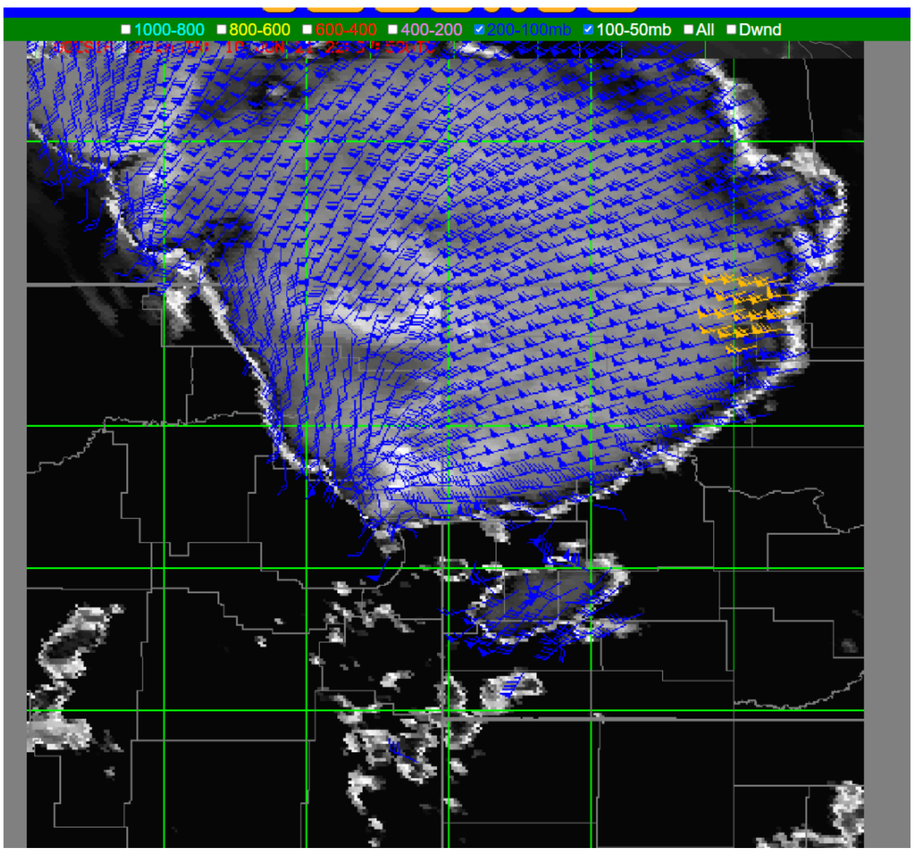

Figure 11: A view of the Optical Flow Winds product used to diagnose and monitor a cloud top divergence signature from the blog post 'Using Optical Wind Flow'.

> 'Convective storms that developed across northeast Montana showed the speed and direction of the outflow. Note the divergent flow outward from the overshooting top [Figure 11], to the outer parts of the anvil. This image also provides the pressure level associated with the wind speed and direction of the flow aloft. This can help determine the strength and possible intensity as it evolves through the life cycle of the convection storm.'

15 June 2021 Blog Post, Using Optical Wind Flow
https://inside.nssl.noaa.gov/ewp/2021/06/15/using-optical-wind-flow-2/

When forecasters were asked to compare the Optical Flow Winds product to the GOES AMVs in AWIPS in the daily and weekly surveys, they at first appear to offer contrasting results. In the daily surveys, two thirds of responses (29/43) indicated that the Optical Flow Winds did not provide additional information that can be accessed from the GOES AMVs. Explanations of the responses from this question revealed over 20 cases where the forecaster responded 'No' because they did not use the Optical Flow Winds product during the event. A few forecasters indicated they do not regularly use AMVs when assessing convective cloud tops. Additionally, there were two days in the testbed where the selected CWAs did not coincide with the location of the GOES-16/17 mesoscale scenes.

