GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2024 Spring Experiment Final Evaluation

Project Title: GOES-R and JPSS Proving Ground Demonstration at the 2024 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 Severe and High-Impact Weather Research and Operations (CIWRO)

Duration of Evaluation: 13 May 2024 – 7 June 2024

Prepared By: Kevin Thiel (OU/CIWRO and NOAA/SPC)

Submitted Date: 20 September 2024

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 2023 Spring Experiment, which took place in-person and virtually at the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) in Norman, Oklahoma from 13 May to 7 June 2023. This year featured 16 participants in the EWP experiment. All the participants were National Weather Service (NWS) forecasters from Weather Forecast Offices (WFOs) in four NWS regions. This group evaluated five experimental GOES-R products in the real-time simulated short-term forecasting, decision support service (DSS), and warning environment of the Experimental Warning Program (EWP). Additionally, they used cloud-based instances of the second-generation Advanced Weather Interactive Processing System (AWIPS-II) and web-based interfaces to interact with the products.

Forecaster feedback during the evaluation was collected through daily and weekly surveys, daily and weekly debriefs, blog posts, a warning and DSS reporting form, public forecast graphics, and informal conversations during the testbed. Typical feedback included suggestions for improving the algorithms, display techniques, training, and awareness of product applications or limitations. Most of the products evaluated in 2024 were advancements of previous product iterations from the 2023 GOES-R/JPSS Proving Ground (Thiel 2023). This included data from the Optical flow Code for Tracking, Atmospheric motion vectors, and Nowcasting Experiments (OCTANE) suite, the Polar Hyperspectral Sounder and Microwave Imagery in the Advanced Baseline Imager (PHS) model, and the Probability of Severe (ProbSevere) LightningCast model. New to the testbed were the Geostationary Lightning Mapper (GLM) Background and Data Quality Product (DQP), along with the GOES Radar Estimation via Machine Learning to Inform NWP (GREMLIN).

Over 22 visiting scientists attended the experiment over the three weeks to provide additional product expertise and interact directly with operational forecasters. Organizations represented by those individuals included three NOAA Cooperative Institutes, five federal partners, and two external partners. The Storm Prediction Center (SPC) and HWT Satellite Liaison Kevin Thiel (OU/CIWRO and NOAA/SPC) provided overall project management and subject matter expertise for the HWT Satellite Proving Ground efforts. Technical support for AWIPS-II were provided by Jonathan Madden and Justin Monroe (OU/CIWRO and NOAA/NSSL).

2. Introduction

GOES-R Proving Ground demonstrations in the HWT have provided users with a glimpse into the capabilities, products, and algorithms available on the GOES-R satellite series since 2011. The education and training received by participants in the HWT fosters interest and engagement with new satellite data and promotes the continued readiness of GOES-R data and products. 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 developers understand how forecasters use these products and what improvements might increase product utility in NWS operations. Feedback received from participants in the HWT has proven invaluable to the continued development and refinement of GOES-R algorithms since its inception in 2009. Furthermore, the EWP (Calhoun et al. 2021) facilitates the testing of satellite-based products in the AWIPS-II data processing and visualization system currently used at NWS Weather Forecast Offices (WFOs).

The first two weeks of the testbed (13-17 May and 20-24 May) were conducted in-person at the National Weather Center's HWT in Norman, Oklahoma with four NWS forecasters participating each week. The third week of the experiment (3-7 June) was conducted virtually using Google Meet and Slack, hosting eight NWS forecasters. In-person testbed activities were reintroduced for the 2023 experiment for one of the three demonstration weeks, with the other two weeks conducted virtually. Feedback from forecasters and developers in this experiment (Thiel 2023) motivated the format for the 2024 experiment with one more in-person demonstration week and one less virtual demonstration week. Before the testbed forecasters were provided user guides, PowerPoint presentations, and online learning modules through Google Drive for each of the products demonstrated. Additionally, the Satellite Liaison worked with each product developer group to formulate R2O questions that would guide their evaluation in the experiment.

The Monday of each week began with introductions, an orientation session, and product summaries from developers, followed by familiarizing forecasters with their cloud-based AWIPS instances. Tuesday, Wednesday, and Thursday began with a discussion between the developers and forecasters of the previous day's operations. After a brief forecast discussion each day by the Satellite Liaison, forecasters were placed into two to three groups localized to various NWS WFOs (hereafter simulated WFOs) across the United States to begin operations. The simulated WFOs were positioned to maximize the probability of severe thunderstorm activity each day, and to provide adequate evaluation of all demonstrated products. Mock-Decision Support Service (DSS) events were created for a majority of simulated WFOs to investigate how the experimental products could also be utilized in communicating hazards to NWS partners. In three cases during the experiment, the Satellite Liaison asked forecasters at a simulated WFO to avoid looking at all available radar fields (WSR-88D and MRMS) to assess how the utility and applications of the experimental products changed in this scenario. At the end of each operations period, forecasters were given a daily survey regarding product performance and utility during the day's events. Forecasters were also encouraged to fill out an online form after submitting a convective warning, along with any forecaster who wished to provide DSS messaging for their assigned event. Responses from this form were then examined to identify how the experimental products were incorporated into the communication of convective hazards by the participants. Additionally, forecasters had the option to create social media graphics using the experimental products, further showcasing their ability to be conveyed to the public.

Two new elements were added to the 2024 experiment. The first element was an archived case, created in collaboration with the Radar Convective Applications Experiment. Forecasters and developers from the 2023 experiment wished for an archived case to avoid weeks with little-to-no intense convection, and to ensure multiple evaluation opportunities were presented with each developer group's R2O objectives. All five developer groups submitted data from 1800 Z on 11 May 2023 to 0300 Z on 12 May 2023 focused on the Norman, Oklahoma WFO. The case featured initiating convection in a dynamic mesoscale environment with multiple convective hazards expected, along with the transition from day to night. This case was used once during the experiment in the third week, and feedback from the event will be embedded throughout the report. The second element included this year was intentionally removing radar data from specific WFOs in a data-denial exercise. Radar data-denial occurred in five of the eighteen simulated WFOs, with the goal of assessing experimental satellite product applications in regions with little to no radar coverage.

Forecasters viewed GLM Background and DQP, GREMLIN, OCTANE, PHS Model, and ProbSevere LightningCast data in the cloud-based instances of AWIPS for the in-person and virtual demonstrations. Prior to the testbed, AWIPS procedures were built by the Satellite Liaison and product developers for each product in AWIPS, so forecasters could quickly access the products and leverage best display practices as described in their training. Within operations forecasters had several tasks, such as building procedures of their own to integrate experimental products with the ones they currently use, discussing their interpretation and experiences with the subject matter experts and developers, writing blog posts, and issuing warnings and DSS messages. 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 which were published online to the HWT EWP Blog (https://inside.nssl.noaa.gov/ewp/) and the GOES-R HWT Satellite Proving Ground Blog (http://goesrhwt.blogspot.com/).

The first day of the in-person testbed weeks began at 11am CDT (16 Z) and ended at 7 pm CDT (0 Z). Tuesday through Thursday activities could begin any time between 10am CDT (15 Z) and 1 pm CDT (18 Z) and end eight hours later based on the convective environment, and was decided on the previous day by the Satellite Liaison. No operations occurred Friday, as forecasters held their weekly debrief with the developers and were encouraged to participate in the R2O O2R Tales (ROOTs) webinar hosted by the NWS Warning Decision Training Division (WDTD). During the virtual week, activities on Monday began at 1 pm CDT (18 Z) and ended at 6pm CDT (23 Z). Tuesday through Thursday could begin between 12pm CDT and 2pm CDT and end five hours later. On Friday, an end of week survey was sent to the participants in the morning, followed by a two-hour final discussion with developers, observers, and SMEs to summarize the week's events and encapsulate key product themes from 1 to 3 om CDT. More details from these schedules can be found in Figure 1 and Figure 2.

On-Site Demonstration Schedule

Hour (From start)	0:00	0:30	1:00	1:30	2:00	2:30	3:00	3:30	4:00	4:30	5:00	5:30	6:00	6:30	7:00	7:30
Monday	Orientation				Product Training				Operations							Daily Survey
Tuesday	Discussion/ Forecast		Operations													Daily Survey
Wednesday	Discussion/ Forecast		Operations													Daily Survey
Thursday	Discussion/ Forecast		Operations										Daily Survey	Webinar Prep		Weekly Survey
Friday	Weekly Debrief		Webinar													

Figure 1: An approximate schedule of the in-person GOES-R Proving Ground HWT Experiment, outlining the major activities from each day.

Virtual Demonstartion Schedule

Hour (From start)	0:00	0:30	1:00	1:30	2:00	2:30	3:00	3:30	4:00	4:30
Monday	Orientation			Product Training/Operations						Daily Survey
Tuesday	Discussion/ Forecast		Operations							Daily Survey
Wednesday	Discussion/ Forecast		Operations							Daily Survey
Thursday	Discussion/ Forecast		Operations							Daily Survey
Friday	Weekly Survey	Weekly Debrief								

Figure 2: An approximate schedule of the virtual GOES-R Proving Ground HWT Experiment, outlining the major activities from each day.

The collective feedback from the 2024 GOES-R Proving Ground (daily surveys, weekly surveys, blog posts, and daily debrief discussions) are summarized in this report. Each of products evaluated in the following subsections begin with a summary of the product and its intended applications, followed by science questions from each product developer group. Next, applications and feedback from the forecasters are summarized across all three weeks of the experiment. These are supported by forecaster questions from the surveys, forms, and blog posts throughout each section. Product recommendations are listed at the end of each section as ‘recommended’, ‘strongly recommended’, and ‘highly recommended’ in an ascending order of apparent significance from the forecasters.

3. Products Evaluated

3.1 GLM Background and Data Quality Product

The GLM Background and Data Quality Product (DQP) (Bruning and Bitzer 2023) was demonstrated for the first time in the Satellite Proving Ground HWT, motivated by a request from the NWS for a GLM data quality product. Every 2.5 minutes the GLM captures a background image over its entire field of view (FoV) to use as a brightness reference for lightning flash detections on the pixel level. The GLM records background imagery at a frequency of 777.4 nm, and at the same horizontal resolution as GLM lightning flash events (8 km at nadir, 14 km at FoV edge). Along with the background image itself, a data quality product is derived from the background imagery consisting of flash detection efficiency and pixel-level quality flags. Local flash detection efficiency values are based on relationships between background imagery and flash detection efficiency (DE) from the Lightning Imaging Sensor on the Tropical Rainfall Monitoring Mission, relative to lightning mapping array flash detections (Cummins 2020). The quality flags are eight discrete tags that may be applied to pixels based on their characteristics. Currently only pixels that are near saturation or at saturation are available, but future slots may become available for solar glint, solar intrusion, dropped events (hardware/software), obscured pixels, and dead pixels.

When applying these data operationally, background imagery from the GLM may serve as an operational backup if visible imagery from the GOES-R Advanced Baseline Imager (ABI) were unavailable. Users outside of the ABI CONUS domains could also receive more frequent updates of visible imagery, albeit at a lower spatial resolution. In the testbed only 16 relative levels of brightness were made available to forecasters. The GLM DQP is expected to provide additional context for the operational GLM gridded imagery (Bruning et al. 2019) and increase confidence in these products when observing convection. Both the GLM Background and DQP were provided every five minutes for the experiment. Before the testbed participants were provided with a one-page training guide describing how each product is created, its intended applications, and examples of the products visualization. Additionally, a four-panel procedure was created in AWIPS containing the GLM Background, DQP, Flash Extent Density (FED), and two ABI products to provide the forecasters a baseline display.

GLM Background and Data Quality Product R2O Questions

- Are the GLM Backgrounds useful (resolution, band, in event of ABI outage) in operations?
- Does the GLM Data Quality Product design convey the kinds of information you need to understand systematic and episodic data quality?
- When used alongside the operational GLM products, did the DQP help clarify unusual behaviors in an observed storm?

Use of GLM Background and Data Quality Product in the HWT

Throughout the testbed, forecasters used the GLM Background and DQP to learn more about the GLM instrument and its data, identify features related to convection and data quality, along with exploring optimal display methods of the products. Observed flash DE values by the participants at their simulated WFOs were most often between the >40% and >70%, with approximately half of respondents (16/34) marking that they saw the >60% bin in the daily surveys. Amongst the data quality flags, 44% (24/55) of forecasters did not observe any quality flags during their daily operations period. Only 8 of the 55 responses (15%) noted pixels that were at or near saturation, while 23 responses mentioned observing GLM subarray boundaries. Subarray boundaries were not explicitly defined by the DQP, but rather by the localized increase in flash DE values or the more frequent appearance of nearly saturated/saturated pixels. In group discussions each week, nearly all forecasters stated they were previously unfamiliar with the location of GLM subarray boundaries with the instrument's FoV and its potential impacts on GLM flash detections.

Once forecasters were made aware of GLM subarray boundaries and the operational significance of pixels that were at or near saturation, they were able to identify thunderstorms crossing these areas. As convection initiated around or traversed a subarray boundary, forecasters scrutinized the GLM FED values from each storm and compared its flash rate signals to radar and ground-based lightning networks. When forecasters were asked how the GLM DQP impacted their confidence interpreting operational GLM gridded imagery each day, around half (21/44) noted no change in confidence while almost as many (19) stated the product gave them slightly greater confidence. These results may correspond with the large fraction of forecasters who did not observe any data quality flags during their daily operations period. Examples and applications of the GLM DQP are highlighted in the following blog posts from forecasters.

> 'We noticed a change in the storm on GLM as it tracked northeast. GLM seemed to decrease as the storm passed through the subarray region, but the number of flashes remained relatively the same according to the ground network. Knowing the location of this subarray and also comparing GLM to the ground network gave us confidence that the dip in GLM was not due to a reduction in flashes/storm intensity.' [Figure 3]
> 21 May 2024, Blog Post: *Storm Hopping Over Subarray*
> https://inside.nssl.noaa.gov/ewp/2024/05/21/storm-hopping-over-subarray/

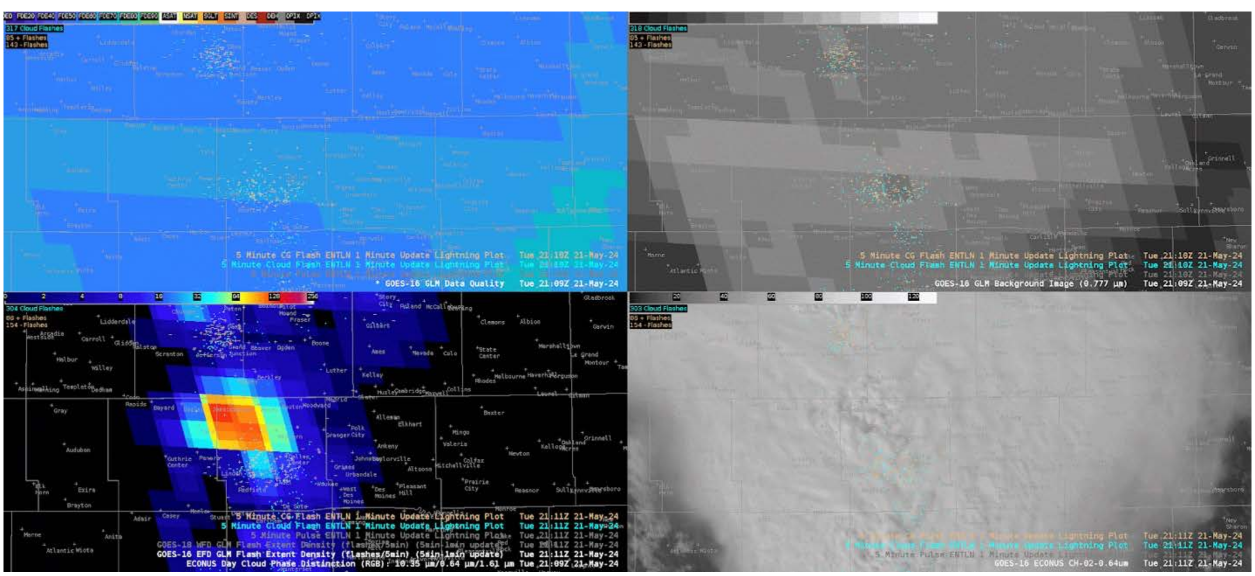

Figure 3: A four-panel image of the GLM data quality product (upper left), GLM background image (upper right), GLM flash extent density (lower left), and ABI visible image (lower right) from a thunderstorm crossing a subarray boundary at 2110 Z on 21 May 2024. Animation.

‘You should be able to make out a sub-array boundary going horizontally (upper left panel) AND also in the GLM background image (upper right panel). In the top right and both bottom panels you can make out strong convection taking place with two cells (one in the southern portion of the CWA, and one just to the south and along the CWA border)… In this example with the relatively close proximity of the two cells one cannot be sure that the GLM data is incorrect, but with the GLM returns showing up on the sub array boundary this does increase uncertainty around this portion of the GLM flash extent density data.’ [Figure 4]
6 June 2024, Blog Post: *Using the GLM Background & Data Quality Product (DQP)*
https://inside.nssl.noaa.gov/ewp/2024/06/06/using-the-glm-background-data-quality-product-dqp/

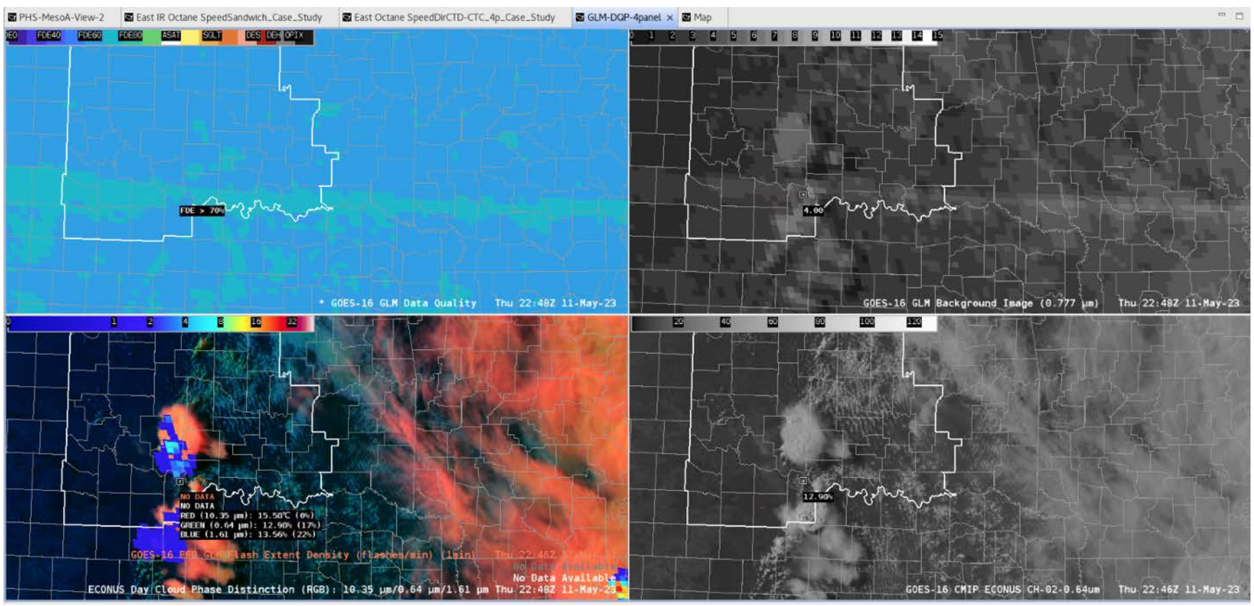

Figure 4: Same display as in Figure 3, but for storms near a subarray boundary at 2248 Z on 11 May 2023.

