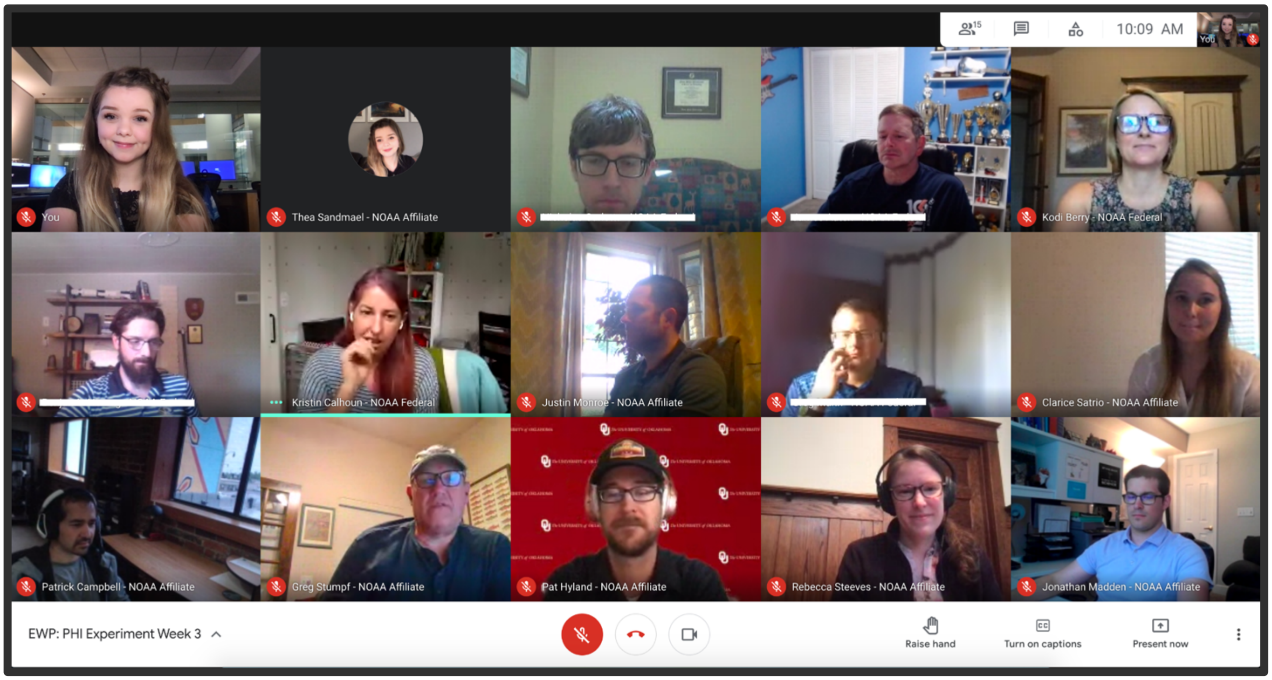
2021 EXPERIMENT SUMMARY

NOAA Hazardous Weather Testbed, Norman, OK

29 November 2022



Kodi Berry[1], Katie Wilson[2,1], Thea Sandmael[2,1], Brandon Smith[2,1], Kristin Calhoun[1], Kevin Thiel[2], G. Stumpf [3,4], C. Ling[5], K. Klockow-McClain[2,1], Jack Friedman[6], Michelle Saunders[7], and Daphne LaDue[7]

[1] NOAA/OAR/National Severe Storms Laboratory
[2] University of Oklahoma/Cooperative Institute for Severe and High-Impact Weather Research and Operations
[3] Colorado State University/Cooperative Institute for Research in the Atmosphere
[4] NOAA/NWS/Meteorological Development Laboratory
[5] University of Akron
[6] University of Oklahoma/Center for Applied Social Research
[7] University of Oklahoma/Center for Analysis and Prediction of Storms

TABLE OF CONTENTS

1. INTRODUCTION

The NOAA Hazardous Weather Testbed (HWT) is a joint project of the National Weather Service (NWS) and the National Severe Storms Laboratory (NSSL). The HWT provides a conceptual framework and a physical space to foster collaboration between research and operations to test and evaluate emerging technologies and science for NWS operations. The HWT emerged from the "Spring Program" which, for more than a decade, tested and evaluated new forecast models, techniques, and products to support NWS Storm Prediction Center (SPC) forecast operations. Now, the HWT consists of two primary programs: the original Spring Program, which is part of the Experimental Forecast Program (EFP), and the Experimental Warning Program (EWP).