In contrast, 47% (7/15) of forecasters in the weekly surveys responded that the Optical Flow Winds were more useful than the GOES AMVs, with 20% (3/15) responding that they were less useful and 33% (5/15) responding that they were overall the same. Results from the daily and weekly surveys indicate that forecasters generally found the Optical Flow Winds more useful than the GOES AMVs. However, the limited use of satellite-derived winds by forecasters reveals that training regarding the utility of satellite-derived wind fields in convective forecasting may be needed to maximize the utility of this product. Additional comments are provided below:

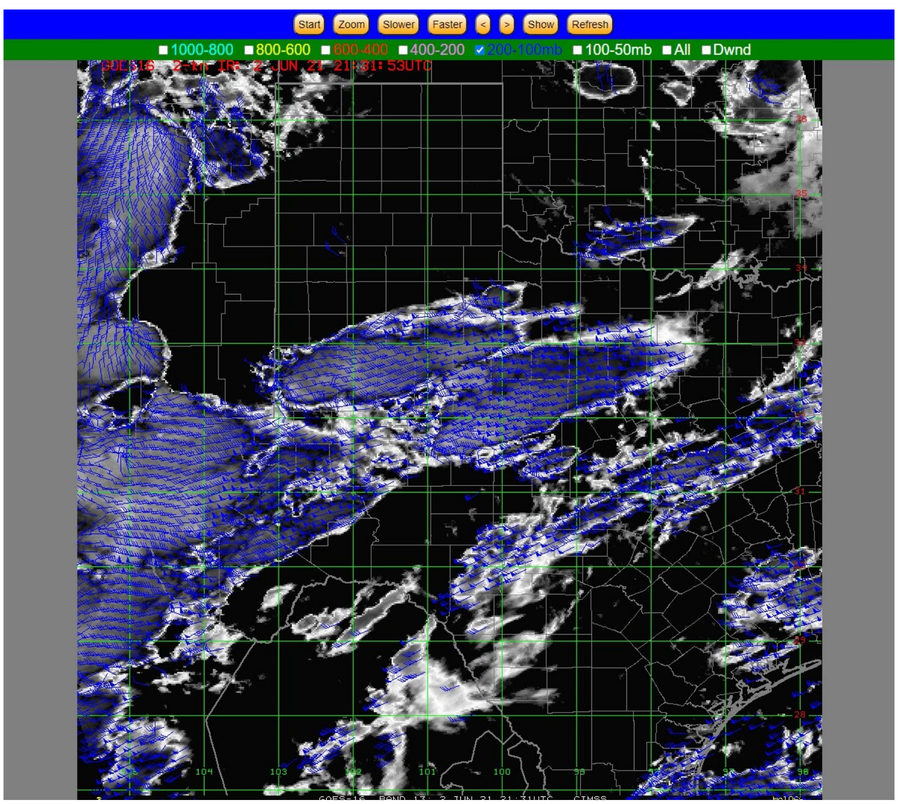

Figure 12: A view of the Optical Flow Winds product over central Texas on 2 June 2021 at 2131 UTC.

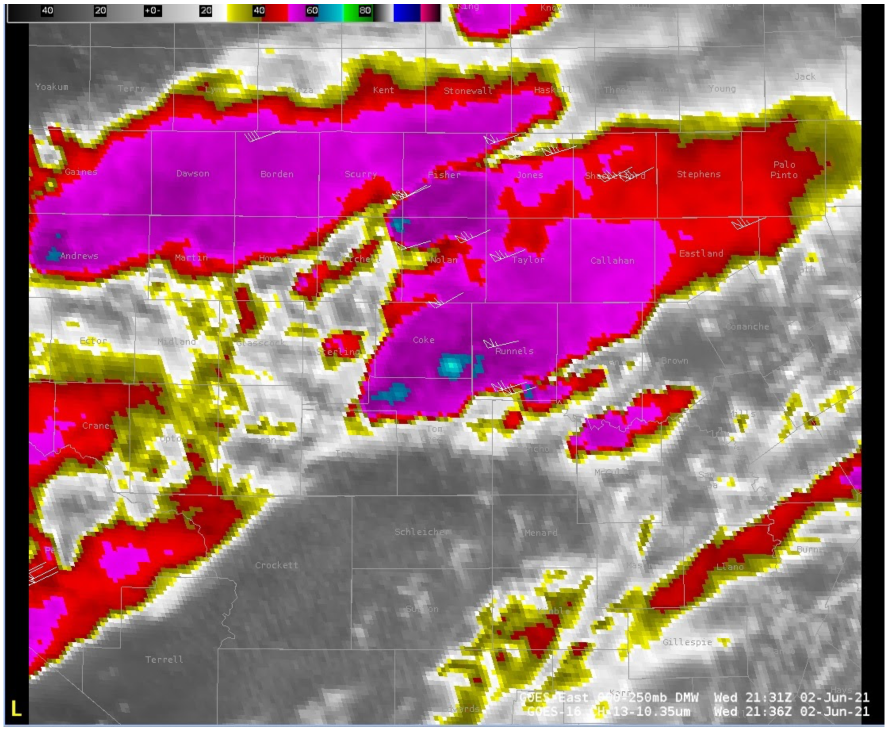

Figure 13: A view of the GOES-East DMVs over central Texas on 2 June 2021 at 2131 UTC, as a direct comparison to Figure 12.