‘It's course resolution background imagery that GLM uses to detect rapid changes in brightness that can be used to diagnose data quality issues. Commonly used to identify locations of subarrays due to the product being sensitive at these boundaries detecting photons, causing dropout in GLM data.’
NWS Forecaster – End of Week Survey

‘This product tells you the efficiency of how well GOES can detect lightning flashes. You're able to see that GLM is less sensitive at harsher angles near the edge of the satellite view.’
NWS Forecaster – End of Week Survey

During the first week of the experiment, two forecasters worked with visiting scientists to create alternative displays that condense the original four-panel display to a single panel. Forecasters mentioned that they became less concerned with viewing GLM background and DQP as convection initiated and intensified, and that much of the information regarding local flash DE and subarray boundary locations could be obtained before the event began. The expectation with these alternative displays was to retain the necessary information from the GLM background and DQP, while increasing the amount of screen space available to load other products and optimize situational awareness. The first alternative display included overlaying a semi-transparent background image on the DQP, with the brightest values set to an alpha of 0 such that forecasters could clearly observe flash DE and data quality tags near subarray boundaries. In the second option, the DQP was overlaid on the background image, with all flash DE values made transparent and data quality tags set to pink to clearly define pixels with these tags when used with the GLM FED product. Lastly, a semi-transparent red was added to all flash DE values less than the FED >40% bin using the previously described display, to also highlight areas of poor GLM performance within its FoV. The following blog post by a forecaster describes these displays and their initial feedback [Figure 5].

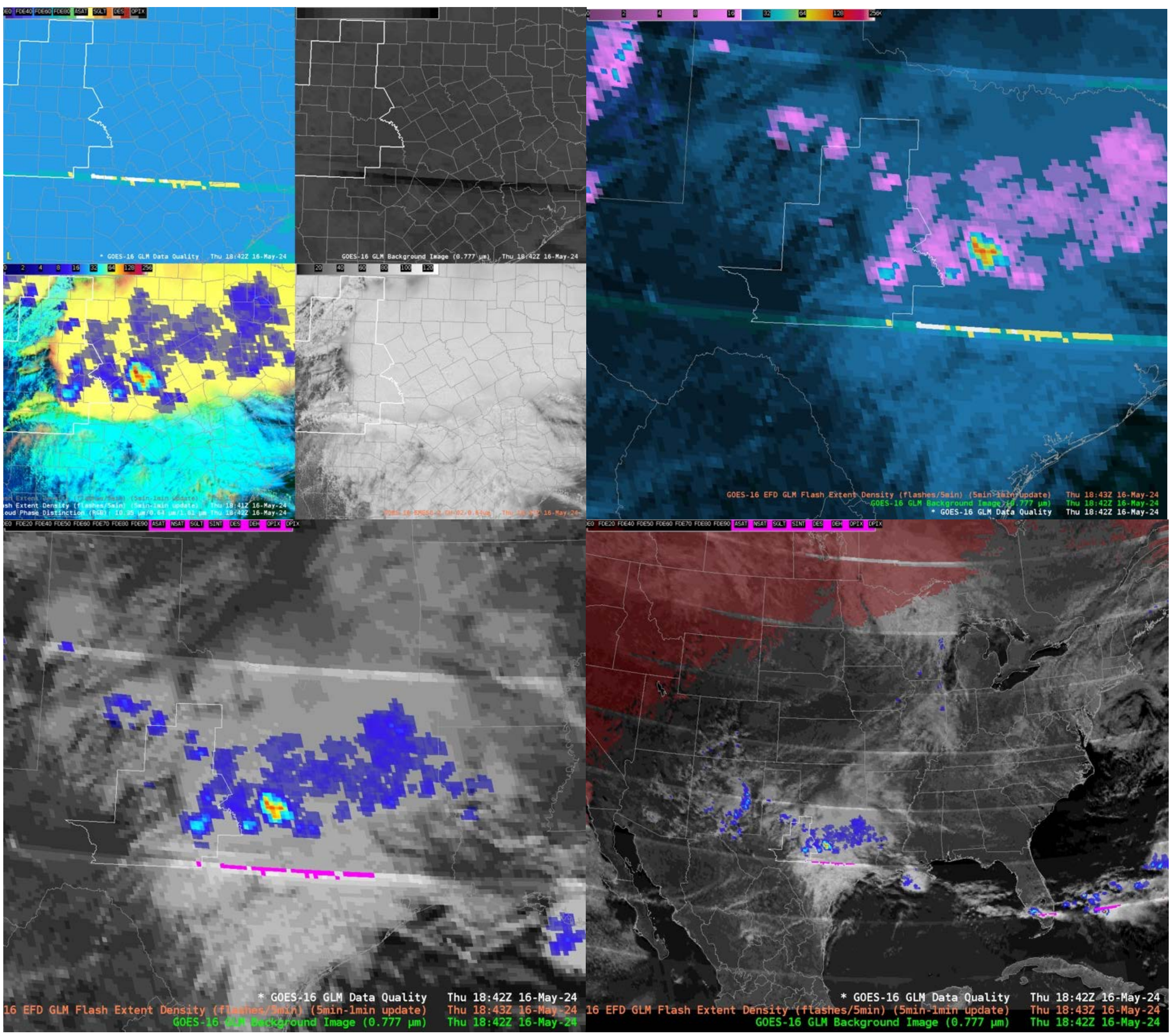

Figure 5: All four display options created in the testbed for the GLM background and data quality product from 1842 Z on 16 May 2024. Upper left animation. Upper right animation. Lower left animation. Lower right animation.

‘When we started the testbed the GLM products were being displayed on a four panel display [Figure 5, upper left]. This display works fine to find areas of poor data quality, like where the white and yellow pixels pop up in the top left of [the image]. However, in its four panel setup, I felt that the display took up too much space. As a group, we worked to merge them into a single panel… I thought there were ways for ways to merge the two designs. My desire would be to have the GLM background image provide texture to the data quality product like in [Figure 5, lower left], but to have the data quality product maintain the sharp good/bad color curve shown in [Figure 5, upper right]. Even more preferred is a color bar like in [Figure 5, lower right], where the “could be poor” but not nearly or fully saturated values are highlighted in another color (red in this example).’
16 May 2024, Blog Post: *Making Four Panels Into One With GLM Data Quality*
https://inside.nssl.noaa.gov/ewp/2024/05/16/making-four-panels-into-one-with-glm-data-quality/

Even though the GLM DQP received most of the forecasters’ attention throughout the experiment, participants often compared the GLM background image to the ABI channel 2 visible band before convection initiated or enter their forecast area. These scenarios frequently included identifying synoptic and mesoscale cloud features, anvil locations, and overshooting tops. When asked at the end of each week how confident the forecasters felt using the GLM background image half (8) responded with the second lowest option ‘A little confident’, six responded with the next highest ‘Moderately confident’, and two responded with the second highest option ‘Very confident’. Similarly, when asked in the same survey how often they would use the GLM background imagery all forecasters selected either ‘Would use as a last resort’ or ‘Would feel comfortable using if I had to (e.g. ABI outage)’. Along with the reduced horizontal resolution of the GLM background image (8 km) when compared to the ABI channel 2 visible band (0.5 km), forecasters noted that subtle cloud features related to textures and gradients became more difficult to distinguish. This may be driven by the fact that GLM background image radiances were divided into 16 levels of brightness to accommodate bandwidth considerations of a potential operational product.

‘I strongly advised using GLM background imagery during an ABI outage as this channel will readily identify mesoscale systems, sea-breeze features, large cumulus/moisture fields, and ongoing deep cellular convection. On the other hand, it will struggle to identify sparse cu fields, small and isolated cloud features, and even early stages of convective initiation.’
NWS Forecaster – End of Day Survey

‘I feel like the one day I did use it in detail, I was able to see the cloud structure with a couple of decaying squall likes that moved through our forecast area earlier in the day and could really see the clear air behind them.’
NWS Forecaster – End of Week Survey

‘As a last resort would it works, yes. But I feel like being able to see an option with a few more bins and possibly a smoothing option would be nice.’
NWS Forecaster – End of Week Survey

Recommendations for Operational Implementation

Based upon the evaluation of the GLM Background and Data Quality Product in the 2024 HWT Satellite Proving Ground, the following items have been recommended:

- **It is recommended that forecasters view the GLM Data Quality Product to become familiar with spatial and temporal variations in GLM flash detection efficiency, the locations of GLM subarray boundaries, and situations which may cause pixels approach saturation within their local forecast area.** Familiarization may occur prior to a convective event through training, or during the event with a situational awareness display.
- **It is strongly recommended that training regarding GLM flash detection efficiency, subarray boundaries, and flagged artifacts be provided with the GLM Data Quality Product to familiarize forecasters with their operational significance and the most common situations where these products may apply.**
- **It is recommended that new condensed display options of the GLM Background and the Data Quality Product be developed, which allow forecasters to observe significant values in GLM flash detection efficiency, quality flags, and subarray boundaries.** These displays may allow forecasters to conserve screen space and enhance situational awareness during severe convective events.
- **Is it recommended that the GLM Background imagery be leveraged in situations where ABI data are unavailable to observe synoptic scale cloud features and convection, along with mesoscale boundaries and features in more limited situations.** Optional displays to explore the impact of limiting imagery to 16 levels of brightness may be useful to explore the full applications of these data.

3.2 GREMLIN

The GOES Radar Estimation via Machine Learning to Inform NWP (GREMLIN, Hilburn et al. 2021) model was demonstrated for the first time the 2024 Satellite Proving Ground HWT. GREMLIN is a convolutional neural network which uses spatial information from three ABI bands (Band 7: 3.9 µm, Band 9: 6.9µm, and Band 13: 10.3µm) and the GLM Group Extent Density to produce an estimation of MRMS composite reflectivity across the ABI Full Disk scene for daytime and nighttime scenarios. The model is trained against MRMS composite reflectivity east of 105°W during the warm season, and its reflectivity estimates are constrained to values between 0 and 60 dBZ. GREMLIN was intended to provide a spatially uniform parameter for initializing convection with numerical weather models, however it is expected that forecasters may leverage GREMLIN reflectivity values in areas of little-to-no radar coverage. Additionally, convective features from satellite imagery associated with initiating convection may precede those from radar to increase lead time.

Within the 2024 Satellite Proving Ground, GREMLIN data were provided over the GOES-East and -West ABI Full Disk scenes, updating every ten minutes. The initial resolution of GREMLIN was set to 4 km, but after two days within the testbed the resolution was increased to 2 km. Before the experiment forecasters were provided a one-page QuickGuide, and two AWIPS procedures were created to help answer the R2O questions listed below. The first was a side-by-side comparison between GREMLIN composite reflectivity and MRMS composite reflectivity, and the second was a four-panel display with GREMLIN, GLM Flash Extent Density, MRMS composite reflectivity, ABI channels 7 and 9, and ABI channel 13 so forecasters could interrogate GREMLIN model inputs alongside its output and validation fields.

GREMLIN R2O Questions

- Does GREMLIN provide useful information to increase confidence for issuing warnings in areas lacking good ground-based radar coverage?
- Does GREMLIN provide timely information for initiating convection when compared to local radar?
- How closely does GREMLIN resemble observations from local radar?

Use of GREMLIN in the HWT

Forecasters used GREMLIN simulated composite reflectivity values (hereafter referred to as 'reflectivity') throughout the experiment to interrogate convection, validate estimates, and leverage these data in situations with little-to-no radar coverage. When comparing GREMLIN reflectivity to available MRMS composite reflectivity values within convection, approximately half (25/49) of participants responded in the daily surveys that estimates from GRELMIN were slightly less (5-10 dBZ) than MRMS. GRELMIN appeared similar (+/- 5 dBZ) to MRMS for around one-fifth (10/49) of all respondents. The product developer noted that GREMLIN output is constrained to values between 0 and 60 dBZ, which may lead GREMLIN to underestimate the maximum reflectivity within intense convection. Additionally, GREMLIN's ability to estimate composite reflectivity as a machine learning model can depend on the features present from the ABI and GLM imagery, and obscurations of these features by dense cirrus, convective cloud debris, anvils, or nearby convection may influence GREMLIN's reflectivity values.

‘The thunderstorm in the southwest is more isolated from other convective cloud debris, while also displaying a more intense thermal couplet and cloud top divergence signal than the thunderstorm to the northeast. It seems likely that this is the reason why the northeast storm has a lower ceiling for GREMLIN reflectivity than the southwest storm, even though the MRMS composite includes a 60 dBZ core in the northeast storm.’ [Figure 6]
22 May 2024, Blog Post: *Comparing GREMLIN Across Two Storms*
https://inside.nssl.noaa.gov/ewp/2024/05/22/comparing-gremlin-across-two-storms/

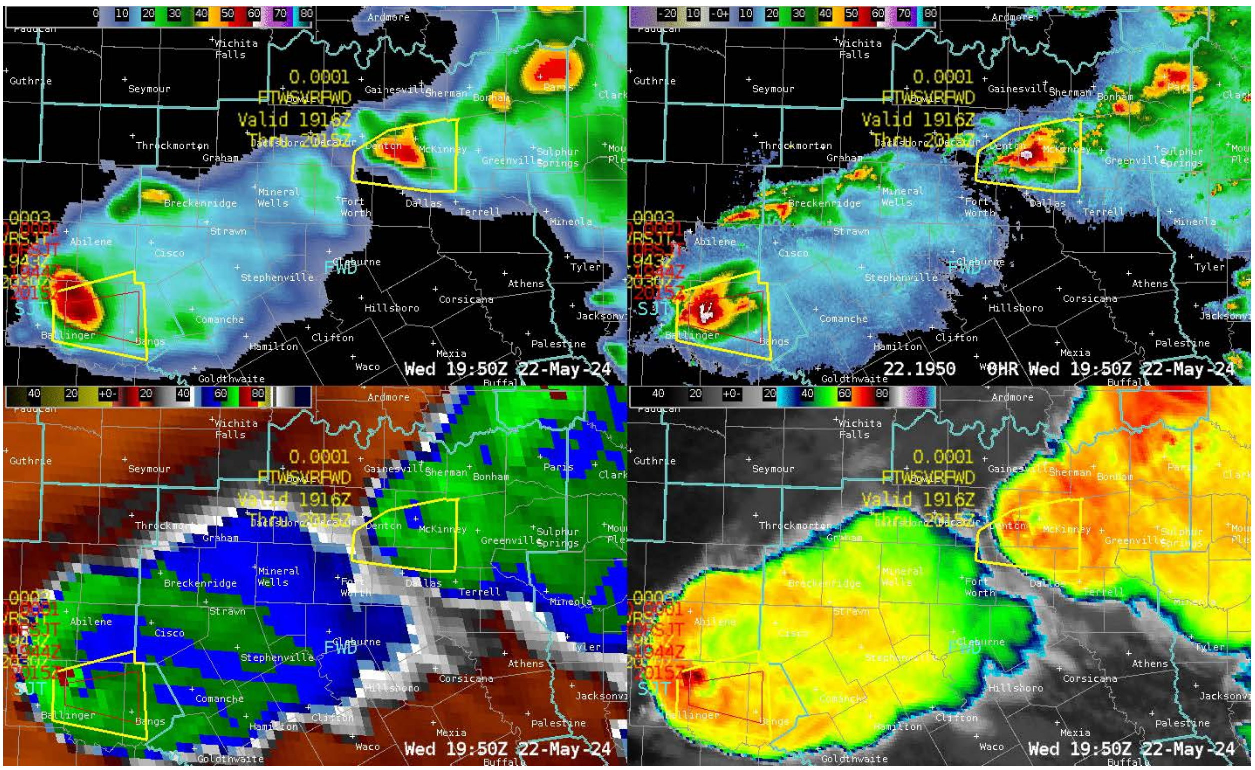

Figure 6: GREMLIN reflectivity (upper left), MRMS composite reflectivity (upper right), ABI water vapor IR imagery (lower left), and ABI ‘clean’ longwave infrared imagery from two severe-warned storms at 1950 Z on 22 May 2024. Animation.

Forecasters also compared GREMLIN’s output on the same storms from the GOES-East and GOES-West Full Disk scenes. Similar to imagery from the ABI, GREMLIN data for the experiment were not parallax-corrected and were noted by a few forecasters, but these concerns were limited. However, the differing perspectives of convection from the GOES-East and GOES-West ABIs meant that GREMLIN reflectivity values could differ when viewing the same storms. Several forecasters noted these differences and had discussions with the product developer to aid their interpretation of GREMLIN. Most often, these cases involved splitting supercells where the anvil from predominant supercell would overspread onto the other storm and obscure its view from satellite.