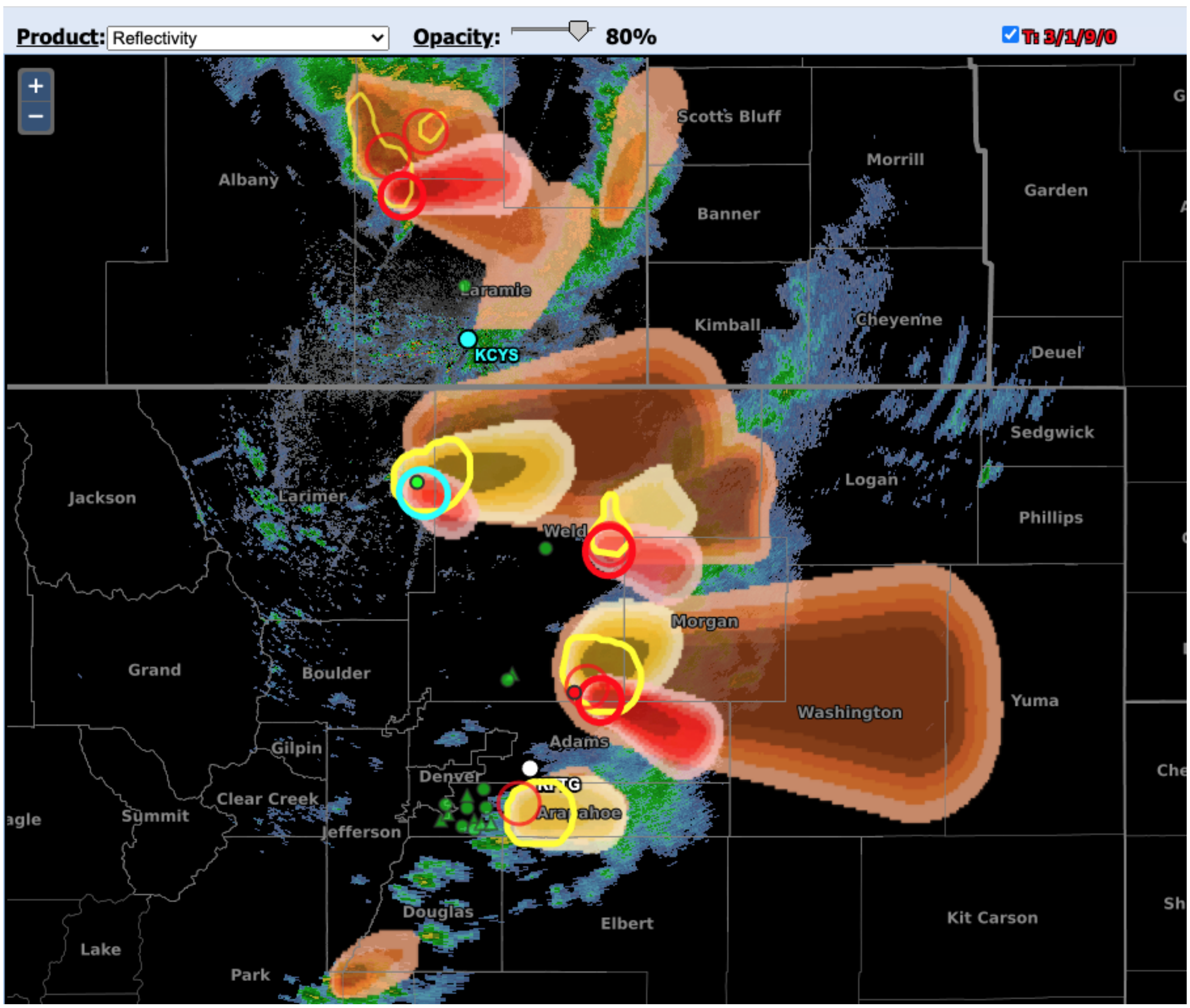


Figure 1. A forecaster's view of the PHI Prototype Tool as it is used to manage storm objects and probability plumes for an archived case.