> 'The optical flow winds product [Figure 12] had much better areal coverage than the GOES-East 000-250 mb DMW product [Figure 13] for strong thunderstorms in the San Angelo CWA. However, it would have been very nice to have both products in virtual AWIPS to compare.'
> *4 June 2021 Blog Post: Optical Flow Winds Versus GOES-East DMW*
> https://inside.nssl.noaa.gov/ewp/2021/06/04/optical-flow-winds-versus-goes-east-dmw/
>
> 'Divergence was depicted, but it was still hard to make out exactly what areas were correlated with certain storms.'
> *Forecaster – End of Day Survey*
>
> 'At this point, not really sure how to incorporate this information into the forecast.'
> *Forecaster – End of Day Survey*

Discussions with forecasters, along with survey responses, consistently showed an interest in using the Optical Flow Winds product to create additional derived products. Fields such as horizonal divergence and vertical wind shear were mentioned as products that may hold additional value in operations in terms of direct forecast applications and data visualization. The ability to quickly identify these features at high spatial and temporal resolutions, like those provided by the Optical Flow Winds product, were mentioned as ways this product may help to boost forecaster confidence in warning decisions. Additionally, derived products were expressed as a way to more directly correlate convective features with those from radar and other satellite platforms. Quotes from the forecasters in blog posts, daily surveys, and weekly surveys show this interest in additional derived products, and are featured below:

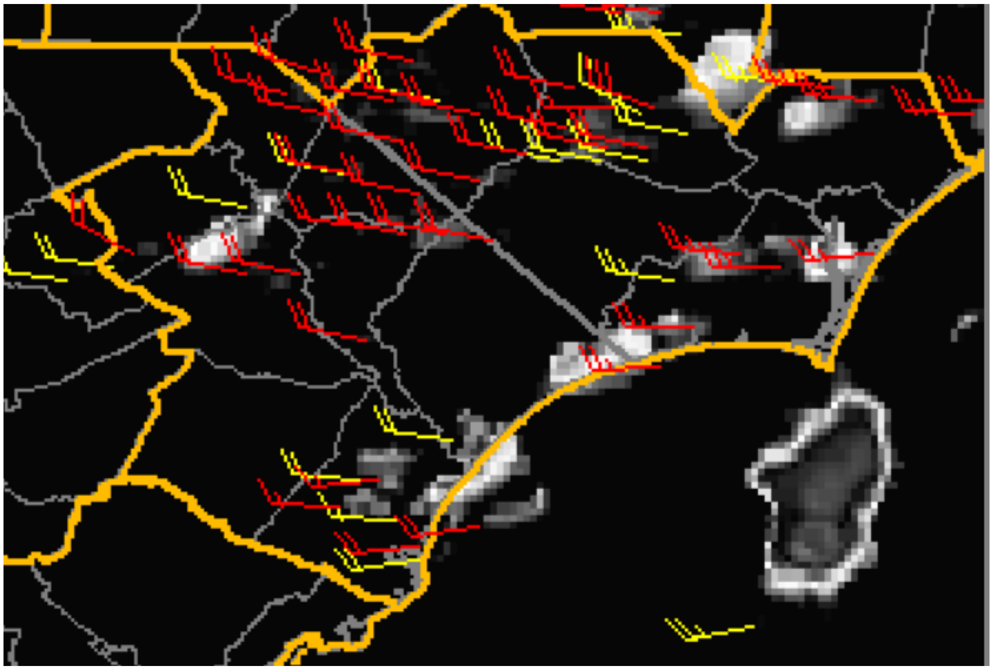

Figure 14 'Optical winds for the ILM CWA on 6/15 at 18Z showing little difference in the winds between the 800-600mb and 600-400 mb levels.'