‘The storm split is apparent on the KMAF lowest slice and comp reflectivity, as early as 2020Z [Figure 7]. On GOES-E GREMLIN, the split does not start to show up until

2040Z, and is not readily apparent and clear until 2100Z…While GOES-E data masks the updraft of the left-mover under the anvil from the right-mover, GOES-W has a better view of the left-mover updraft…Comparing GOES-W GREMLIN and GOES-E GREMLIN, it's clear that GOES-W had the better view of the left-mover updraft, and picked up on the split much more accurately (though it was low on dBZ values). On the contrary, GOES-E GREMLIN did a much less consistent job in handling the right-mover.'
22 May 2024, Blog Post: *Developing and Splitting Cells in Sterling County TX*
https://inside.nssl.noaa.gov/ewp/2024/05/22/developing-and-splitting-cells-in-sterling-county-tx/

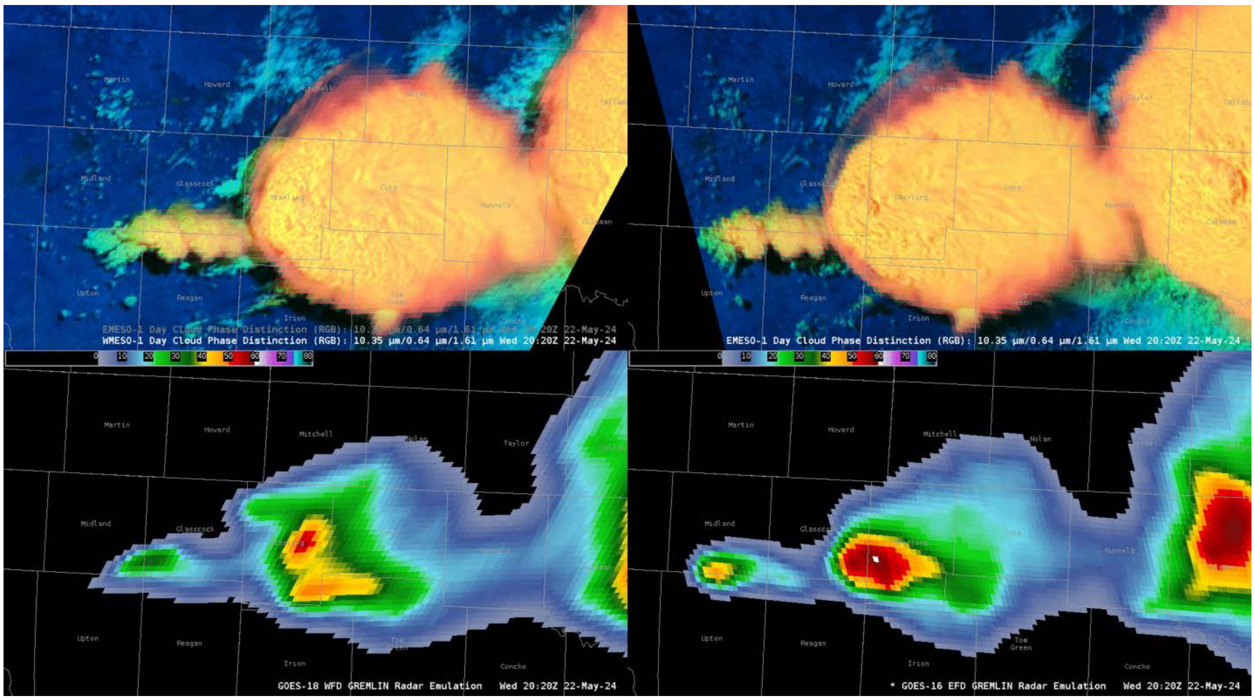

Figure 7: ABI Day Cloud Phase Distinction RGB and GREMLIN reflectivity from the GOES-West (left) and GOES-East (right) perspective at 2020 Z on 22 May 2024. Animation.

The apparent utility of GREMLIN changed throughout the experiment, depending on a forecaster's tasks, the coverage and intensity of convection, and radar data quality. From both the data-denial and organic scenarios where forecasters had little-to-no radar coverage, the participants noted in group discussions that they relied more heavily on GREMLIN reflectivity to monitor convective evolution and identify the most intense storms. This observation was further supported in the daily surveys, when forecasters were asked how they would describe their local radar coverage and how useful they found GREMLIN in their scenario (Figure 8). From the 55 participants who answered both questions, a majority who identified their radar coverage as 'Great' or 'Excellent' found GREMLIN ''Slightly useful' or 'Not at all useful'. On the other hand, when radar coverage was characterized as 'Poor' or 'Terrible/None' forecasters were more likely to view GREMLIN as more useful. Forecasters frequently commented that they felt comfortable interpreting outputs from GREMLIN because they were already keenly familiar with observing convection from radar, and when faced with little-to-no radar coverage GREMLIN appeared as an adequate substitution. These comments were supported in the weekly survey responses, as nearly all participants felt

GREMLIN was much more impactful in scenarios with little-to-no radar coverage for monitoring initiating convection (16/16) and issuing warnings (13/16).

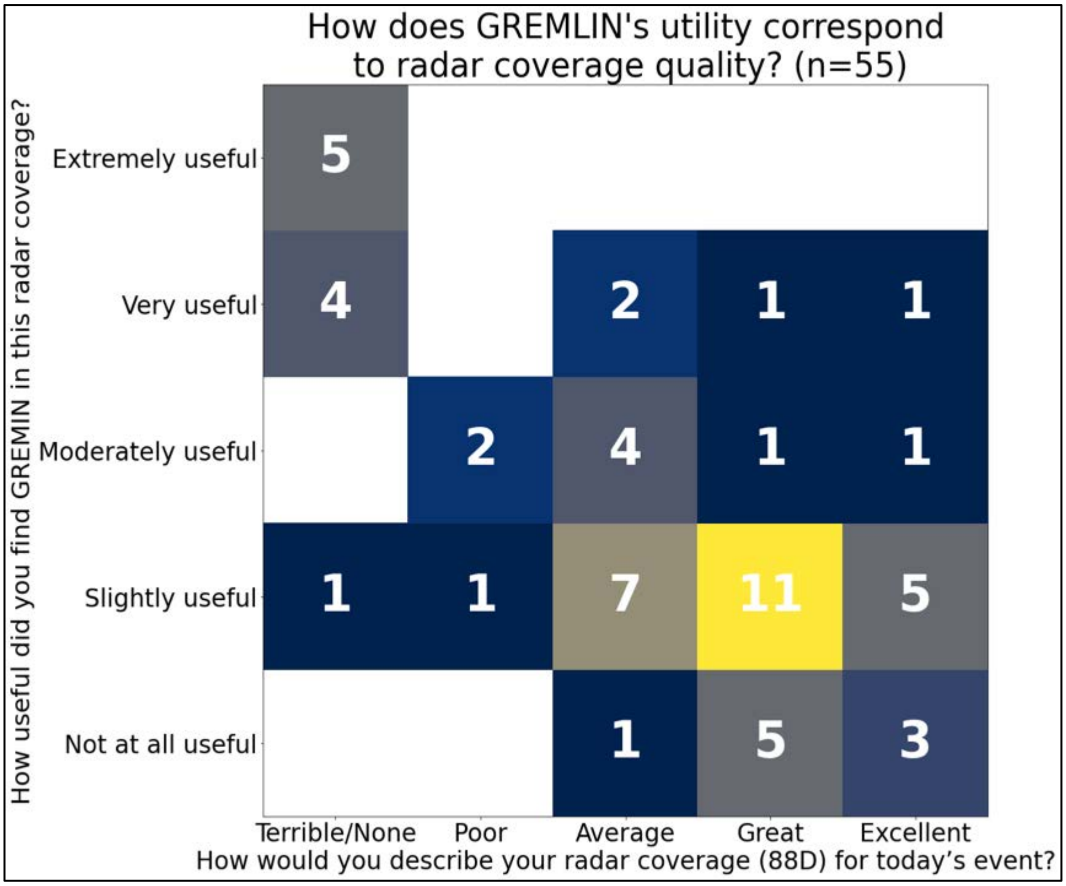

Figure 8: A two-dimensional histogram of daily survey responses (55) showing participant's apparent radar coverage quality and GREMLIN utility.

> 'Overall, GREMLIN performs fairly well with the zoomed out shape of the precipitation returns compared to radar [Figure 9], but perhaps a bit larger extent. Picks up on a few heavier cells more so versus heavier stratiform rain. Since the overall radar mosaic of GREMLIN looks so close to current radar, it would increase my confidence in using it for the big picture (especially if the radar data were to go out for some reason).'
> 20 May 2024, Blog Post: *GREMLIN Versus Radar*
> https://inside.nssl.noaa.gov/ewp/2024/05/20/gremlin-versus-radar/

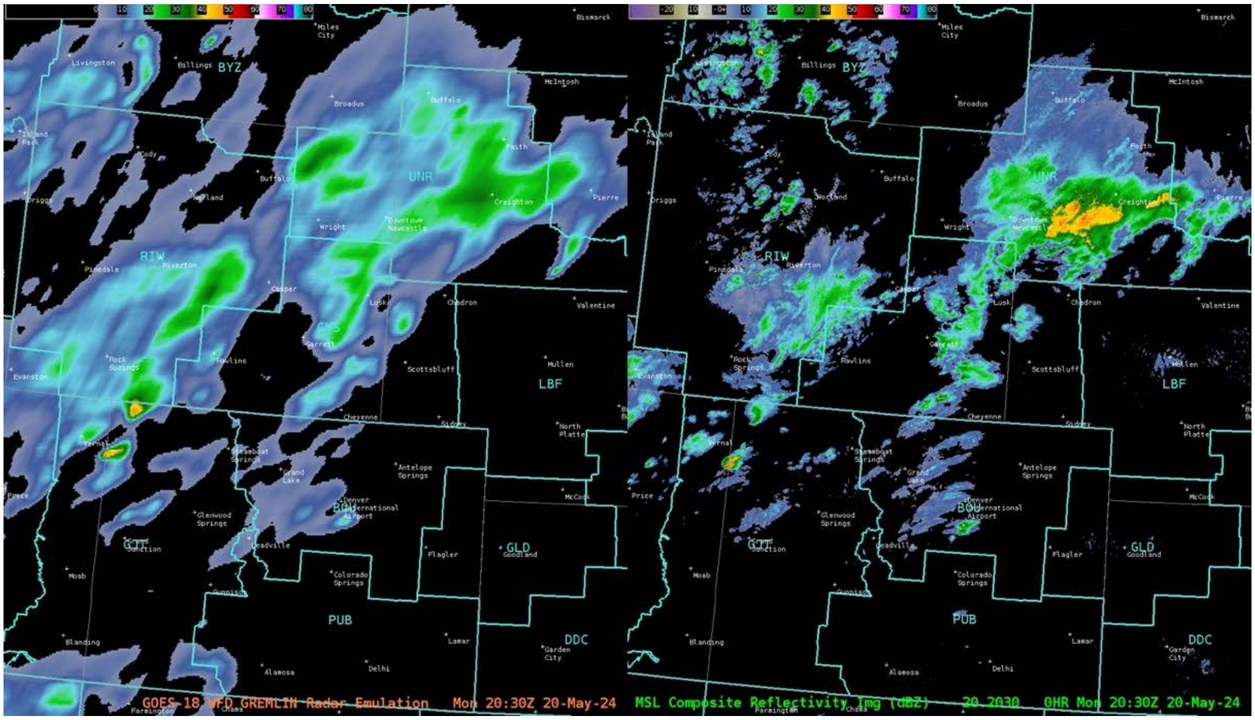

Figure 9: GOES-West GREMLIN reflectivity compared against MRMS composite reflectivity from 2030 Z on 20 May 2024. Animation.

As forecasters monitored initiating convection and made warning decisions, GREMLIN reflectivity was included in these analyses with mixed feedback. Over 90% of forecasters in daily surveys (31/34) noted that GREMLIN observed initiating convection with +/– 10 minutes of local radar, with over two-thirds (27/46) finding GREMLIN 'Slightly useful' or 'Moderately useful' when monitoring for initiating convection. When making warning decisions, forecasters stated inherent limitations using a composite reflectivity product such as needing altitude information to interpret updraft depth, width, and intensity. During the weekly surveys, forecasters were asked how confident they would feel using GREMLIN in an area with little-to-no radar coverage to monitor initiating convection, along with issuing severe thunderstorm and tornado warnings. A majority of participants (9/16) responded that they would feel 'Somewhat confident' (level three of five) issuing warnings with GREMLIN in these scenarios, while five more forecasters responded with 'A little confident' (level two of five). While a similar majority of forecasters (10/16) also felt 'Somewhat confident' with monitoring initiating convection, four responded with 'Very confident' (level four of five). Overall, GREMLIN was referenced as an influential product in 8 of 52 reported warning decisions.

> '[GREMLIN] provided information that I can not otherwise get from traditional satellite products. It even allows me to issue warnings with more confidence than I otherwise would with only satellite.'
> *Forecaster – End of Week Survey*

> 'I think GREMLIN has a long way to go before I'll feel confident in using it to issue warnings without supplementing it with other satellite data. But, it does provide a great base product to make decisions in a scenario with little to no radar coverage.'

Forecaster – End of Week Survey

‘I noticed that the GREMLIN Emulated reflectivity in Briscoe County was higher than the MRMS [Figure 10]. The GREMLIN emulation was producing a dBZ of 50+ while the MRMS was showing a 30 dBZ. To me, this difference was operationally significant as I would pay closer attention to the developing 50 DbZ feature than I would the 30 dBZ feature. So far in our experiment, this was the opposite of what I had experienced and I initially thought the emulated radar was wrong… My opinion changed 10 minutes later. Seen in image one, the MRMS with the next update showed a similar storm intensity to the GREMLIN emulation. This is impressive when one considers the latency of GREMLIN is greater (about 10 min) than the latency of MRMS (about 2 min). The GREMLIN product actually delivered a more operationally useful product sooner, despite having a greater lag.’
15 May 2024, Blog Post: *The Machine Learning Foretold the Future*
https://inside.nssl.noaa.gov/ewp/2024/05/15/the-machine-learning-foretold-the-future/

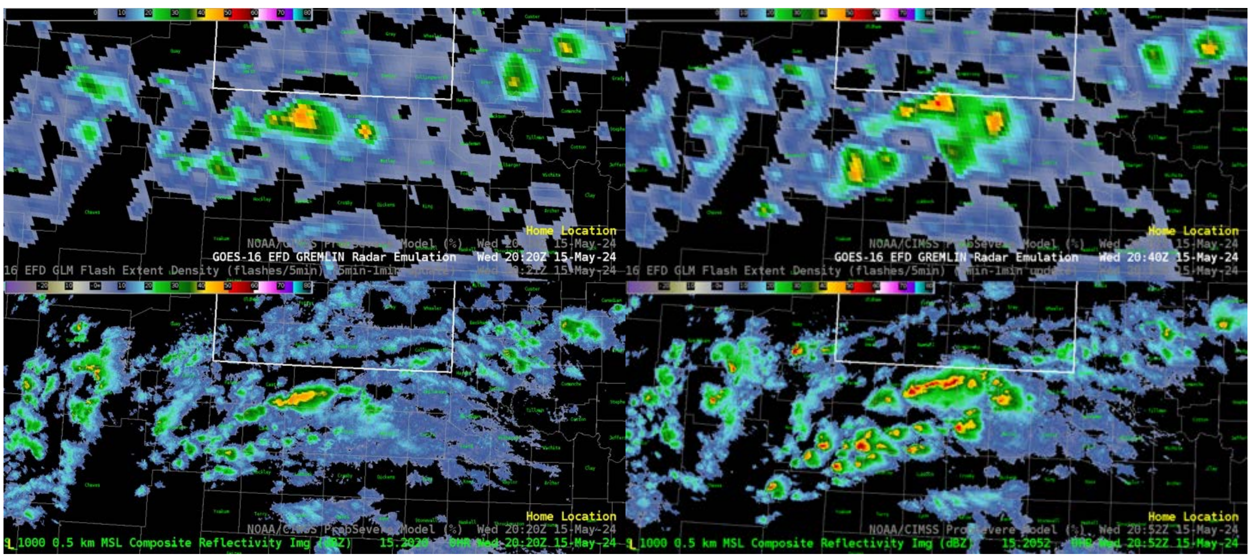

Figure 10: Comparisons of available GOES-East GREMLIN and MRMS composite reflectivity values from 2020 Z (left) and 2052 Z (right). [Animation](#).

During the first week of the experiment forecasters noticed considerable latency in the GREMLIN product of up to twenty minutes from the previous file. Even though GREMLIN is produced over the ABI Full Disk domain in ten-minute increments, the ABI bands used by GREMLIN take approximately ten minutes to scan the entire scene to be available for GREMLIN, which introduces an effective 20-minute delay from the previous GREMLIN reference time until the next data are made available. Forecasters across all weeks suggested in group discussions that GREMLIN produced on the ABI CONUS or Mesoscale scenes may decrease its effective latency and increase the utility of the product and appeared to be the most frequent idea for improvement. Forecasters also suggested GREMLIN output tailored to other radar products such as maximum expected size of hail, layered reflectivity, and echo tops. Additional model inputs from ground-based lightning networks and the OCTANE divergence and cloud-top cooling fields were also suggested, along with retraining GREMLIN in different weather scenarios (weak convection, rainfall estimates, winter weather, etc.) to increase its applications in other operational scenarios.

‘Would love if 5-min GREMLIN data could be made available, but only if it were not overly jumpy. Sometimes even the 10-min data was jumpy and struggled with consistently depicting storm mode.’
Forecaster – End of Week Survey

‘Training GREMLIN on some other products like vertically integrated ice or MESH could be interesting and help with thresholds in the GREMLIN for warning issuance.’
Forecaster – End of Week Survey

‘I would be interested to see it more often in radar denial cases or in mountain region applications where beam blockage is an issue.’
Forecaster – End of Week Survey

Recommendations for Operational Implementation
Based upon the evaluation of the GREMLIN model in the 2024 HWT Satellite Proving Ground, the following items have been recommended:

- **It is strongly recommended that GREMLIN imagery be produced on the ABI CONUS and Mesoscale scenes, to evaluate the impact of more rapidly updating information when monitoring initiating convection and its evolution.** Additional areas of investigation may include the production of other convection-focused MRMS products, or model inputs such as ground-based lightning networks or cloud-top divergence fields.
- **It is recommended that forecasters use GREMLIN in regions lacking adequate ground-based radar coverage to interpret the intensity, coverage, and evolution of mid-latitude deep convection.**
- **It is recommended that GREMLIN training include information such as the model’s training dataset, coverage, and latency.** Known strengths, biases, and limitations related to GREMLIN’s training dataset and applications may assist operational implementation.

3.3 OCTANE

The Optical flow Code for Tracking, Atmospheric motion vectors, and Nowcasting Experiments (OCTANE) products (Apke et al. 2022) were demonstrated again in the 2024 Satellite Proving Ground HWT. Near-pixel level brightness motions in cloud textures from the ABI visible band (0.64 μm) during the day and the clean-infrared band (13.3 μm) at night are calculated using the 1-minute and 30-second mesoscale scenes available from the GOES-R series ABI. Magnitudes are calculated in m s^{-1}, a 5-minute median filter is applied to mitigate jitter signals, and new outputs are available with a latency of 2.5 to 5 minutes. OCTANE speeds and directions are calculated using the visible (infrared) band when the solar zenith angle is less than (greater than) 80 degrees. A Hue Saturation Value (HSV) display method is used to visualize the OCTANE speed and direction products, creating 'sandwich' imagery that leverages reflectance (brightness temperature) values from the ABI visible (infrared) band. Based on forecaster feedback from the 2023 experiment, Cloud-Top Divergence (CTD) and Cloud-Top Cooling (CTC) fields were derived from time trends in OCTANE horizontal wind retrievals and ABI infrared imagery. Both products are also produced as a sandwich with visible imagery and attempt to provide more concise cloud-top information when compared to the OCTANE speed and direction fields.