The EWP tests and evaluates new applications, techniques, and products to support Weather Forecast Office (WFO) severe convective weather warning operations. This was the fourteenth year for warning activities in the HWT. Feedback was gathered from NWS operational meteorologists. The experiment participants issued experimental warnings, published live blogs, engaged in focus groups, and completed surveys. User comments were also collected during shifts, which have been used to inform product development. This input is vital to improving the NWS warning process, which ultimately leads to saved lives and property.

2. OVERVIEW

Due to the COVID-19 pandemic, all in-person HWT activities were conducted virtually. The HWT EWP hosted seven experiments over 19 calendar weeks to improve NWS severe weather warnings.

Cloud services were procured using NOAA's cloud contract vehicle. This contract allows for the investigation and implementation of remote HWT experiments using AWIPS in a cloud environment. With NOAA's inclusiveness of remote working employees and partners we feel that investigating methods to improve the opportunity and collaboration with as many NOAA employees and partners for the HWT is paramount.

EWP Experiment	Dates	Length	Number of Participants
Warn-on-Forecast	Feb 9 - Feb 12 Mar 2 - Mar 5 Mar 9 - Mar 12 Mar 30 - Apr 2	4 weeks	16 forecasters
Radar Convective Applications	Apr 19 - Apr 23 May 3 - May 7 May 17 - May 21	3 weeks	16 forecasters 2 military
PHI Prototype	Apr 26 - Apr 30 May 10 - May 14 May 24 - May 28	3 weeks	9 forecasters
Satellite Convective Applications	Jun 1 - Jun 4 Jun 7 - Jun 11 Jun 14 - Jun 18	3 weeks	17 forecasters
Hazard Services - Threats-in-Motion	Jul 19 - Jul 23 Aug 2 - Aug 6 Aug 30 - Sep 3	3 weeks	6 forecasters
Threats-in-Motion End Users	Sep 20 - Sep 24	1 week	24 broadcast meteorologists 21 emergency managers
Brief Vulnerability Overview Tool	Jul 12 - Jul 16 Jul 26 - Jul 30 Dec 6 - Dec 10	3 weeks	18 forecasters 17 emergency managers

Table 1. Details for the 2021 Experimental Warning Program.

3. PROJECT DETAILS AND RESULTS

Warn-on-Forecast Experiment

Summary by Katie Wilson

Overview

Since 2017, research efforts with local Weather Forecast Offices (WFOs), the Weather Prediction Center (WPC), and the Storm Prediction Center (SPC) have provided opportunities to learn about Warn-on-Forecast System (WoFS) applications in NWS operations. These research efforts were conducted separately (i.e., collaborating with WFOs, SPC, or WPC at one time), and all three NWS entities demonstrated utility for using WoFS in their forecast process. Given the storm-scale, 0–6-h probabilistic guidance provided by WoFS, its utility to NWS operations occurred across the traditional watch-to-warning (W2W) spatiotemporal scales.

Despite the separation in collaborations with WFOs and national centers, overlapping applications of WoFS guidance was evident. While national centers used WoFS to provide more refined guidance in the several hours preceding a weather event (i.e., through mesoscale discussions), WFOs demonstrated used WoFS to provide slightly broader yet still storm-focused guidance for weather events ahead of the traditional warning product. As a result, a research need to understand the forecast processes and responsibilities of these NWS entities in a joint sense, and what they mean for the future operational integration of WoFS guidance, became a top priority and motivated the 2021 Warn-on-Forecast (WoF) Experiment.

In the 2021 WoF experiment, a unified approach brought together 16 NWS forecasters from nine southern region WFOs, the Storm Prediction Center (SPC), and the Weather Prediction Center (WPC). This experiment was unique in its first attempt to explore forecast challenges that span the watch and warning spatiotemporal scales, and addressed research goals related to both NOAA HWT Experimental Forecast and Warning Programs. This experiment explored three research questions:

1) How do forecasters envision WoFS guidance fitting into their existing forecast process?
2) How can WoFS guidance be used most effectively across the current watch-to-warning forecast process?
3) How can WoFS guidance fit into a visionary forecast process?