'Looking at multiple levels of optical winds can be useful in analyzing the amount of wind shear over an area in near-real time. In this case [Figure 14], the tool shows limited wind shear, so one would expect storms to be a bit more short lived. Would it be possible to add wind shear fields directly into this tool for quicker analysis?'
17 June 2021 Blog Post: Optical Wind, GLM, And ProbSevere Use in Convective Environments in North Carolina and South Carolina
https://inside.nssl.noaa.gov/ewp/2021/06/17/optical-wind-glm-and-probsevere-use-in-convective-environments-in-north-carolina-and-south-carolina/

'I think there is value in the OFW, especially if there were layer-based contours indicating areas of divergence and shear, etc.'
Forecaster – End of Day Survey

'Overall I struggled to find much use for the raw winds themselves, but perhaps some derived products (e.g. divergence, bulk shear enhancements) could be useful to view and also go into ProbSevere.'
Forecaster – End of Week Survey

'I know it was mentioned that gridded Optical Flow Winds is a little tricky mathematically, but I do wonder if there could be some utility in terms of machine learning and if it were assimilated into some models to see how it would perform.'
Forecaster – End of Week Survey

Recommendations for Operational Implementation

Based upon the evaluation of the Optical Flow Wind product in the 2021 HWT Satellite Proving Ground, the following items have been recommended:

- It is strongly recommended that future research and development efforts of the optical flow winds product involve derived fields such as horizontal divergence and vertical wind shear, to increase the applicability of high resolution, satellite-derived wind fields.
- It is strongly recommended that additional training be provided to NWS forecasters regarding the use of satellite-derived winds when interrogating pre-convective environments and cloud top characteristics of convection. Detailed examples of how GOES AMVs and the Optical Flow Winds product can be effectively used in convective environments through blog posts, recorded presentations, and one page 'quick guides' may increase forecaster understanding.
- It is recommended that integration of the Optical Flow Winds product into AWIPS be explored, along with effective visualization techniques when used in concert with ABI imagery.

4. Summary and Conclusions

The GOES-R and JPSS Satellite Proving Ground conducted three weeks of remote satellite product evaluations during the 2021 Spring Experiment in a virtual format of the Hazardous Weather Testbed. Seventeen NWS forecasters evaluated four GOES-R and JPSS products and interacted with multiple algorithm developers and subject matter experts during the experiment. Quantitative feedback was collected primarily through surveys administered at the end of each day and week, with more qualitative feedback coming from daily discussions with forecasters and blog posts. Participants received training materials prior to the testbed for each product through a combination of user guides, PowerPoint presentations, and online learning modules. Products were also summarized at the beginning of each week by their developers, which included product applications, limitations, locations in AWIPS, and recommended display practices. During daily discussions a few forecasters did comment that pre-built procedures may have been helpful for increasing how soon forecasters could begin looking at products in a condensed operations schedule this year, along with following best display practices to fully leverage the experimental products.

One new addition this year was the creation of simulated DSS events within operations, with forecasters providing useful and timely information for public safety. As a growing mission within the NWS, the creation of these events created a new evaluation opportunity that challenges forecasters to use the data beyond warning decisions. Interpretation and communication of experimental products within this framework further revealed forecaster understanding, which directly impacts a product's ability to be integrated into operations and its associated training. This year the experiment only featured two simulated DSS events, however expansion of DSS applications within the evaluation process may provide additional insight into direct applications of experimental products within the testbed.

Like all other HWT activities during the 2021 Spring Experiment, the Satellite Proving Ground was held entirely in a virtual format due to the COVID-19 Pandemic. This provided new logistical and technological challenges that had to solved by the project managers and technical support staff, Product developers and NWS forecasters were also required to adapt to a different testing and evaluation environment compared to all previous iterations of the experiment. When asked about their experiences in a virtual testbed, both developers and forecasters generally felt that the experiment still provided a valuable opportunity to test new products. However, those with experience in previous iterations of the satellite testbed consistently said that there were some limitations to this year's entirely virtual experiment that impacted the quality and quantity of feedback received. One of the most obvious limitations was the compressed operations schedule with static start and end times. This was done to reduce forecaster screen time and fatigue throughout the week, with still a few forecasters commenting that mild physical and mental fatigue was noticeable by the end of the week. Time in operations was further reduced by various technical difficulties associated with each participant's unique computing setup, and often required constant technical support from CIMMS IT staff involved in the experiment. Adapting to the cloud-based AWIPS-II instances also took time on the first day but was often a minimal disruption, as the instances were reliable throughout nearly all the experiment.

One positive aspect of holding a virtual testbed was the added flexibility this allowed forecasters with effectively no travel required to participate. Some participants noted the convenience of not having to arrange the logistics of traveling to Norman, Oklahoma for the Spring Experiment. Additionally, a few forecasters noted that they typically were unable to participate in an in-person experiment due to personal commitments that limited the time they could spend away from their area of residence. Under these scenarios, a testbed with more remote elements may allow for a more diverse pool of applicants across the NWS or other sectors than previously observed.