The suite of OCTANE products provides wind information at cloud top within each mesoscale scene, highlighting environmental shear, and cloud-top divergence (cooling) from developing and mature convection. For the 2024 experiment, all the OCTANE sandwich products were made available in AWIPS through procedures created by the developers, and forecasters were encouraged to optimize their display within their current convective environment. The CIRA SLIDER webpage also provided a default view of the OCTANE speed sandwich, CTC and CTD sandwich, and a wind barbs product.

OCTANE R2O Questions

- Do divergence and cloud-top cooling products help you identify stronger versus weaker updrafts over initiating and mature deep convection over speed/direction alone?
- How does using IR-only optical flow calculations impact your perception of these products, and are they still useful for your operations?
- How often would you use these products if made available at your WFO?

Use of OCTANE in the HWT

Within the testbed, forecasters used OCTANE in all phases of convection, and in tandem with satellite, radar, and mesoanalysis data, to interrogate convective updraft strength and rapidly evolving mesoscale environments. In the daily surveys forecasters frequently reported using satellite imagery, lightning data, and radar data alongside the OCTANE products. Mesoanalysis fields were also leveraged from the SPC and several models including the HRRR and PHS. These fields were most often related to vertical wind shear (e.g. 0-1km, 0-3km, 0-6km) and convective instability (e.g. CAPE, CIN). Confirmation of OCTANE wind speeds and direction against mesoanalysis wind fields improved confidence in the OCTANE products themselves, while complimentary signals between OCTANE and the other products improved confidence in thunderstorm intensity. During the daily discussions, participants from each week also remarked about how the satellite-based signals and trends from the OCTANE product would lead complimentary radar-based signatures and trends. This bolstered forecaster confidence in thunderstorm behavior along with making warning and DSS messaging decisions.

> ‘A strong storm with a well defined mid-level mesocyclone entered the western portion of the MKX CWA at around 4:00 PM CDT. At the time, the OCTANE speed product showed a well defined gradient and the divergence product showed fairly high values, indicating that the updraft was quite strong. We decided to issue a severe thunderstorm warning with a tornado possible tag on this cell as a result… In all, the OCTANE products seem to be very useful in assessing the strength of a storm’s updraft.’
> 3 June 2024, Blog Post: *OCTANE Speed Product Shows Weakening Trend in Storm Well*
> https://inside.nssl.noaa.gov/ewp/2024/06/03/octane-speed-product-shows-weakening-trend-in-storm-well/

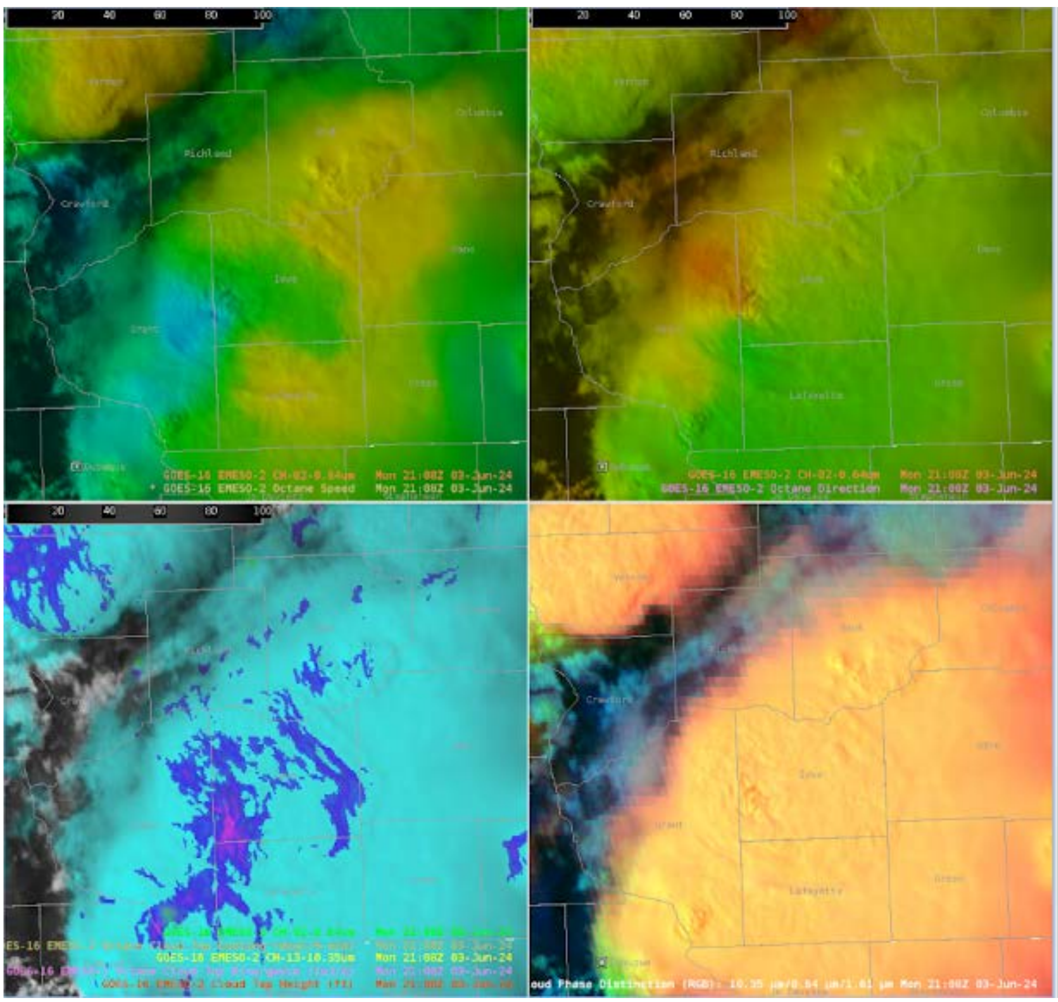

Figure 11: OCTANE Speed (upper left), Direction (upper right), Cloud-Top Cooling (lower right) overlaid on Cloud-Top Divergence (lower left), and ABI Day Cloud Phase Distinct RGB (lower right) at 2100 Z on 3 June 2024.

> ‘The OCTANE wind and speed gradients provided substantial confidence in warning issuance given the meager radar coverage. Both on strengthening convection and in monitoring weakening convection.’

Forecaster – End of Day Survey

‘Octane speed was probably the most useful product today. Did a great job of highlighting the min speed on the upshear side of updraft. This was about the same time a TBSS was noted in radar reflectivity.’
Forecaster – End of Day Survey

When applying the OCTANE products during their daily shift, nearly 90% (54/61) identified thunderstorm updraft trends through cloud-top divergence signals, the most out of all recommended uses. The next highest categories were identifying initiating convection in sheared (28/61) and unsheared (17/61) environments, along with recognizing regions of mesoscale ascent in strong low-level shear (16/61). In more isolated cases forecasters also noted features related to supercells such as updraft splitting and inflow notches. When issuing severe thunderstorm or tornado warnings, a majority of forecasters (33/52) noted that OCTANE influenced their warning decisions. They most often cited signatures in the cloud top speed and direction products such as divergence or above-anvil cirrus plumes, followed by CTC and CTD products. Additionally, OCTANE was cited as a factor in the forecaster’s DSS messaging for over one-third of all messages sent (19/53). Forecasters reported most often using the CTC and CTD products to anticipate thunderstorm trends and subsequent hazards, along with their communication to the simulated partners.

‘A combination of cloud top cooling in OCTANE and subsequent divergence aloft was a helpful clue in assessing the potential of a storm that was distant from the radar [Figure 12]. It was caught a little later in analysis, but OCTANE proved helpful in diagnosing the storm and deciding to pull the trigger.’
21 May 2024, Blog Post: *Warning vs DSS: WFO DMX*
https://inside.nssl.noaa.gov/ewp/2024/05/21/warning-vs-dss-wfo-dmx/

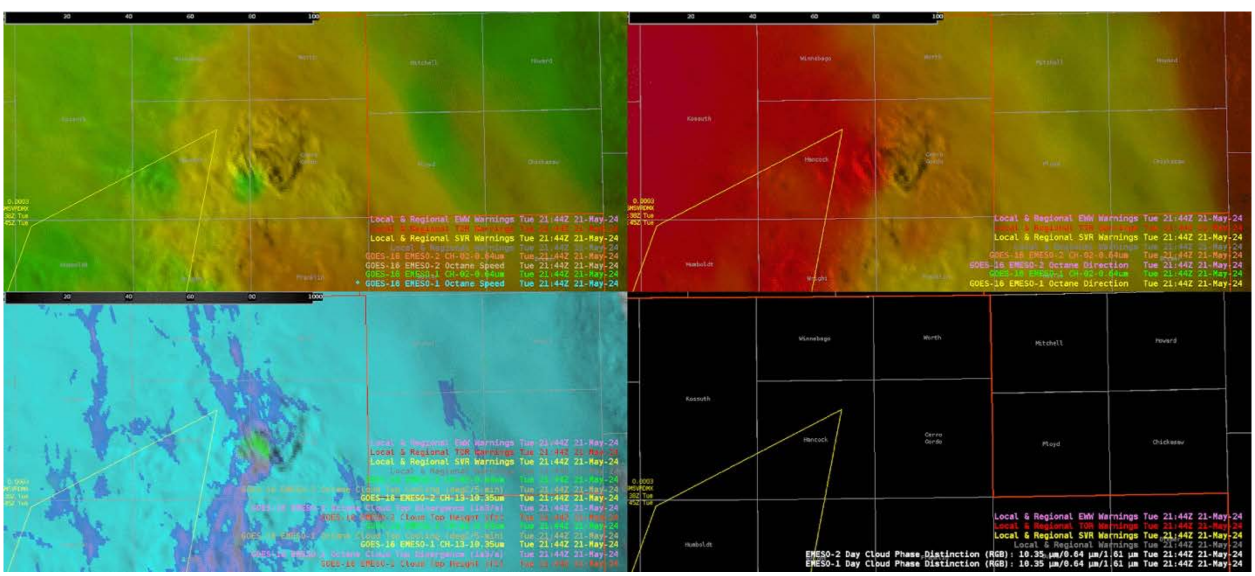

Figure 12: OCTANE Speed (upper left), Direction (upper right), and Cloud-Top Cooling overlaid on Cloud-Top Divergence and (lower right) at 2144 Z on 21 May 2024. Animation.

'learning to use the CTC and CTD portion of OCTANE – it's been helpful to keep track of the newer updrafts (shows up really well to the west with the outflow, but also in the green where there is more anvil/cirrus overflow). Also, it's been a great tool to see where storms are maintaining strength (where the divergence signature in the pinks/purples hold on)… These products have aided in confidence in warning ops, especially with lack of primary radar.'
22 May 2024, Blog Post: *OCTANE…The GOAT for Warning Ops Today!*
https://inside.nssl.noaa.gov/ewp/2024/05/22/octane-the-goat-for-warning-ops-today/

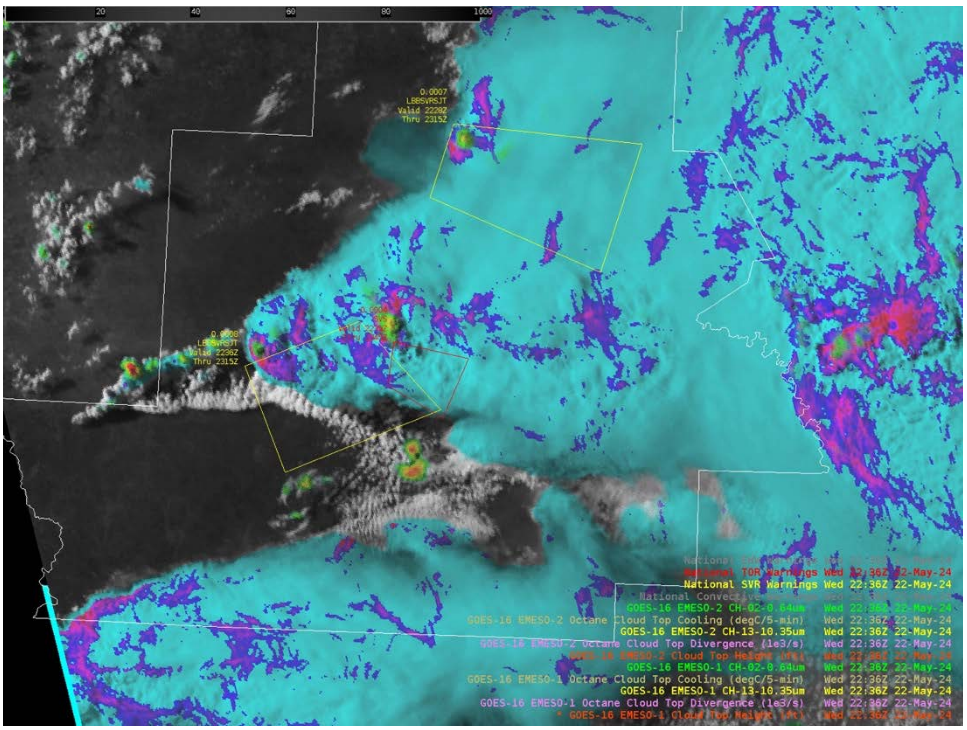

Figure 13: OCTANE Cloud-Top Cooling overlaid on Cloud-Top Divergence at 2236 Z on 22 May 2024. Simulated severe thunderstorm warnings (yellow) and tornado warnings (red) are also overlaid. Animation.

Several AWIPS display procedures were made for the forecasters to quickly leverage the OCTANE product, and they provided feedback on these displays along with avenues for display improvement and customization. Similar to the 2023 demonstration of OCTANE speed and direction at the HWT, forecasters consistently adjusted the color bar ranges on the speed product. This modification allowed the forecasters to more easily pick out gradients in speed, often related to cloud top divergence, and tune the product to the ambient deep-layer shear. When demonstrating the OCTANE CTC and CTD fields, forecasters provided mixed reactions to the developer's recommendation of overlaying the CTC field over the CTD field. Some forecasters wanted to have the CTC displayed separately from CTD, or in a few cases switch the color bar tables between the two products. Lastly, forecasters worked with the OCTANE developers to introduce more texture in the nighttime OCTANE speed and direction products. This involved modifying the infrared "value" layer for each sandwich in the hue-saturation-value settings to highlight cold overshooting tops in darker colors with the clean-infrared ABI band. The default setting for this value layer is intended to increase the brightness with colder temperatures. Nearly all forecasters who experimented with the modified value layer, however, felt it improved their ability to distinguish cloud-top textures at night near regions of cloud top divergence.

> 'OCTANE products were specifically helpful, especially by modifying the colorbar settings for OCTANE Speed. Decreasing the MAX from 200 to 100 and increasing the MIN from 0 to 15 gave a greater contrast and "bullseye" to help diagnose strengthening divergence… Several additional DSS notifications were sent to the site to alert them of not only the approaching activity, but how long the activity might last over the next following hour… OCTANE gave great situational awareness to support alongside with radar to lead to proactive warning decision.'
> 4 June 2024, Blog Post: *Monitoring Convective Trends Across WFO MPX*
> https://inside.nssl.noaa.gov/ewp/2024/06/04/monitoring-convective-trends-across-wfo-mpx/

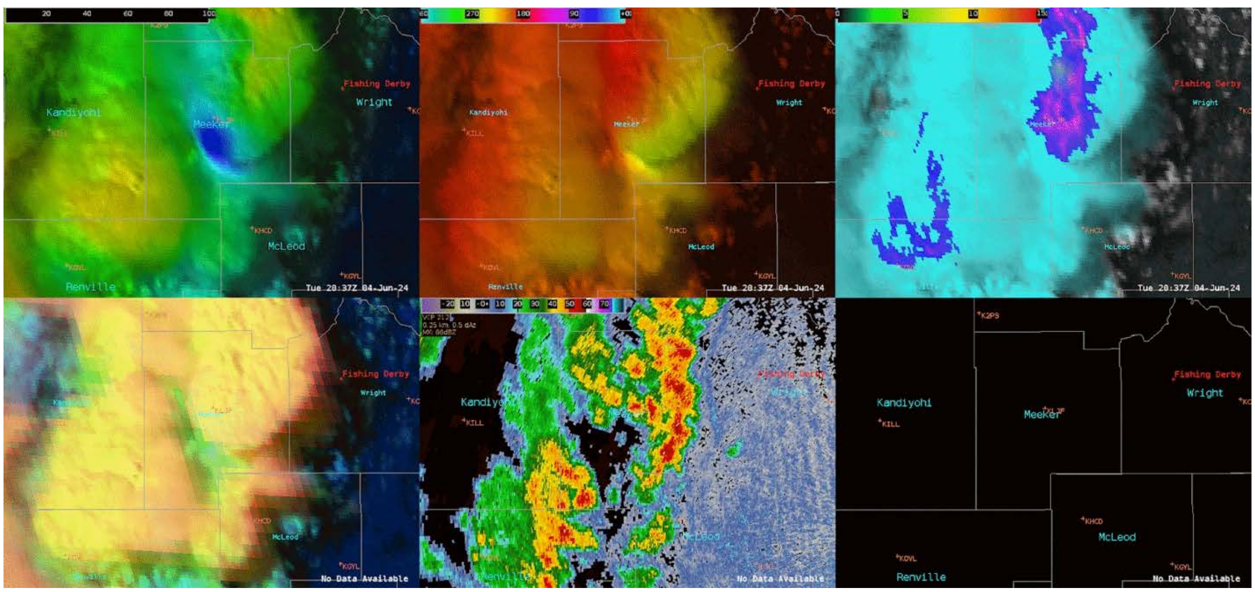

Figure 14: OCTANE Speed, Direction, CTC, and CTC (top row) at 2037 Z on 4 June 2024 along with ABI Day Cloud Phase Distinction RGB and MPX Base Reflectivity (bottom row). Animation.

Throughout the experiment, forecasters were also asked to compare the intended applications of the OCTANE speed and direction products from the visible (daytime) and infrared (nighttime) retrievals. While the nighttime OCTANE fields were available even during the daytime, forecasters viewed these products less often in operations and therefore they received less feedback in the daily surveys compared to the daytime products. In the end of week survey, forecasters were asked how often they would use each OCTANE product if they were made available at their WFO. For both daytime and nighttime products, a majority of forecasters responded with highest three categories of 'All of the time', 'Most of the time', or 'Half of the time', while only the daytime products received a majority of 'All of the time' or 'Most of the time' responses. This sentiment was supported in a follow-up question on the weekly survey, as well as daily discussions. Forecasters expressed utility in the daytime and nighttime products, but overall felt the daytime products provided more detailed information from the higher-resolution optical wind retrievals. One exception to this statement was the participants' reaction to the nighttime direction product during ongoing convection. A handful of participants noted stronger gradients in direction at cloud top in the nighttime direction product, which were easier to observe and monitor when compared to the daytime direction product.