Experiment Details and Results

The 2021 WoF experiment consisted of two short pre-experiment activities and a four-day (Tuesday-Friday) virtual engagement including overview presentations, simulated real-time events, focus groups, a debrief, and an opportunity to provide post-experiment anonymous feedback. Findings from the feedback are weaved throughout this summary. This experiment was repeated over four weeks, each time with a different group of participants. Each group included two WFO forecasters, one SPC forecaster, and one WPC forecaster.

i) Survey and Flowcharts

The pre-experiment activities consisted of a survey and a flowchart exercise. The survey collected information on participants' current forecast roles, career experience, familiarity with WoFS guidance, and existing expectations for WoFS use in operations. The flowchart exercise was designed to document participants' existing workflows that aid prognosis and diagnosis of typical warm-season severe weather events (Fig. 1). In the post-experiment feedback survey, 93% of participants agreed or strongly agreed that the flowchart exercise was helpful for thinking about and sharing their forecast and decision-making process. These flowcharts were used later in the first focus group session, and were found to bring self-awareness to an otherwise automated process, facilitate discussion with other forecasters, and highlight when and how WoFS guidance fits into existing workflows.

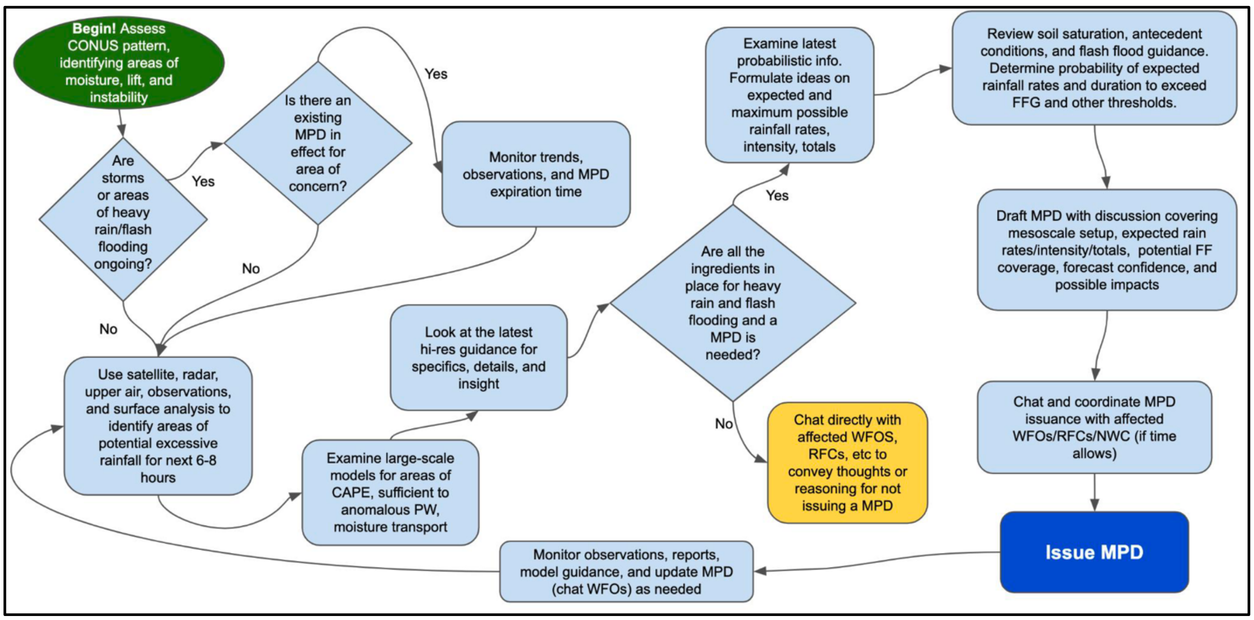


Figure 1. An example of a flowchart created by a WPC forecaster describing their workflow associated with Metwatch Desk responsibilities.

ii) Simulations

To prepare participants for WoFS guidance use during the experiment, overview presentations included information on how the system works, how to view and interpret different products, and operational examples of its utility for different weather events. Participants were then prepared to work two simulated real-time events. These events were made available using AWIPS-2 in-the-cloud (a first for the NOAA HWT), and a 90-min familiarization session was provided to introduce participants to the cloud interface and the processes required to load, view, and interact with data. This session was particularly important for the SPC and WPC forecasters since they do not use AWIPS-2 in routine operations. Furthermore, the familiarization session was important for learning how to load and view WoFS guidance in AWIPS-2, which was another first for this experiment.