Recommendations for Future GOES-R/JPSS Satellite Proving Ground Experiments

Based upon the 2021 GOES-R/JPSS Satellite Proving Ground, the following recommendations for future testbeds are included below:

- It is strongly recommended that pre-built AWIPS-II display procedures be created prior to the satellite experiment, with the purpose of allowing forecasters to leverage the evaluated products quickly and efficiently.
- It is strongly recommended that simulated Decisions Support Service events continue within the satellite experiment, with the consideration of expanded use also recommended.
- It is recommended that integration of virtual elements be explored for future experiments including video conferences, collaborative blog posts through file storage and synchronization services, cloud-based AWIPS-II instances, and other relevant forms. Any implementation should be done in coordination with other activities in the Hazardous Weather Testbed's Spring Experiment.

Acknowledgements

Multiple NWS meteorologists participated in this experiment and provided detailed feedback that went into the recommendations in this report. It would not be possible to complete research-to-operations without their willingness to adapt to a virtual environment this year, work with experimental data, ask questions, and participate in crucial discussions with product developers. Johnathan Madden and Justin Monroe provided critical technical support during all phases of the experiment, which allowed for forecasters to effectively use the cloud-based AWIPS instances. The following people also participated as focal points and subject matter experts throughout the week for the products demonstrated in the HWT: Chris Schultz (NASA – Marshall Space Flight Center), Scott Rudlosky (NOAA – Center for Satellite Applications and Research), Joseph Patton (NOAA – Center for Satellite Applications and Research), Justin Sieglaff (University of Wisconsin – CIMSS), John Cintineo (University of Wisconsin - CIMSS), Jack Dostalek (Colorado State University – CIRA), Rebekah Esmaili (Science and Technology Corporation), Brian Kahn (Jet Propulsion Laboratory), Emily Brendt (NASA – Marshall Space Flight Center), Peter Kalmus (NASA – Jet Propulsion Laboratory), Jeff Szkodzinski (Science and Technology Corporation), Bob Rabin (NOAA – National Severe Storms Laboratory).

5. References

Berndt, E., and Coauthors, 2020: Gridded Satellite Sounding Retrievals in Operational Weather Forecasting: Product Description and Emerging Applications. *Remote Sensing*, **12**, https://doi.org/10.3390/rs12203311.

Blumberg, W. G., K. T. Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPpy: An Open-Source Sounding Analysis Toolkit for the Atmospheric Sciences. *Bulletin of the American Meteorological Society*, **98**, 1625–1636, https://doi.org/10.1175/BAMS-D-15-00309.1.

Brox, T., A. Bruhn, N. Papenberg, and J. Weickert, 2004: High Accuracy Optical Flow Estimation Based on a Theory for Warping. *Computer Vision - ECCV 2004*, T. Pajdla and J. Matas, Eds., Berlin, Heidelberg, Springer Berlin Heidelberg, 25–36.

Bruning, E. C., and Coauthors, 2019: Meteorological Imagery for the Geostationary Lightning Mapper. *J. Geophys. Res. Atmos.*, **124**, 14285–14309, https://doi.org/10.1029/2019JD030874.

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. *Weather and Forecasting*, **29**, 639–653, https://doi.org/10.1175/WAF-D-13-00113.1.

——, ——, ——, L. Cronce, and J. Brunner, 2020a: NOAA ProbSevere v2.0?ProbHail, ProbWind, and ProbTor. *Weather and Forecasting*, **35**, 1523–1543, https://doi.org/10.1175/WAF-D-19-0242.1.

——, ——, ——, A. Wimmers, J. Brunner, and W. Bellon, 2020b: A Deep-Learning Model for Automated Detection of Intense Midlatitude Convection Using Geostationary Satellite Images. *Weather and Forecasting*, **35**, 2567–2588, https://doi.org/10.1175/WAF-D-20-0028.1.

Esmaili, R. B., and Coauthors, 2020: Adapting Satellite Soundings for Operational Forecasting within the Hazardous Weather Testbed. *Remote Sensing*, **12**, https://doi.org/10.3390/rs12050886.

Gambacorta, A., 2013: The NOAA Unique CrIS/ATMS Processing System (NUCAPS): Algorithm Theoretical Basis Documentation.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmospheric Research*, **125–126**, 34–49, https://doi.org/10.1016/j.atmosres.2013.01.006.

Rabin, B., 2021: Wind Estimates in Storms from an Optical Flow Technique Using GOES Imagery.