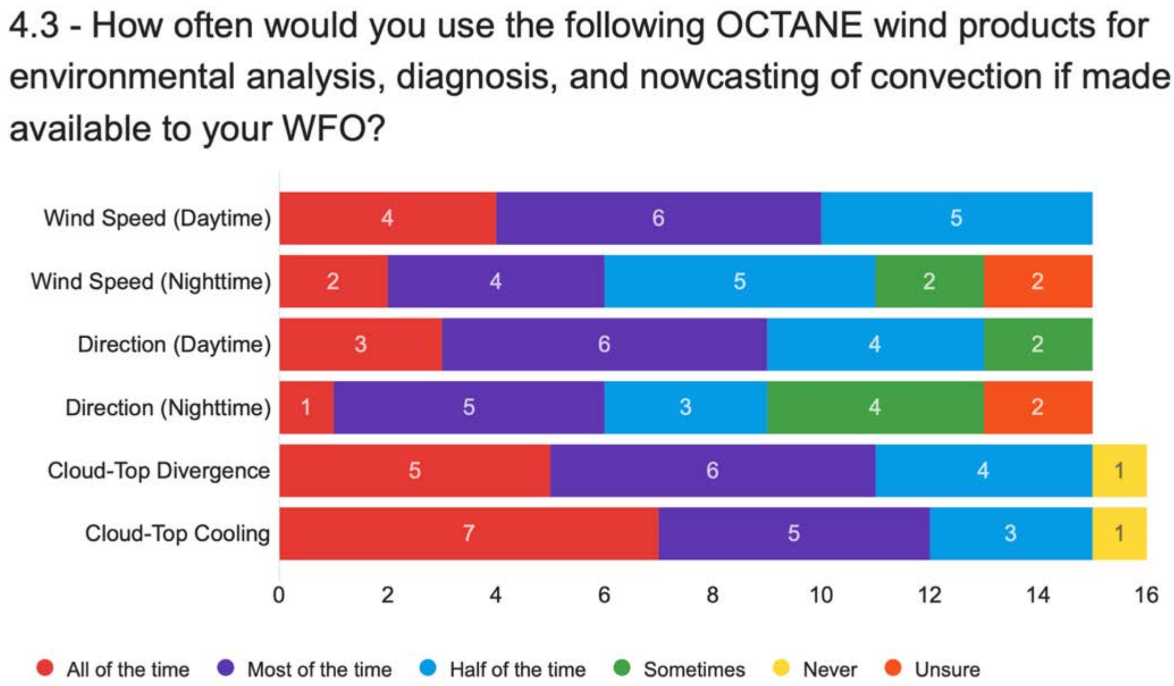

Figure 15: End of Week Survey responses from forecasters when asked how often they would use the OCTANE products if available at their WFO (n=16).

> ‘Although some amount of fine detail was lost, the nighttime product actually provided more pronounced depictions of the changes in speed/direction. Thus, I think the nighttime product proved very useful, almost as useful as the daytime version.’
> *Forecaster – End of Week Survey*

> ‘There was a huge loss of detail with the clouds when going to the nighttime OCTANE products. Everything became merged and it was harder to pick out details of distinct storms or MCS. Changing the IR overlay colormap helped greatly to define some of the details that were lost.’
> *Forecaster – End of Week Survey*

> ‘The day version is [top row], and the night version is [bottom row] [Figure 16]. There is detail in the day version that does not carry through to the night version. Of note: the anvil-top cirrus (slower motions) do not show up on the night version. Also, the magnitudes of the north and south lobes of higher speeds are muted somewhat on the night version…Also want to note that the night version of the direction product (bottom right) shows a more significant shift in direction than the day version (top right).’
> 22 May 2024, Blog Post: *Comparing OCTANE Day and Night Products with KSJT Convection*
> https://inside.nssl.noaa.gov/ewp/2024/05/22/comparing-octane-day-and-night-products-with-ksjt-convection/

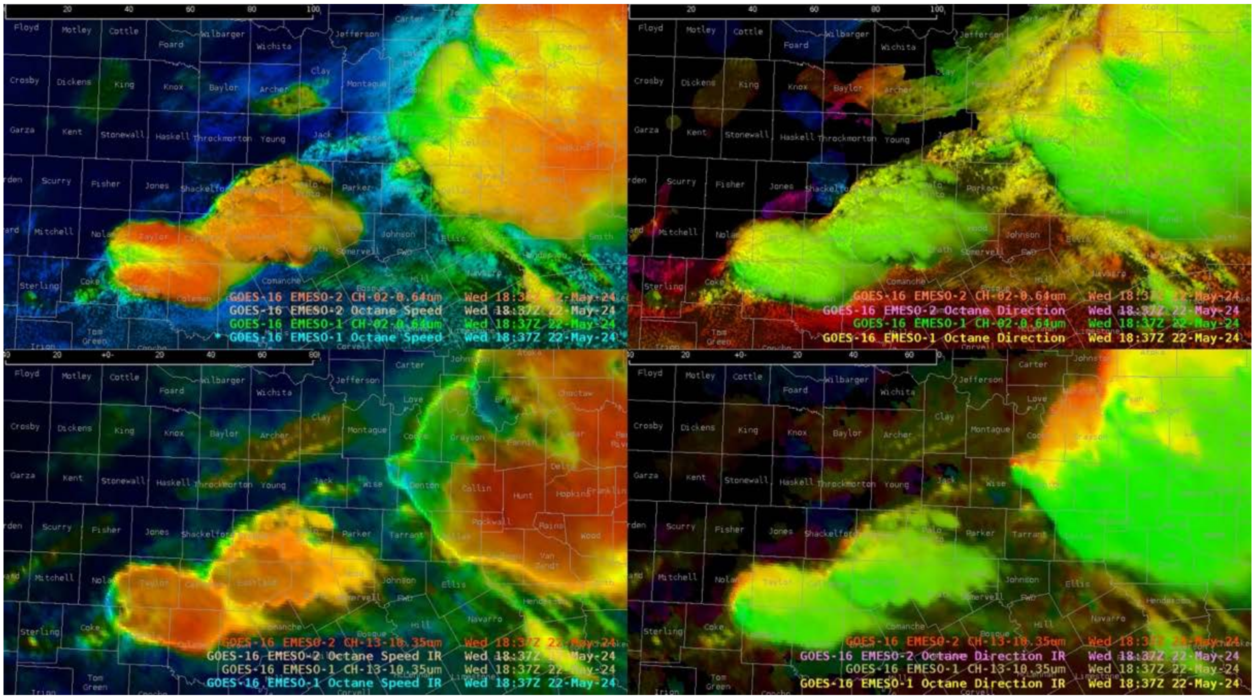

Figure 16: Daytime OCTANE Speed (upper left) and Direciton (upper right) compared against the nighttime Speed (lower left) and Direction (lower right) products at 1837 Z on 22 May 2024.

Forecasters provided mixed reactions when asked on their preference between the base-level OCTANE products (speed and direction) and the derived fields of cloud-top cooling and divergence (CTC and CTD). Forecasters were also asked after each shift how confident they felt using the OCTANE wind speed and direction products when compared to the derived cloud top divergence product, and these results were also mixed. Nearly identical proportions of forecasters felt more confident in the speed and direction products (23/57) when compared to those who felt more confident in the CTD product (22/57), while over 20% (12/57) felt the same level of confidence between the two products.

When monitoring initiating convection, forecasters overall found the CTC product the most useful, with approximately 87% of forecasters surveyed (48/55) responding with the two highest levels of utility 'Very useful' and 'Extremely useful' in the daily survey. Most forecasters (33/56) also felt the CTD product was 'Very useful' or 'Extremely useful' when monitoring initiating convection. Participants stated in daily discussions and blog posts that the appearance of strong cooling/divergence from the CTC/CTD products often coincided with radar and satellite imagery signatures of initiating deep convection, increasing their confidence in the derived products and their understanding of the convective environment. Similarly, when asked about each product's utility in monitoring ongoing convection a majority of forecasters selected 'Very useful' or 'Extremely useful' for wind speed (daytime), direction (daytime), CTD, and CTC.

Recommendations for Operational Implementation

Based upon the evaluation of the OCTANE suite of products in the 2024 HWT Satellite Proving Ground, the following items have been recommended:

- **It is highly recommended that NWS forecasters leverage the OCTANE Speed, Cloud-Top Divergence, and Cloud-Top Cooling fields to monitor initiating convection, trends in updraft strength through rapid cooling and cloud-top divergence signatures, and regions of ascent along mesoscale boundaries.** These signatures can increase forecaster confidence in thunderstorm intensity trends when making convective warning decisions. Additional applications include monitoring inflow notches and updraft splitting in supercells.
- **It is strongly recommended forecasters use the OCTANE Cloud-Top Cooling product to identify initiating convection and increasing updraft intensity through rapid changes in cloud-top temperature.** These data may be overlaid with the Cloud-Top Divergence product or individual ABI bands.
- **It is recommended that NWS forecasters leverage the nighttime (infrared imagery-based) OCTANE Wind Speed and Direction retrievals similar to the daytime (visible imagery-based) Wind Speed and Direction retrievals when monitoring initiating convection and cloud-top divergence signatures.** Training may be required to better calibrate forecasters to these signatures at night, such as the increased gradients in the Direction product and changes in resolution.
- **It is recommended that additional display modification to the OCTANE Cloud-Top Divergence and Cloud-Top Cooling products be explored to increase their potential utility in NWS operations.** Suggestions include varying scales, gradients, and colors to more readily identify cloud-top divergence with respect to severe weather potential. Reducing the visibility of artifacts such as around thin cirrus or cloud tops moving over snow covered terrain may also be considered.

3.4 PHS Model

The Polar Hyperspectral Soundings and Microwave Data and ABI (PHS) Model was demonstrated in the 2024 Satellite Proving Ground HWT, building upon lessons learned from the 2022 and 2023 experiments. Hyperspectral infrared sounders and microwave imagers on the NOAA-20/-21 and MetOp-B/-C satellites in low-Earth orbit collect temperature and moisture information at various levels, which are then associated with data from the GOES-R ABI. These fused observations are stepped forward in time and used to create soundings which are assimilated into a 4 km Weather Research and Forecasting (WRF) model (Powers et al. 2017). The PHS model is initialized every hour, with hourly output available out to six hours, and covering nearly all the central and eastern United States. The numerical model leverages the high spectral resolution of the hyperspectral infrared sounders and microwave sounders in low-Earth orbit with the high spatiotemporal resolution of the ABI and is expected to provide more accurate forecasts of the mesoscale environment coincident with deep convection. Increased accuracy of the moisture distribution within a profile is anticipated to provide most of the anticipated improvements.

In 2023, passive microwave soundings were included in the PHS model framework to improve retrievals in cloudy scenes, along with the routine updating of observational error covariance and hydrometeor fields. Numerous model fields were included in AWIPS-II to allow more direct comparisons to data from SPC mesoanalysis and the operational High-Resolution Rapid Refresh (HRRR) model. These included instability parameters, vertical wind shear, composite indices, and thermodynamic variables at 850, 700, and 500 mb. Data flow disruptions and unexpected model latency limited the ability for a thorough evaluation in the 2023 experiment, driving the need for a follow-up evaluation in 2024. Prior to the testbed, participants were provided with a training video on the PHS model, along with an instructional video and one-page document describing how to compare the PHS and Rapid Refresh (RAP) model outputs online. Forecasters could access the PHS model products in AWIPS-II, with a larger suite of model output available online in a web display.

PHS Model R2O Questions

- Does the assimilation of hyperspectral sounding data into numerical models provide timelier and more useful nowcasts of severe weather than the traditional assimilation of observations into operational convective allowing models?
- Do the PHS observations and numerical model nowcast/forecast products help forecasters in the issuance of tornado warnings?
- How does including microwave soundings from cloud-to-partly cloudy scenes impact PHS model performance when compared to model performance in clear scenes?

Use of PHS in the HWT

Forecasters used the PHS model throughout the experiment to anticipate the location, timing, and evolution of convection, and provided feedback on model performance, output characteristics, and display needs. Most often the PHS model contributed to the forecasters need for mesoscale data in the pre-storm environment. When asked in the daily surveys what PHS model fields the forecasters used during their shift, instability products (MU-CAPE, ML-CAPE, ML-CIN, etc.) were most frequently mentioned, followed by model composite reflectivity, composite indices (STP, SCP, SHIP, etc.), and shear (0-1 km, 0-6 km, SRH, etc.). Traditional datasets for mesoanalysis from the HRRR, RAP, SPC Mesoanalysis, radar, surface observations, and satellite imagery were paired with information from the PHS to validate and interpret the mesoscale environment related to convection. These tasks most often consolidated around identifying potential regions for convective development, storm mode, and storm hazards. Forecasters also were asked at the end of each shift which model forecast hours they frequently viewed, and approximately 87% (39/45) viewed forecast hours 2-3. The next highest category was forecast hours 0-1 with 56% (25/45) and hours 4-5 with 40% (18/45).

> ‘Here we compare convective initiation in the PHS model [left] versus the HRRR [center]. Both models have relatively similar SBCAPE (around 1000-20000 J/kg) but the HRRR does carry CIN longer through the afternoon. The 16Z (and 17Z, not shown) PHS model produces convection by 21Z to 22Z, while the 17Z HRRR waits until 23Z. [Right] is the 1854Z GOES-East day cloud phase RGB which shows some shallow cumulus growth in western to north central areas of the CWA. This may lead to more confidence in the earlier convective development from the PHS model.’ [Figure 17]
> 23 May 2024, Blog Post: *Convective Initiation Timing Between the PHS and HRRR*
> *https://inside.nssl.noaa.gov/ewp/2024/05/23/convective-initiation-timing-between-the-phs-and-hrrr/*

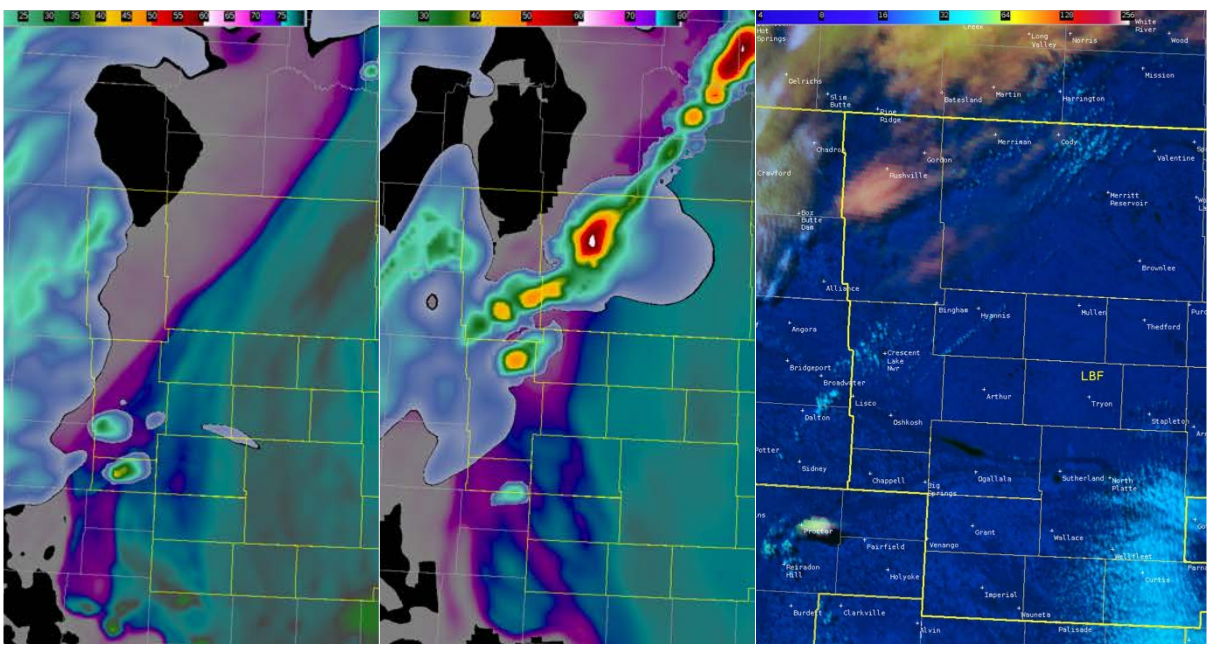

Figure 17: Model output at 22 Z from the 17 Z HRRR (left) and 16 Z PHS (center) models on 23 May 2024 in Nebraska with simulated composite reflectivity overlaid on MU-CAPE. Verification is provided from the GOES-East ABI Day Cloud Phase Distinction Imagery at 1854 Z.

> ‘In simulated composite reflectivity (bottom right), there is an indication of a supercell (perhaps right-moving) tracking eastward toward the CO/NE/KS triple point. PHS indicates that this supercell will be tracking into an increasingly favorable environment, and by 02Z, it will be entering an area with much larger CAPE and a nearby local maxima in STP.…To me, the PHS data provides additional confidence in the potential

for this storm to become quite strong a few hours from now. Based on mesoanalysis, this storm could be capable of producing all hazards, with hodographs (SPC meso) and STP (both SPC meso and PHS) suggesting a tornado threat is absolutely there. I would probably use this information to start adjusting messaging, in sort of that in-between watch-and-warning paradigm.' [Figure 18]
20 May 2024, Blog Post: *PHS Forecast for Supercell in Northeast Colorado*
https://inside.nssl.noaa.gov/ewp/2024/05/20/phs-forecast-for-supercell-in-northeast-colorado/

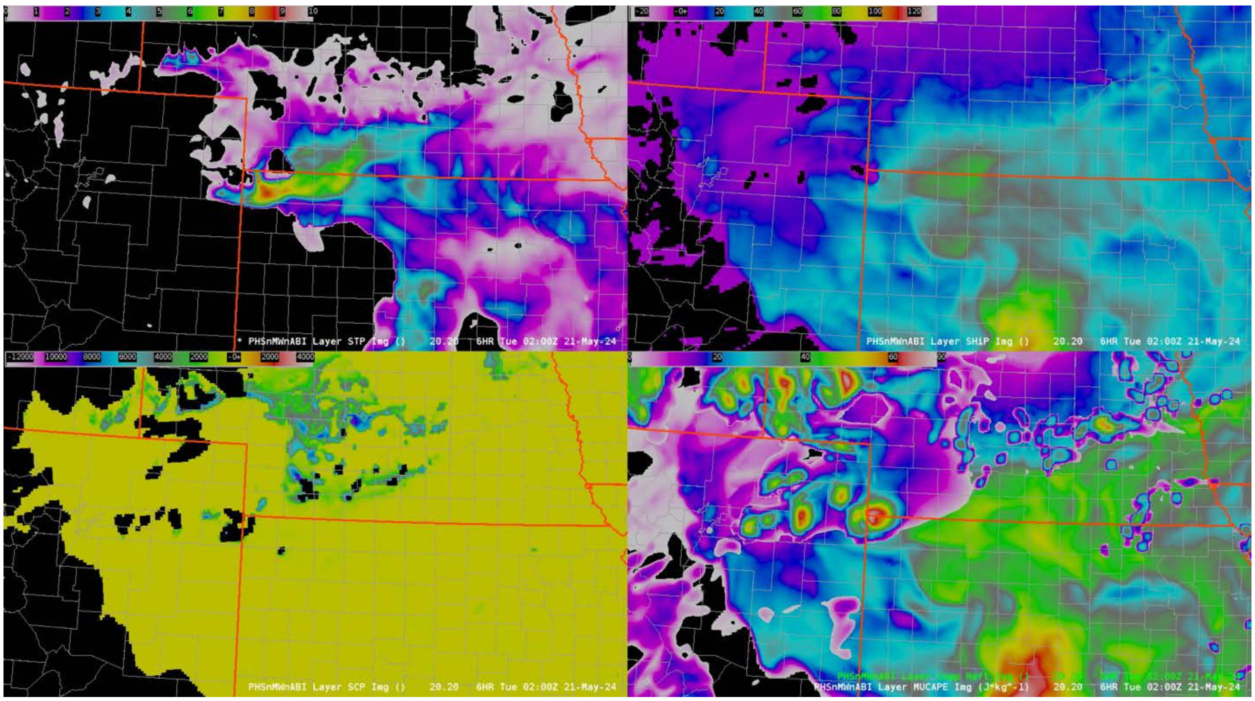

Figure 18: Convective indices from the 20 Z PHS model run at forecast hour 6 (2 Z) on 21 May 2024 in northeast Colorado, northwest Kansas, and southwest Nebraska. Fields displayed include STP (upper left), SHiP (upper right), SCP (lower left), and simulated composite reflectivity overlaid on MU-CAPE. Animation.