Participants worked two weather events. The first was a shorter simulation (~3.5 hours long) of the 19 July 2018 tornado event in Iowa, and the second was a day-long simulation (~8 hours) of the 28 June 2018 northern plains supercells and mesoscale convective system event. A forecaster from the Norman WFO created videos to brief participants on the ongoing weather. These briefings were watched immediately prior to beginning each simulation. All participants were assigned the same goal when working the simulations: *"To immerse yourself in your forecast process, while exploring ways in which the WoFS guidance can be applied to your roles and responsibilities."*

Participants were provided a one-page document describing the scenario, their roles, and the associated responsibilities and tasks they were expected to carry out. The two WFO forecasters alternated between mesoanalyst and decision support service positions, while the WPC forecaster fulfilled the metwatch desk position in both cases, and the SPC forecaster fulfilled the lead desk position in both cases. The WFO mesoanalyst and decision support service forecasters were expected to share, create, and disseminate information within and beyond their WFO, as they would in normal operations. To facilitate the full WFO forecast process, each week an additional non-participant forecaster served in a support role to complete the task of issuing warnings. The WPC forecaster was asked to focus on mesoscale precipitation discussions, and the SPC forecaster was asked to focus on both watch and mesoscale convective discussions. All forecasters also had the opportunity to create ad-hoc graphics and issue statements via a constructed NWS chat room.

Figure 2. A screenshot highlighting the different forecast roles, presence of IT support, and facilitation by WoF scientists during a simulation.

Throughout both cases, participants remained in a google meet room and could speak and use the sidebar chat to communicate with other forecasters and with the WoF team. They could also use the google chat room to share information with one another (including snapshots of WoFS guidance). Furthermore, participants shared a google slide document, in which they all documented, and in some cases completed, any product that was issued (e.g., watches, warnings, mesoscale discussions, and decision support graphics).

All of these data were recorded, yielding 46 hours of simulation video, text logs, and a chronology of issued products (ranging between 10–80 products per case). The simulation videos have since been transcribed and are undergoing rigorous thematic analysis by four research scientists. This analysis is capturing themes including discussion of WoFS guidance, focus on storm-scale vs mesoscale analysis, warning operation activities, and the occurrence of inter vs. intra office collaboration. The ebb and flow of these themes within each simulation will provide insight on how WoFS guidance impacts the standalone vs. overlapping forecast processes of WFOs, WPC, and SPC.

In the post-experiment feedback survey, 100% of participants agreed or strongly agreed that the simulations helped them develop an understanding of how WoFS guidance can be applied in operations. They felt the simulations mimicked in-office distractions, yet the no-risk environment allowed them to explore the WoFS guidance in greater depth. Through trial and error, participants were able to identify the most useful products and how to most effectively interject the WoFS guidance into the internal forecast process as well as to external receivers (e.g., via NWS chat or a decision support graphic).

iii) Focus groups

Three semi-structured, moderator-led focus group sessions were conducted during Thursday and Friday morning of each experiment week. These focus groups lasted on average two hours each, were conducted in google meet chat rooms, and were video recorded. Each focus group session addressed a particular topic that drew on participants' experiences from using WoFS during the simulation exercises and directly addressed the three research questions outlined in the study overview above. The topics included: 1) The existing forecast process, 2) Comparison of national center and local office WoFS guidance use, and 3) The visionary forecast process.

Each focus group began with a question that was answered through an independent hands-on activity followed by group discussion. In the first focus group, the pre-experiment flowcharts were revisited, and participants were asked to modify their workflows to embed use of WoFS guidance. Their flowcharts were then described one-by-one, and this discussion provided an important foundation for following questions focused on WoFS guidance use in the existing forecast process. In the post-experiment feedback, 67% of participants agreed or strongly agreed that the flowchart modification activity was helpful for capturing changes to their forecast and decision-making processes resulting from WoFS use guidance. For these participants, they felt the activity helped to reinforce where WoFS guidance can fit into their existing workflows, the existing gaps it fills, and potential best practices for adjusting current workflows. For the remaining participants who did not find this activity helpful, they reported that they either did not envision WoFS altering their workflows given that they already use convection-allowing models, or that they had already embedded WoFS into their pre-experiment flowchart because they used WoFS routinely when it is available in real-time operations.