'Was looking at the tornado probability for the storm moving out of Runnels County since the midlevel meso was strengthening [Figure 19]. So, I took a look at the environmental parameters from PHS (namely, the SRH). It looked like the storm was ingesting 800-900 m2/s2 SRH, which is a bit high, but the idea that the storm was ingesting more SRH/modifying the environment to possibly increase the tornado potential was useful. Interestingly, the RAP from the SPC Mesoanalysis page shows an area of enhanced 0-3 km SRH (just to the NW of the storm).'
22 May 2024, Blog Post: *PHS Environmental Parameters vs RAP and Radar for Warning Ops*
https://inside.nssl.noaa.gov/ewp/2024/05/22/phs-environmental-parameters-vs-rap-and-radar-for-warning-ops/

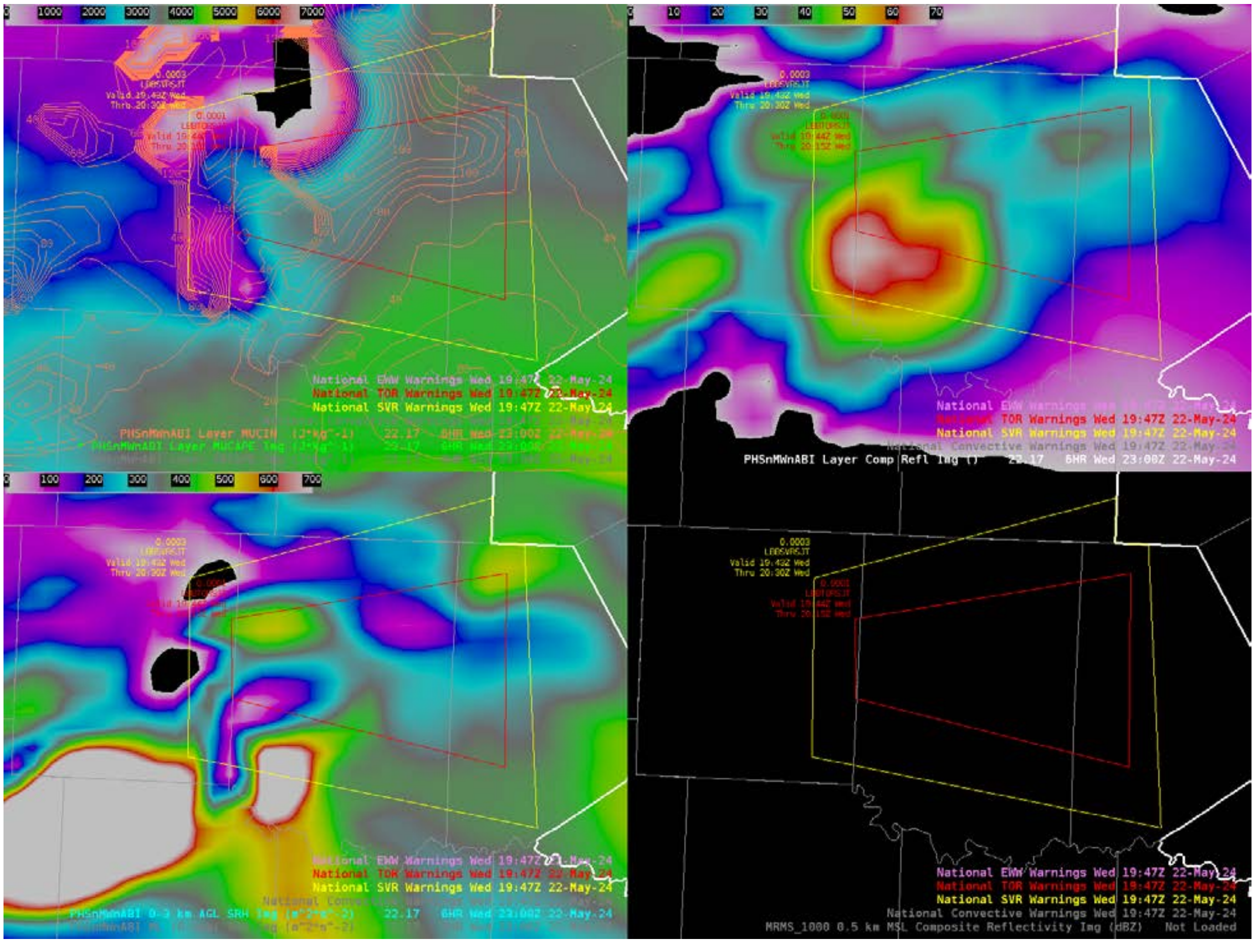

Figure 19: Output from the 17 Z PHS model run at forecast hour 6 (23 Z) on 22 May 2024 in central Texas. MU-CIN and MU-CAPE (upper left), simulated composite reflectivity (upper right), and 0-3 km SRH (lower left) are shown with the testbed-issued severe thunderstorm and tornado warnings from 1947 Z overlaid.

When validating the stability fields in the PHS model against the HRRR and SPC mesoanalysis each day, nearly half of the forecasters (46%, 18/39) felt the values were nearly identical. Approximately 23% (9/39) felt the PHS model was slightly less accurate, while 21% (8/39) felt the PHS model was slightly more accurate. In the daily discussions, surveys, and blog posts, forecasters most often mentioned how they compared the magnitudes, trends, and gradients of instability fields from each forecast hour, along with any observed convection in these environments. Most model errors when compared against observations were driven by differences in magnitudes, speeds, and locations of values and gradients in the environmental fields. This would impact the timing, location, primary mode, and development of convection. There were also a few instances where model artifacts, represented by non-physical boundaries and features, were apparent when viewed in AWIPS. Forecasters noted that they could calibrate for these features with additional time and proper training materials, however in this setting it may decrease a forecaster's confidence in the PHS model output. Participants were asked at the end of each shift how the PHS model impacted their confidence in where severe convection was more likely. Approximately 45% (21/47) of respondents felt slightly more confident, while 30% (14/47) noted no change in confidence due to the PHS model.

'I feel it was a mixed bag this week. On one day it wasn't quick enough with convective initiation and progression, even compared to HRRR. Another day it more accurately depicted initiation timing while the HRRR was late. I was also unable to find a trend in how it handled the pre-storm environment compared to other models.'
Forecaster – End of Week Survey

'The PHS oftentimes was too aggressive in its values. While this may just be a function of high resolution data getting assimilated into the models, the output would be misleading to forecasters if not calibrated to some degree.'
Forecaster – End of Week Survey

'At 22Z (5hr forecast) the PHS SBCIN product was showing a terraced appearance in SW Kansas. Maximum values were above 350, dropping to around 100 (blue/purple) a couple counties to the north. Between the two, there is the appearance that SBCIN decreases then increases again.' [Figure 20]
20 May 2024, Blog Post: *Terraced SBCIN PHS Product in SW Kansas*
https://inside.nssl.noaa.gov/ewp/2024/05/20/terraced-sbcin-phs-product-in-sw-kansas/

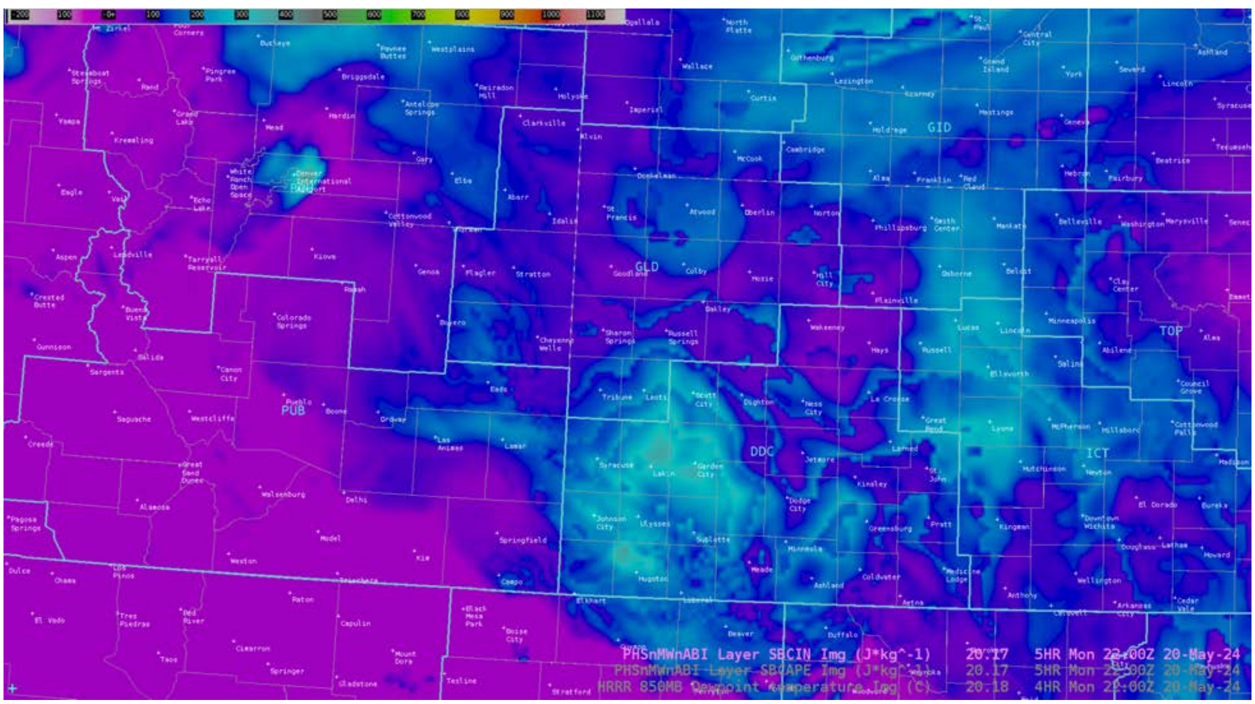

Figure 20: SB-CIN from the 17 Z PHS model run at forecast hour 5 (22 Z) on 20 May 2024 in western Kansas.

Forecasters reported in the daily surveys that they viewed the PHS model data online nearly as often as in AWIPS. On the PHS model website developers provided tools to compare PHS and RAP output of select parameters along with Skew-T plots with PHS, RAP, and radiosonde profiles at CONUS RAOB sites. Forecasters remarked in the daily discussions that these data were overall helpful when verifying the PHS data, however these applications were limited by the website interface and the graphics themselves. Most often the remarks centered on the lack of common geographic markers in the data including state boundaries, leading to confusion in the location of magnitudes, trends, and gradients in the PHS model fields with respect to other model and mesoanalysis fields. Zooming into the online plots did not improve the resolution of the data and resulted in the loss of details like the forecast hour and model run time.

> 'The model sounding graphics don't allow for enough detail to be seen. The lines and dashes could be made more fine (increase density) allowing for better takeaways. Adding a link to the user guide for each product on the products page. It would be best if there was a simpler way to compare the PHS and HRRR data.'
> *Forecaster – End of Week Survey*

> 'I would like to see more variables displayed while following similar maps to the SPC mesoanalysis. For example, pressure level maps without simulated reflectivity with a strong emphasis on temperature, humidity, and other variables, such as wind, PW, etc...'
> *Forecaster – End of Week Survey*

> 'The [online] interface is very clunky and it is hard to see the output until you zoom in...but then you lose the ability to see the model times if zoomed in. To be a resource that is used frequently, it needs to be easy to use and be able to quickly view the variables in your area of interest.'
> *Forecaster – End of Week Survey*

Developers were also interested in the PHS model performance in clear/mostly-clear and cloudy scenes. In 2023 microwave sounder information was included in the PHS model's assimilated profiles, which allowed for increased retrievals in cloudy scenes when infrared sounder retrievals were unavailable. Limited opportunities to evaluate the PHS model in these scenarios throughout a single week of the experiment made it difficult for forecasters to draw consistent conclusions in the daily surveys and blog posts. Forecasters were asked during the weekly survey if the model performed better in clear/mostly-clear or cloudy scenes, and showed some signs that including microwave information in the assimilated profiles from cloudy scenes was benefitting the PHS model. Of the 12 total responses five responded that the model was slightly better in clear/mostly-clear scenes, four responded that there was no difference between the clear/mostly-clear and cloudy environments, and three said the model performed much better in clear/mostly-clear scenes. No forecasters felt that the model performed better in cloudy scenes than clear/mostly-clear scenes.

When asked about potential avenues for product research and development, forecasters expressed in surveys and group discussions that a more direct baseline would be helpful when interpreting the impact of the assimilated satellite observations on the model's performance. Suggested baselines from group discussions often included a standard 4 km WRF run without the hyperspectral sounding observations or assimilating an operational model such as the RAP or HRRR. Display improvements to the webpages that compared the PHS model to the RAP or specific fields that could be viewed in AWIPS. Additionally, forecasters expressed interest in viewing model output on the order of 30-minutes.

> 'I think it does add value, especially in areas that don't have upper air sounding data or sparse data. However, I think we need to dig deeper into what and how the WRF uses

this, and how it impact the model input. Maybe having a tool to compare the PHS-WRF to non-PHS WRF model runs over the same domain.'
Forecaster – End of Week Survey

'It was difficult to know how much sounding information was included in the model. So some sort of comparison between non-PHS and PHS model output could be neat.'
Forecaster – End of Week Survey

'This data may help to provide additional detail on the storm scale, but overall I found it hard to evaluate whether or not this model was a sizable improvement over other CAMs we use every day.'
Forecaster – End of Week Survey

Recommendations for Operational Implementation
Based upon the evaluation of the PHS model in the 2024 HWT Satellite Proving Ground, the following items have been recommended:

- **It is strongly recommended that more tools be developed to explicitly show forecasters the impact of assimilating hyperspectral sounding information when compared to a baseline model.** Forecasters may show increased trust and better leverage information from the PHS model based on signals regarding where the assimilated information contributes to the analysis field and subsequent forecast hours.
- **It is strongly recommended that improvements be made to remove non-physical features and artifacts within the PHS model, along with continued improvements to its web display.** Both recommendations may increase the utility of the model to NWS forecasters in its current configuration. Output provided every 30 minutes may also be considered to provide more consistent signals in timing guidance.
- **It is recommended that NWS forecasters leverage the PHS model when conducting mesoanalysis in pre-convective environments.** These data should be used alongside other convection-allowing models, surface observations, radar, and satellite imagery to better understand the development, evolution, and potential hazards related to severe convection.

3.5 Probability of Severe (ProbSevere) LightningCast Model

The NOAA/CIMSS Probability of Severe (ProbSevere) LightningCast model (Cintineo et al. 2022) was evaluated again in the 2024 Satellite Proving Ground HWT, to accompany its continued transition to operations and DSS applications. LightningCast is a machine learning model trained on the reflectances and brightness temperatures of four predefined ABI bands as input (Band 2: 0.64 µm, Band 5: 1.6µm, Band 13: 10.3µm, and Band 15: 12.3µm) and 60-minute accumulations of the Geostationary Lightning Mapper (GLM) Flash Extent Density (FED) product as truth. Spatial and multi-spectral features from the four GOES-R ABI spectral bands are used in a machine-learning model to predict the probability that the GLM will observe lightning in the next 60 minutes. At night the model input is limited to Band 13 and Band 15. Output probabilities are displayed as contours by default in AWIPS-II with the option for parallax-corrected data, available for the GOES-16/-18 CONUS and mesoscale scenes, and accessible both day and night. LightningCast output for individual points are also visualized in a time-series format through a web-based lightning dashboard, available by filling out a request form for a user-specified location.

Continued demonstrations of ProbSevere LightningCast in the 2024 SPG are motivated by the need to evaluate the online lightning dashboard tool, along with demonstrating a new product featuring the probability of 10 or more GLM flashes (per five minutes) over the next hour per the GLM FED product to target more intense convection. Forecasters within the testbed were encouraged to request LightningCast meteograms for the lightning dashboard tool for each simulated DSS event to assess their effectiveness in communicating lightning-related information to their partners. Prior to the testbed, forecasters viewed a training video describing how the ProbSevere LightningCast model generates its output, the definition of the lightning related probabilities, and example applications including forecasting the initiation of lightning and decision support services. AWIPS-II procedures overlaying the LightningCast model on the Day Cloud Phase Distinction RGB, GLM FED, and the MRMS Isothermal Reflectivity at -10 °C were provided to forecasters. The default probability contours were set to 10%, 25%, 50%, and 75% for the P (≥ 1 flash in the next 60 minutes) product (hereafter 'P (≥1)'), and to 10% and 25% for the P (≥ 10 flashes per 5 minutes in the next 60 minutes) product (hereafter 'P (≥10)') with the ability to modify the contour values and colors by modifying the style rules in the AWIPS localization.

ProbSevere LightningCast R2O Questions

- How much do the lightning dashboards improve forecasters' ability to provide IDSS to partners? What components of dashboards should be implemented into operations?
- How helpful is having a product aimed towards predicting a stronger flash rate (e.g., 10 flashes)?

Use of ProbSevere LightningCast in the HWT

Throughout the Satellite Proving Ground experiment, NWS forecasters used information from ProbSevere LightningCast to anticipate the onset of lightning production and its movement as convection initiated and matured during each shift. When forecasters were asked in the daily surveys how they used LightningCast over two-thirds (43/63) responded with 'CI or convective maintenance situational awareness'. This was followed closely by 'Impact-Based Decision Support Service' (37/63) and 'Severe convective weather situational awareness' (35/63). The prepared AWIPS-II procedures and their corresponding ABI, GLM, and MRMS products were frequently used to monitor convection. Nearly all of these applications mirrored those identified in the 2022 and 2023 demonstrations of ProbSevere LightningCast. One additional update to ProbSevere LightningCast in 2024 was that the GOES-West LightningCast model had been trained on GOES-West data, where previously it has used the same model trained on GOES-East data. Several forecasters previously used LightningCast at their home office and were familiar with its performance. While no forecasters commented on a change in LightningCast performance from the GOES-West perspective, no opportunities to view data from WFOs in the western United States may have limited this aspect of the evaluation.

Forecasters most often responded that LightningCast P(≥1) product provided 40 or more minutes of actionable lead time when monitoring for the onset of lightning (initiation) and the advection of lightning activity (Figure 21), garnering nearly one-half (17/35) and over one-third (13/39) of all responses respectively. When it came to the initiation of lightning activity, over one-half (21/39) of all forecasters felt LightningCast provided actionable lead times between 10 and 30 minutes. The increase in responses for actionable lead time between the onset of lightning activity and its advection suggests that forecasters are more comfortable providing increased lead times with LightningCast for ongoing convection.

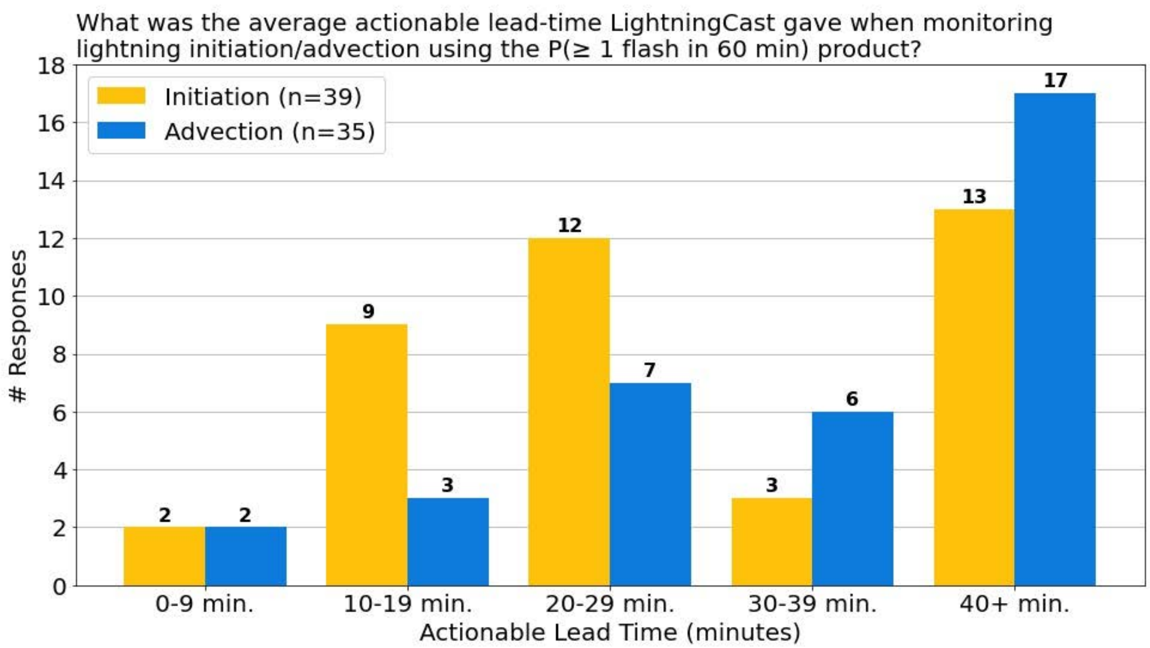

Figure 21: A bar chart showing forecaster reported actionable lead time of the ProbSevere LightningCast P (≥ 1) product for the initiation and advection of lightning activity.

> ‘Wanted to provide an assessment of LightningCast on the cells developing on the KS/MO border…Lower probabilities for Lightning Cast (10% to 25%) began appearing for these particular storms as early as 1856Z. Probabilities for 10-flash began appearing after 1922Z. The first cloud flash detected by ENTLN occurred at 1924Z. The first flash detection by GLM was at 1928Z. The first CG strike (NLDN) occurred at 1946Z. All in all, Lightning Cast provided a considerable amount of lead time, which I found to be useful.’

21 May 2024, Blog Post: *LightningCast at Initiation Near St. Joseph MO*
https://inside.nssl.noaa.gov/ewp/2024/05/21/lightning-cast-at-initiation-near-st-joseph-mo/

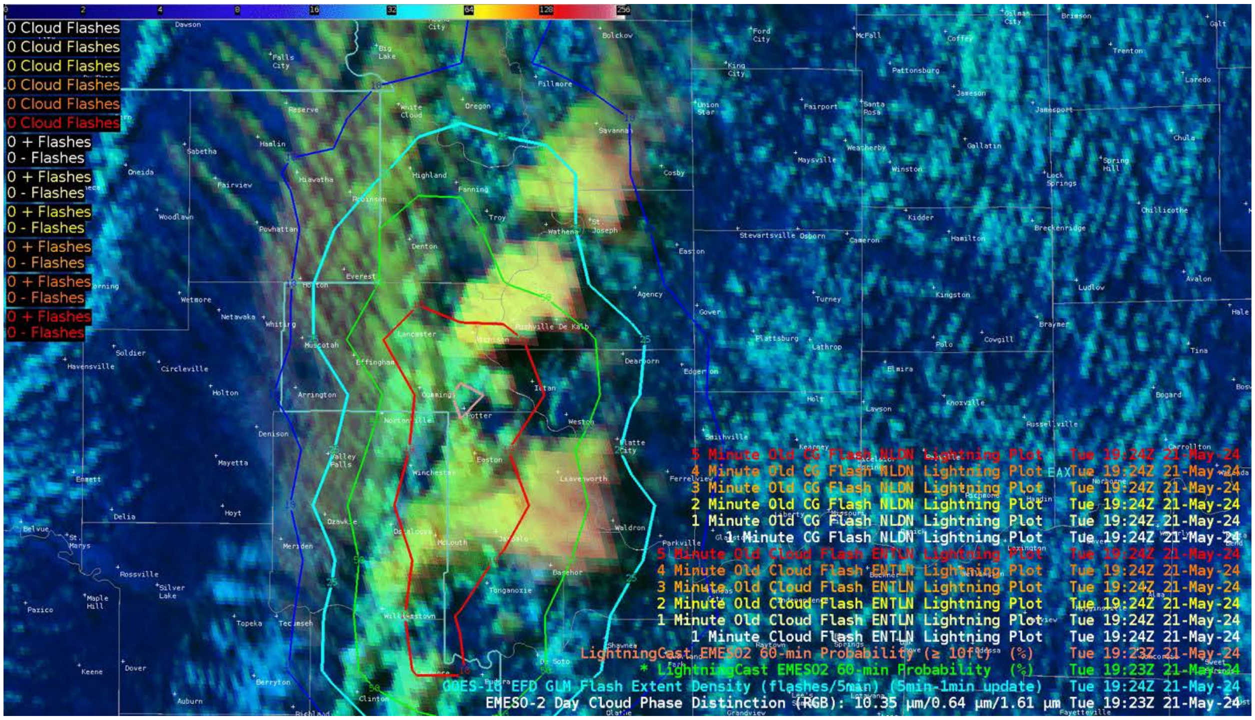

Figure 22: ProbSevere LightningCast P (≥ 1) and P (≥10) contours overlaid on the ABI Day Cloud Phase Distinction RGB in western Missouri at 1924 Z on 21 May 2024. Animation.

Forecasters also leveraged the new LightningCast P (≥ 10) product to monitor for more intense lightning activity, and therefore more intense convection, during the 2024 experiment. Overall the P (≥10) product received less feedback when compared to the P (≥ 1) product and the LightningCast Dashboard. The forecasters may already be familiar with the P (≥ 1) product, even though the product was not operational at the time of the experiment. Additionally, a few participants noted that they were confused about how to apply the P (≥ 10) product in operations, leading them to view it less frequently. Those who did view the P (≥ 10) product monitored for rapid convective development where thunderstorms had not initiated yet, along with identifying regions where ongoing thunderstorms may become more intense. At the end of each week all 16 forecasters were asked in the survey how often they would use the P (≥ 10) product if it was available operationally. This question received mixed responses, with 3 forecasters selecting 'Always or almost always', 5 selecting 'Regularly', 3 selecting 'Occasionally', 4 selecting 'Seldom', and 1 selecting 'Unsure'.

During the simulated DSS events, forecasters noted that LightningCast influenced their messaging in 77% (41/53) of all messaging sent to their partners. Lightning timing guidance provided by forecasters within their simulated DSS events corroborates with the data in Figure 21. Most often the participants communicated lightning activity impacting their site within the next 30 minutes or the next 1-2 hours. Probability thresholds were also noted by some forecasters for approaching thunderstorms, with 50% probabilities within 10 to 25 miles site being most often used. In group discussion forecasters mentioned that LightningCast was most useful in scenarios where lightning was a higher priority, such as during DSS events or when monitoring developing convection. However, as the hazard priority changed from situational awareness in pre-convective environments to warning operations with ongoing convection, forecasters noted that the utility of LightningCast decreased. This is reflected in the 52 recorded severe thunderstorm and tornado warnings throughout the testbed, with only 7 listing LightningCast as an influence in their warning decision.

> ‘Update: Several storms are moving eastwards along the Hwy 20 corridor. While hail and strong winds have been reported with these storms and will be a concern as they move towards Crookston, the lightning threat is the current concern. The probability of lightning has climbed to 50-75% within the 20 mile range of the chemical spill location and storm coverage will increase over the next 30 minutes.’
> *Forecaster – DSS Messaging Form*

> ‘LC probabilities on the SW portion of the storm [Figure 23] at 1958Z ranged between 70-75%. Just before GLM signatures pop up at 2007Z, LC probabilities jump up to around 82%. Not included in the animation, but at 1951Z, LC probabilities were around 50%. The overall trend upward would give me confidence that I can use this product to tell an emergency manager the potential for lightning is medium to high within the next 10-20 minutes (using this case, hypothetically starting at 1951Z).’
> 20 May 2024, Blog Post: *LightningCast and DSS*
> https://inside.nssl.noaa.gov/ewp/2024/05/20/lightningcast-and-dss/

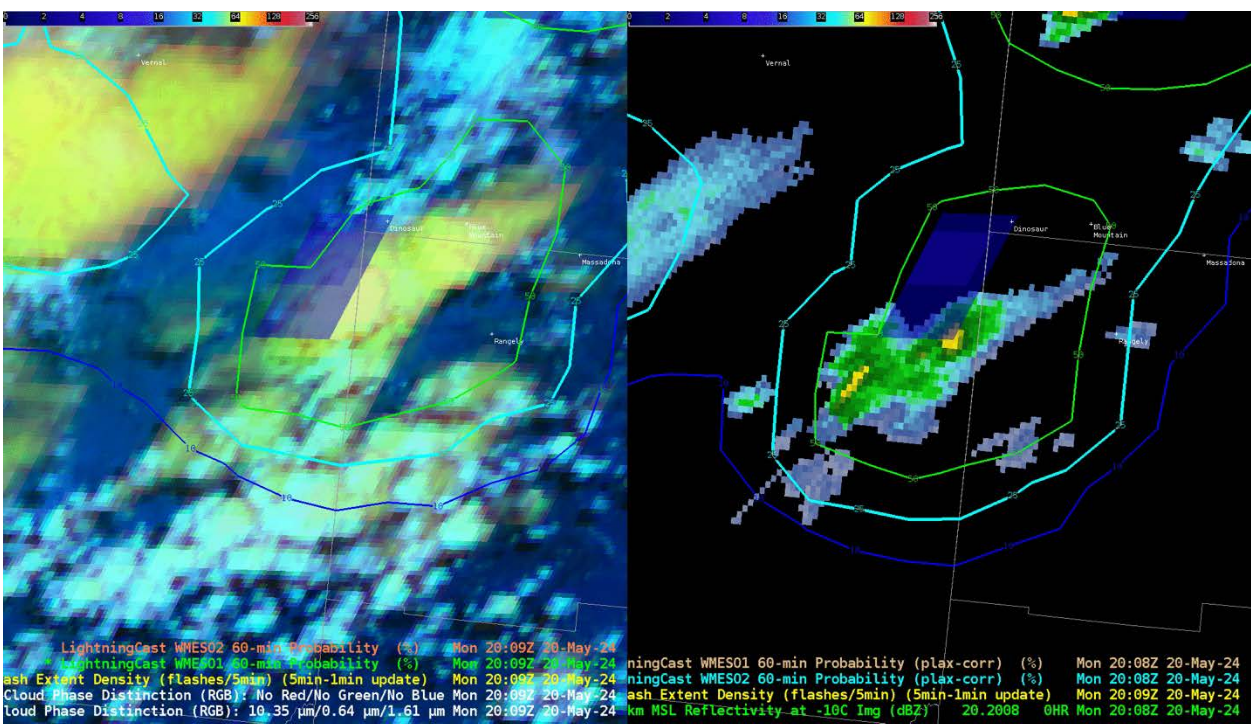

Figure 23: ProbSevere LightningCast P (≥1) product and GLM Flash Extent Density overlaid on ABI Day Could Phase Distinction RGB (left) and MRMS Reflectivity at -10℃ (right) at 2009 Z on 20 May 2024. Animation.

Forecasters frequently used the LightningCast Dashboard throughout the experiment, leveraging the static and on-demand meteograms for DSS events. During the first week of the experiment forecasters noted that probabilities directly over the site were useful, however adding the maximum probability within a specified radius around the site would also be useful. LightningCast developers implemented the maximum probability within 10 miles of the site on the dashboard, which was well received. Even after this change, forecasters consistently recommended improvements to the maximum probability range. Ideas that followed were the ability to input custom ranges other than 10 miles, or the ability to draw polygons around a point to cover larger areas such as wildfire footprints or lakes. Another question raised by the LightningCast developers in group discussions was if these dashboards should be provided and accessed by NWS partners such as emergency managers. In general, forecasters felt that the meteograms without the spatial information provided by the contoured probabilities in AWIPS-II and meteorological information that influences the probabilities may lead to confusion or an incorrect interpretation amongst their partners. Instead, forecasters suggested that this information be presented through direct communications with partners, or that partners access the dashboard be limited to situations when a forecaster is providing on-site support to aid in its interpretation.

> 'From this data, I would be able to let an EM know that we're noticing an uptick in lightning probabilities due to storm cells developing to the south/southeast of the Derby and moving northeast. These are not severe at this time and no lightning has been observed as of yet. However, currently, probabilities of lightning within 10 miles are between 40-50% [Figure 24]... I crafted a DSS message between 310-315pm…and the first flashes were observed at 320p and CGs at 325p. Therefore, this gave a 10-15 minute lead time.'
> 23 May 2024, Blog Post: *LightningCast and DSS*
> https://inside.nssl.noaa.gov/ewp/2024/05/23/lightningcast-and-dss-2/

Figure 24: ProbSevere LightningCast Dashboard for a simulated DSS event on 23 May 2024.

‘If our partners are simply looking to delay, LightningCast may also prove helpful in giving the all clear if you don’t see anything upstream on radar. The values react well as convection clears the site. This is also great in helping you know for sure when the last lightning flash took place, and it can help us give better information.’ [Figure 25]
Using GREMLIN and LightningCast for Warning Ops and DSS
https://inside.nssl.noaa.gov/ewp/2024/05/21/using-gremlin-and-lightningcast-for-warning-ops-and-dss/

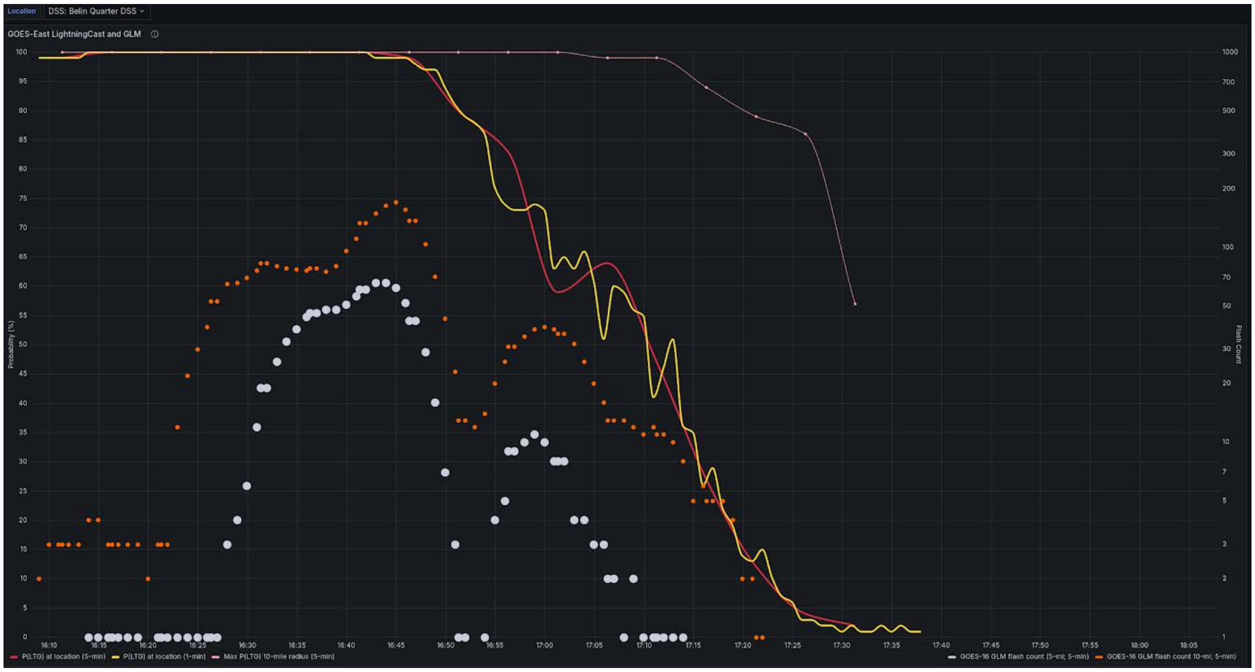

Figure 25: ProbSevere LightningCast Dashboard for a simulated DSS event on 21 May 2024.