The second and third focus group sessions began with use of the google meet tool "jamboard." In the second focus group, participants added ideas to a venn diagram to highlight most useful applications and most challenging aspects of using WoFS guidance. The Venn diagrams helped to identify ideas that were specific to local offices, specific to national centers, and applicable to both (e.g., Fig. 3). In the third focus group, participants posted key words or phrases that captured their expectations for how their role as a forecaster would evolve over the next ten years. Some high-level findings related to the second focus group jamboard activity are described below.

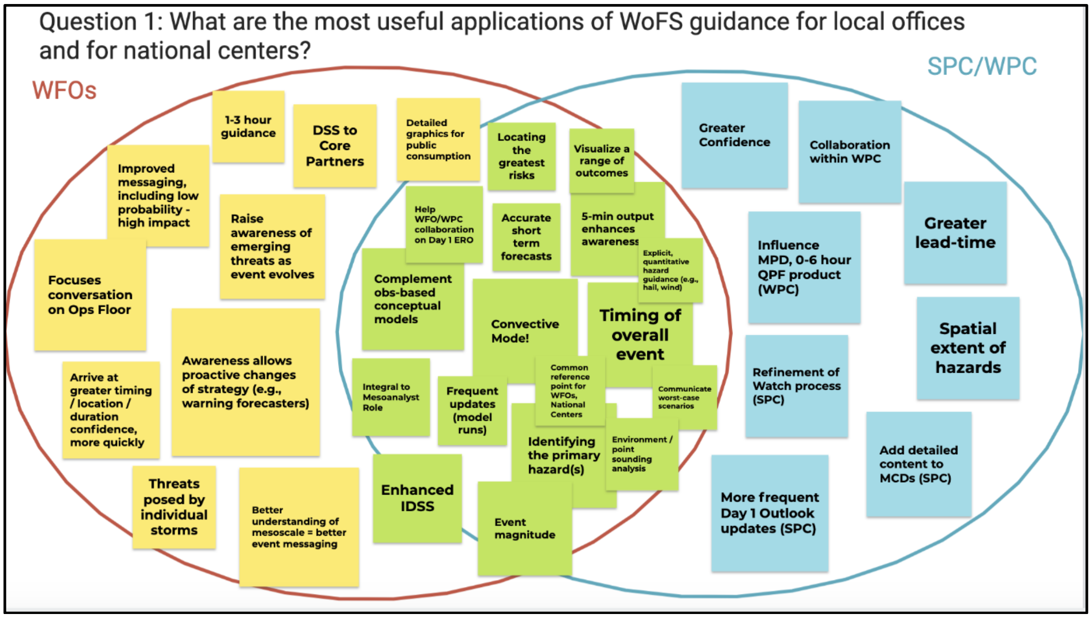


Figure 3. Summary of jamboard input across the four experiment weeks for the second focus group session question inquiring about most useful applications of WoFS guidance (collated by Patrick Burke).

The WFO participants emphasized how WoFS can support the production of decision support service graphics for external users, and that they could share information on the timing, location, duration, and potential magnitude of weather hazards earlier and with more confidence. The WPC and SPC participants emphasized how WoFS serves as an important decision aid for greater product detail, and for SPC in particular, could help to refine the watch process. Overlapping themes between WFOs and national centers included WoFS guidance serving as a common reference point for collaboration, enabling forecasters at different offices to visualize the range of potential outcomes, and be more certain of individual hazard magnitude and timing. Consistency in expectations and communications of forecast information was expected as a result of using WoFS as a common reference point.

For the venn diagram highlighting challenges associated with using WoFS in operations (not shown), WFO participants noted concerns with the varying performance of WoFS predictions for different storms in the same domain (which is a function of a storm's post-CI lifetime). These participants also noted challenges including forecasters' emphasis on using individual members and maintaining a deterministic mindset, as well as establishing a strategy for who would be best suited to analyze and disseminate WoFS guidance information within the office. For the national centers, the largest challenge noted was that the WoFS guidance only goes out to six hours. Overlapping challenges between WFOs and national centers included having sufficient time to analyze each WoFS run, handling data overload, investing too much trust into WoFS guidance because it can be "beguiling," and minimal and/or inconsistent exposure to WoFS guidance due to the current limited domain location and real-time availability. This latter challenge will hopefully be alleviated with the cloud-based WoFS capabilities that are being developed and tested as of spring 2022.

Aside from the post-experiment flowchart and two jamboard activities, participants answered a variety of predetermined questions in the focus groups and engaged in rich discussion. To conduct a qualitative analysis of these discussions, a summer student transcribed the approximate 24 hours of video footage. The research team is beginning a thematic analysis of these qualitative data. In the post-experiment feedback, 100% of participants agreed or strongly agreed that the focus groups were effective for sharing their thoughts and ideas. In particular, participants felt the questions were appropriately designed, that participants had opportunities to engage with one another, and they were able to shed light on their shared goals and visions for WoFS use in the NWS.

Future Plans

The research team's goals for 2022 include completing a thorough analysis of the data collected (forecast flowcharts, simulation discussions and chats, simulation product chronologies, and focus group discussions). A review of initial findings was shared by Patrick Burke at the 2022 AMS Annual Meeting, and further findings will be shared at conferences and submitted to peer-reviewed journals later this year.

Collaborations with WFOs, WPC, and SPC are ongoing to evaluate forecasters' real-time use of WoFS guidance. To enable forecasters' real-time use of WoFS, resources need to be developed to facilitate understanding and effective use of WoFS guidance. The research team will use findings from the 2021 WoF experiment to develop webinar and training materials, and these resources will be made available on a web page in 2022. The WoF research team will also expand real-time collaborations outside of the NWS southern region. Cloud-based WoFS capabilities will enable more year-round real-time runs in 2022, resulting in a broader and more diverse demonstration of WoFS.

Future testbed experiment plans will be finalized upon completion of the analysis of this study. Efforts are underway to develop machine learning prototype products that incorporate probabilistic hazard information from both observational and model guidance, and thus provide probabilistic information from the warning to the watch scales. These products will undergo evaluation in the testbed within the next several years. The unified approach of bringing together national center and local WFO forecasters, as demonstrated in this study, will be adopted in future testbed experiments that span the interests of multiple NWS entities. Participants in the 2021 WoF experiment enjoyed this approach, with 100% of participants agreeing or strongly agreeing that they developed new insight into others' roles in the NWS. In particular, participants were able to get an "inside look" at what others are doing during weather events, become familiar with the challenges they face, and recognize how their roles can flex to better work together in the future. The types of perspectives gained through this unified approach will be instrumental to building a future forecast process that values consistency and continuity across offices.

Web Presence

Warn-on-Forecast Program https://www.nssl.noaa.gov/projects/wof

Real-time and Archived WoFS Guidance https://wof.nssl.noaa.gov/realtime/

Project Contacts

Katie Wilson	CIMMS/NSSL	PI, Warn-on-Forecast Scientist
Patrick Burke	NSSL	Co-PI, Warn-on-Forecast Program Lead
Burkely Gallo	CIMMS/SPC	Co-PI, Warn-on-Forecast Liaison to SPC
Patrick Skinner	CIMMS/NSSL	Co-PI, Warn-on-Forecast Scientist
Jorge Guerra	CIMMS/NSSL	Warn-on-Forecast and AWIPS Developer / Simulation Developer
Todd Lindley	NOAA NWS	Operational Collaborator
Stephen Bieda	NOAA NWS	Operational Collaborator
Chad Gravelle	NOAA NWS	Operational Collaborator
Jonathan Madden	CIMMS/NSSL	Lead IT Support / Simulation Developer
Justin Monroe	CIMMS/NSSL	IT Support/ Simulation Developer
Dale Morris	CIMMS/WDTD	Simulation Developer

Radar Convective Applications Experiment

Summary by Thea Sandmael and Brandon Smith

Overview

The HWT Experimental Warning Program (EWP) Radar Convective Applications (RCA) experiment featured both the New Tornado Detection Algorithm (NTDA; now TORP) and the New Mesocyclone Detection Algorithm (NMDA), which were evaluated over a three-week period in the spring of 2021. Funded by the National Weather Service (NWS) Radar Operations Center (ROC), the main goal of the new algorithms is to replace the aging legacy Tornado Detection Algorithm (TDA) and Mesocyclone Detection Algorithm (MDA) for the WSR-88D radar network with new versions that leverage modern products and advances in radar technology.

The 2021 RAC experiment was conducted over 3 weeks (April 19-23, May 3-7, May 17-21) with six staff members from CIWRO/NSSL, including the technical support team. 16 forecasters from the NOAA NWS and two from the U.S. Department of Defense Air Force participated in the experiment, with six forecasters joining each week through Google Meet video chats (Fig. 1). The experiment consisted of a blend of real-time and archived displaced real-time weather events displayed on instances of AWIPS-II that operated on virtual machines hosted by Amazon Web Services (AWS).