In the weekly surveys forecasters were asked how often they would use static and on-demand demand LightningCast dashboards if they were made operational [Figure 26]. While the same number of forecasters responded with the highest and lowest bins of ‘Always/Almost always’ and ‘Seldom’ respectively, there are differences in the middle two bins. Three more forecasters selected the higher frequency option of ‘Regularly’ for the on-demand dashboard when compared to the static dashboards. For nearly all simulated DSS events forecasters used the on-demand dashboards since the static locations were not often at the default sites of airports and stadiums. While this may impact the perceived utility each dashboard type, forecasters frequently mentioned the utility of the dashboards for additional tasks such as airport weather warnings and recognized how the static dashboards could be useful. On the other hand, forecasters also mentioned how the on-demand dashboards allowed them to tailor information for their partners, with cases like wildfires also mentioned.

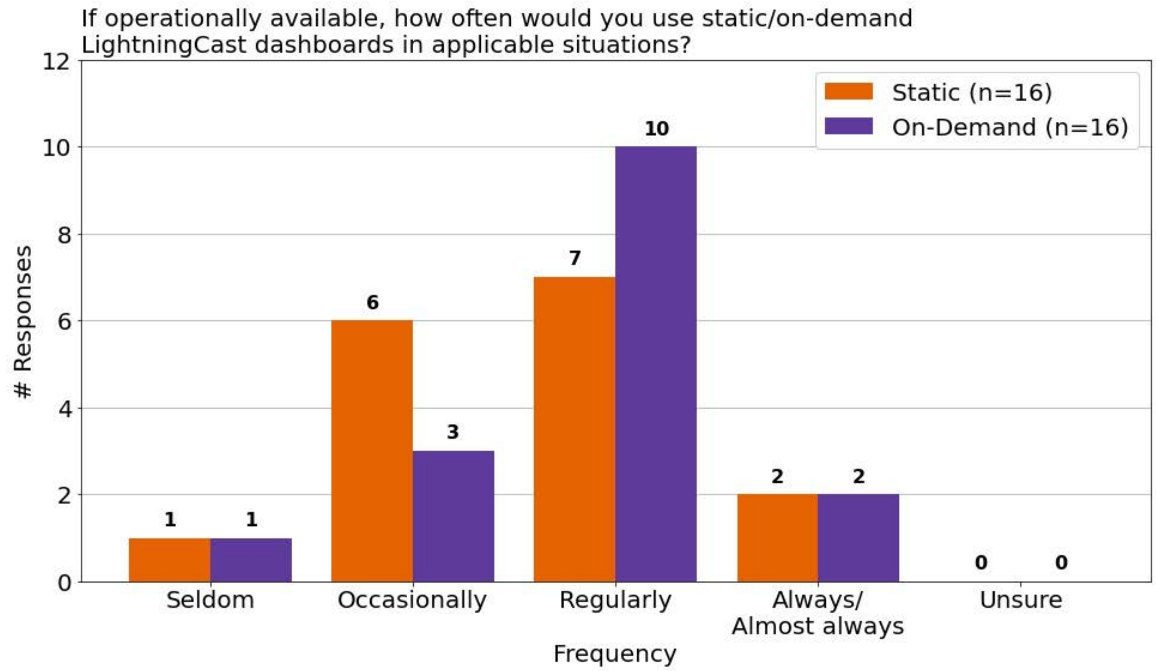

Figure 26: A bar chart showing forecaster's responses when asked how often they would use the static and on-demand LightningCast dashboards. Bins from 'Seldom' to 'Always/Almost always' are in order of increasing frequency.

Participants were also encouraged to experiment with the display of LightningCast probabilities in AWIPS-II for the P (≥ 1) product. Most often forecasters adjusted which probabilities were contoured along with their colors, or the probabilities were loaded as an image and customized. Additional lightning probability contours were often for 2%, 5%, and 90% depending on the forecaster's priorities. The 2% and 5% contours were investigated to improve LightningCast sensitivity and lead time for initiating convection for, while the 90% contour provided a high confidence to communicate to partners during DSS events. The colors of the contours themselves were also changed buy a few forecasting groups during the experiment. One modification included highlighting probabilities that were deemed important for their operational tasks, while another matched the color-coded hazards from NCEP (e.g. SPC convective outlook colors) for consistency when communicating LightningCast information to the public. Additionally, forecasters loaded the P (≥1) product as an image in AWIPS-II. Images of the product were frequently made semi-transparent, interpolated, and overlaid with other products such as radar, satellite imagery, and the LightningCast P (≥10) product contours. When using both LightningCast products in this format, forecasters commented that they were able to monitor for general lightning activity and more intense convection.

> 'the idea came up to re-do the contours such that they better aligned with the style guide the NWS, or SPC more specifically uses [Figure 27]. Image two shows how we added two contours and realigned the colors to add consistency with other operational areas of the NWS…We compared this with the base style from the lightning cast tool and we felt that our updated style better captured our eyes and made it simpler to interpret by us.'
> 14 May 2024, Blog Post: *Consistency with LightningCast*
> https://inside.nssl.noaa.gov/ewp/2024/05/14/consitency-with-lightning-cast/

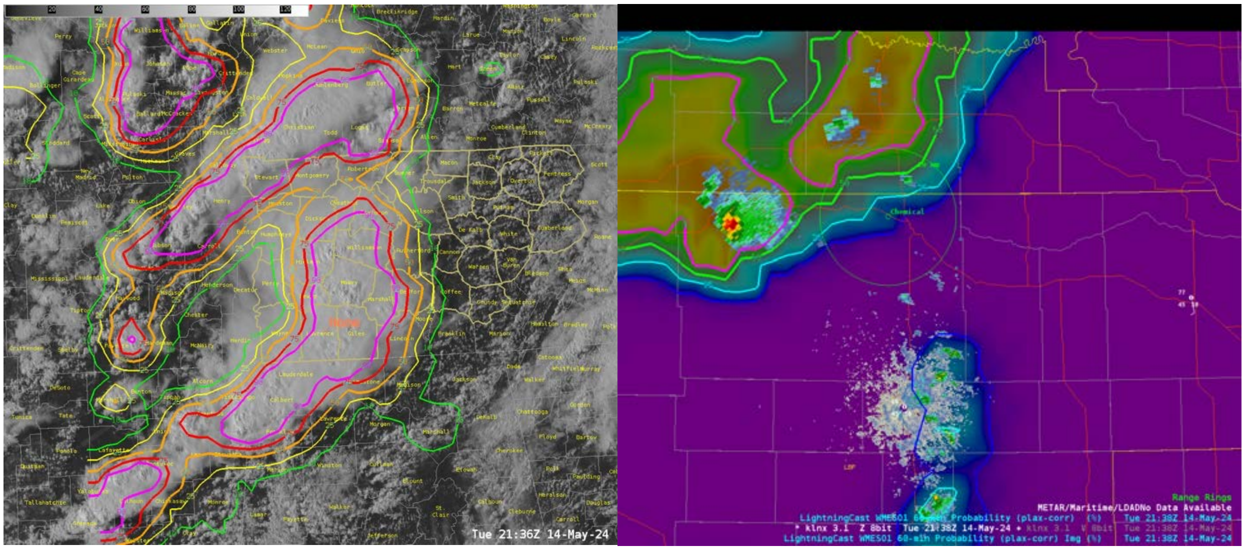

Figure 27: Left: ProbSevere LightningCast P (≥1) contours recolored to match NWS hazard colors overlaid on ABI visible imagery at 2136 Z on 14 May 2024. Right: ProbSevere LightningCast P (≥1) contours overlaid on the same product loaded as an interpolated image and base reflectivity from the KLNX radar at 2138 Z on 14 May 2024.

Recommendations for Operational Implementation

Based upon the evaluation of ProbSevere LightningCast in the 2024 HWT Satellite Proving Ground, the following items have been recommended:

- **It is highly recommended that NWS forecasters leverage static and on-demand ProbSevere LightningCast Dashboards when monitoring and communicating lightning and related hazards to NWS partners.** Dashboards may provide complimentary information when paired with probability spatial contours to diagnose developing convection and approaching/departing lightning activity.
- **It is recommended that ProbSevere LightningCast Dashboards include a maximum probability within a specified range from the point of interest.** For on-demand points, the ability to set this range may allow forecasters to tailor their messaging to a partner's needs.
- **It is recommended that ProbSevere LightningCast training include best practices for user applications and display configurations.** This may include customization of the probability contours, the advantages/limitations of each product, and complimentary products (satellite imagery, radar, lightning data, etc.) to increase situational awareness.

4. Summary and Conclusions

The GOES-R Satellite Proving Ground conducted two in-person weeks and one virtual week of product evaluations during the 2024 Spring Experiment of the Hazardous Weather Testbed. Sixteen NWS forecasters evaluated five experimental GOES-R products and interacted with their algorithm developers along subject matter experts during the experiment. Quantitative feedback was collected through surveys administered at the end of each day and week, along with a warning/decision support service reporting form. Qualitative feedback came from daily discussions with forecasters, blog posts, and public graphics. Along with the standard warning operations, mock decision support service events were created most days to reflect the full range of intended operational applications. 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. This year's experiment featured one more in-person demonstration week and one less virtual demonstration week based on feedback from the 2023 experiment, along with an archived case and radar data-denial events.

2024 Satellite Proving Ground HWT Science Questions

- How does conducting two in-person weeks and one virtual week of demonstrations compare with the 2023 experiment format of one in-person week and two virtual weeks?
- How can using an archived case impact the evaluation of each product?
- How does the utility and evaluation of satellite-related products change when radar data quality is reduced through random occurrence or data-denial exercises?

Forecasters were asked how they would briefly describe their experience in the testbed to someone at their local office. An overwhelming majority of participants stated that it was a positive learning experience, and that they enjoyed getting to contribute in both the in-person and virtual formats. Within the experiment, forecasters each week heavily relied on and liked using the centralized landing page, Slack, pre-made AWIPS-II procedures, and cloud AWIPS instances. When talking with forecasters during the in-person weeks, many expressed the advantages of interacting with product developers who were also in-person to learn and provide feedback on the experimental products. Additionally, several of the forecasters who participated virtually expressed that they were unable to travel from their home area for an entire week, and that the virtual option accommodated this need while allowing them to participate. This has been expressed during each experiment since 2021. The following quotes from forecasters highlight these points.

> 'Overall wonderful experience. Even in virtual format, I feel like there was great discussion and collaboration in working real-time and simulated events.'
> *Forecaster – End of Week Survey*

> 'This testbed is an amazing opportunity to test new products and datasets that may become operational, while being able to give real-time feedback to the developers. Be ready for a lot of information and, at times, a fast-paced environment.'
> *Forecaster – End of Week Survey*

Product developers were also surveyed at the end of the experiment, and described how demonstrating their product within the Satellite Proving Ground benefitted its development and growth toward operations. Developers who participated virtually during the in-person weeks noted a stark difference in the amount of information they learned from only the daily discussions, versus when they could be in the room with forecasters during their shifts. Those who attended on-site during the in-person weeks found the live feedback and discussion during each shift highly informative. When presented with several combinations of in-person and virtual weeks, the eight developers who responded favored options with more in-person weeks. These included the current year's format (2 in-person weeks, 1 virtual week) along with all weeks being in-person. On the other hand, the options with more or all virtual weeks received the most opposition. Hybrid weeks were suggested by at least two developers, with a small virtual group running concurrently with the in-person component if personnel and experiment resources allow.

> 'I was able to get high quality feedback on my R2O questions, particularly on days when radar denial experiments were conducted. I know that was challenging for the forecasters, but it forced them to look at the satellite data instead of focusing entirely on the radar during severe weather.'
> *Developer – End of Testbed Survey*

> The forecasters were very helpful in deciding several key factors relating to display choices for [our] products, as well as future science questions... The forecasters always provide unique opinions that we simply cannot get anywhere else!
> *Developer – End of Testbed Survey*

> 'I thought that in-person was the best given the both operational interaction + after hours conversations which often result in inspiration for new projects/approaches for current products. Virtual also works surprisingly well all things considered, perhaps the biggest benefit is the ability to include more forecasters in the experiment.
> *Developer – End of Testbed Survey*

Both forecasters and developers also had constructive criticism about the experiment design, along with ideas for improving the operational demonstration of the products. New experiment elements this year included the archived case from 11 May 2023, and selective radar data denial exercises. The archived case was only run once during the third week, and while forecasters found this case interesting, its focus on tornadoes drew forecasters toward radar radial velocity products and away from the satellite information being demonstrated. On the other hand, radar data denial exercises forced the participants to rely heavily on satellite data when interrogating convection and they therefore spent more time on the demonstrated products. Both the forecasters and developers noticed this change, and nearly all found it useful to provide a detailed evaluation. When introducing the forecasters to five new experimental products, multiple forecasters mentioned that they felt overwhelmed trying to evaluate all products during live operations each day while also exploring their applications. While forecasters were encouraged to focus on certain products based on their convective situation, it was recommended that participants focus on individual products to start each week to become more familiar with their applications.

'the warning issuers this year were glued to their radar screens, when in reality we want everyone kind of looking at and using the satellite data, which can be a challenge on big severe days. The radar denial experiments should absolutely stay though, those were some of the most useful experiments we had of the bunch'
Developer – End of Testbed Survey

The experiment setup was good. I would recommend a period of time for forecasters to really get their AWIPS set up with their preferences before jumping into interrogation. We ran this experiment with no radar data a couple of times and that was really beneficial to get me to focus on the experimental products and not to revert back to products that I'm comfortable with.
Forecaster – End of Week Survey

'Because there was a lot of new products that I have not seen or used before, it became overwhelming at times trying to look at all of the new products, in addition to radar and surface observations and other normal products I'm used to.'
Forecaster – End of Week Survey

Recommendations for Future GOES-R HWT Satellite Proving Ground Experiments

Based upon experiences and feedback from the 2024 GOES-R HWT Satellite Proving Ground, the following recommendations for future testbeds are included below:

- **It is recommended that a mix of in-person and virtual demonstration weeks continue, with hybrid weeks considered if experiment resources are available.** Forecasters and developers felt that in-person weeks, even with less forecasters, provided more feedback, but also found the virtual weeks were still valuable.
- **It is recommended that additional guidance be provided to participants regarding the products being demonstrated including their intended applications, motivation within the experiment, and evaluation parameters.** Options may include greater emphasis on R2O questions, the ability to view the survey questions before operations begin on the first day, and/or forecasters designated to evaluate specific products each shift for a more thorough evaluation.
- **It is recommended that data denial exercises be considered for future experiments, and that any archived cases focus on a spectrum of convective hazards to increase their utility.** Options may include event times or locations that are less common (e.g. inter-mountain west, northeast, overnight severe convection, etc.) that allow forecasters to fully exploit the intended applications of the demonstrated products.

4.1 Acknowledgements

Multiple NWS meteorologists participated in this experiment and provided detailed feedback that went into the recommendations from this report. It would not be possible to complete research-to-operations without their willingness to work with experimental data, ask questions, and participate in crucial discussions with product developers. The developer teams themselves have continued to leverage the experiment as a collaborative environment to test emerging technologies and transfer impactful knowledge to NWS operations. Johnathan Madden and Justin Monroe (OU CIWRO) provided exceptional technical support during all phases of the experiment, which allowed forecasters to effectively use the cloud-based AWIPS-II instances for the evaluation. Alyssa Bates, Jessica Blair, and Lexy Elizalde (CIWRO/WDTD) organized the WDTD R2O-O2R Tales Webinar. Anthony Lyza (NSSL) provided logistical and technical oversight for the Hazardous Weather Testbed. The following people participated as product focal points, subject matter experts, and observers for the products demonstrated in the experiment. The Satellite Proving Ground at the HWT thanks them for their contributions.

GLM Background and Data Quality Product

Eric Bruning	Texas Tech University
Phillip Bitzer	University of Alabama – Huntsville

OCTANE

Jason Apke	Colorado State University – CIRA
Bill Line	NOAA – NESDIS
Thomas Juliano	Colorado State University – CIRA

PHS

Scott Lindstrom	University of Wisconsin – CIMSS
William Smith	University of Wisconsin – CIMSS
Anthony diNorscia	Analytical Mechanics Associates
Andy Heidinger	NOAA – NESDIS
Lee Cronce	University of Wisconsin – CIMSS

ProbSevere (v3 and LightningCast)

John Cintineo	NOAA – NSSL
Justin Sieglaff	University of Wisconsin – CIMSS
Lena Heuscher-Stewart	University of Wisconsin – CIMSS
Michael Pavolonis	NOAA – NESDIS

Observers and Subject Matter Experts

Dan Bikos	Colorado State University – CIRA
Rebekah Esmaili	NOAA – NESDIS
Dan Lindsey	NOAA – NESDIS
Joseph Patton	University of Maryland – CISESS
Chris Siewert	NWS – SWPC
Christopher Smith	University of Maryland – CISESS
Sara Stough	University of Oklahoma – CIWRO
Javier Villegas-Bravo	University of Maryland – CISESS

5. References

Apke, J. M., Y.-J. Noh, and K. Bedka, 2022: Comparison of Optical Flow Derivation Techniques for Retrieving Tropospheric Winds from Satellite Image Sequences. *Journal of Atmospheric and Oceanic Technology*, **39**, 2005–2021, https://doi.org/10.1175/JTECH-D-22-0057.1.

Bruning, E. C., and P. M. Bitzer, 2023: Characterizing GLM Data Quality.

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

Calhoun, K. M., K. L. Berry, D. M. Kingfield, T. Meyer, M. J. Krocak, T. M. Smith, G. Stumpf, and A. Gerard, 2021: The Experimental Warning Program of NOAA's Hazardous Weather Testbed. *Bulletin of the American Meteorological Society*, 1–51, https://doi.org/10.1175/BAMS-D-21-0017.1.

Cintineo, J. L., M. J. Pavolonis, and J. M. Sieglaff, 2022: ProbSevere LightningCast: A Deep-Learning Model for Satellite-Based Lightning Nowcasting. *Weather and Forecasting*, **37**, 1239–1257, https://doi.org/10.1175/WAF-D-22-0019.1.

Cummins, K. L., 2020: Deriving GLM Group and Flash Detection Efficiencies from Measurable GLM Thresholds.

Hilburn, K. A., I. Ebert-Uphoff, and S. D. Miller, 2021: Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations. *Journal of Applied Meteorology and Climatology*, **60**, 3–21, https://doi.org/10.1175/JAMC-D-20-0084.1.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Thiel, K., 2023: *GOES-R and JPSS Proving Ground Demonstration at the Hazardous Weather Testbed 2023 Spring Experiment Final Evaluation*. Cooperative Institute for Severe and High-Impact Weather Research and Operations, https://doi.org/10.25923/37va-6g77 (Accessed November 15, 2023).

This report was prepared by Kevin Thiel with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.