FIGURE 1. Screenshot of video chat and pictures of staff setup during the HWT experiment.

Experiment Operations Details

The experiment operated Monday through Friday with a training day occurring on Monday, archived and real-time operations on Tuesday-Thursday, and a weekly group discussion and surveys on Friday. On the training day, the forecasters were provided with an archived case featuring supercells in the north Texas area to get them familiar with how to use the NTDA and

NMDA in concert with the meteorological data they are familiar with when in actual warning operations. The training day was also used to set up the real-time systems and make sure all of the forecasters were able to access everything they needed for potential real-time evaluations during Tuesday through Thursday operations. Unfortunately, this was the only opportunity some of the forecasters got to evaluate the products in real time due to inactive weather conditions during some of the experiment weeks.

On Tuesday-Thursday, the forecasters were shown various archived weather events each morning, as well as either a real-time event or another archived case in the afternoon depending on whether there was the potential for severe weather in the CONUS. Table 1 shows the list of cases that the forecasters used to evaluate the products. These cases included a variety of areas in all NWS regions and different storm modes, such as quasi-linear convective systems (QLCSs), supercellular environments, mixed storm modes, or challenging events like the July 2018 Des Moines, IA supercell tornado event. The forecasters were asked to provide feedback through blog posts (available at the HWT EWP Blog: https://inside.nssl.noaa.gov/ewp/), verbal discussions, and online surveys provided after each case. During evaluations, forecasters operated in pairs in a video chat with a researcher that would facilitate and observe the discussion, as well as take notes. These pairs would rotate throughout the week to allow the researchers and participants access to different individuals within the experiment. After each evaluation, there were also PI-led case discussions that preceded the surveys.

On Fridays, the forecasters communicated their experience as a whole through a final group discussion, again led by the principal investigators, and an end-of-week survey focusing on their overall impressions of the products.

While working weather events, the forecasters were encouraged to choose their own methods of evaluation. This meant that sometimes they would just focus on investigating one or both of the products and write blog posts of what they saw, and other times they would simulate real-time warning operations and issue warnings as if they were the warning meteorologist working the event, switching between covering the same county warning areas (CWAs) or splitting up to cover different storms. During live weather events, each forecaster pair was assigned a CWA and issued warnings for different storms within that area.

TABLE 1. List of cases included in the RAC experiment.

Date	Time (UTC)	County Warning Area(s)	Number of Forecasters Evaluated	Type of Case
21 Oct 2019	0100-0300	Fort Worth-Dallas, TX (FWD)	18	Archived training
15 May 2018	1810-2025	Albany, NY (ALY) Binghamton, NY (BGM) New York/Upton, NY (OKX)	17	Archived
28 May 2018	0100-0315	Indianapolis, IN (IND) Northern Indiana (IWX) Wilmington, OH (ILN)	18	Archived
21 Oct 2019	0230-0445	Norman, OK (OUN) Tulsa, OK (TSA)	18	Archived
31 May 2018	2045-2300	Pocatello, ID (PIH)	6	Archived backup
19 Jun 2018	2000-2200	Boulder, CO (BOU) Cheyenne, WY (CYS)	12	Archived backup
19 Jul 2018	1930-2200	Des Moines, IA (DMX)	18	Archived backup
19 Apr 2021	1930-2100	Melbourne, FL (MLB) Miami, FL (MFL)	6	Real-time training
03 May 2021	1900-2015	Columbia, SC (CAE) Greenville-Spartanburg, SC (GSP) Peachtree City/Atlanta, GA (FFC)	6	Real-time training
04 May 2021	1950-2215	Birmingham, AL (BMX) Jackson, MS (JAN) New Orleans/Baton Rouge, LA (LIX)	6	Real time
06 May 2021	2000-2230	Indianapolis, IN (IND) Lincoln, IL (ILX) Memphis, TN (MEG) Paducah, KY (PAH)	6	Real time
17 May 2021	1900-2030	Albuquerque, NM (ABQ) Amarillo, TX (AMA) Austin/San Antonio, TX (EWX) Fort Worth-Dallas, TX (FWD) Lubbock, TX (LUB) San Angelo, TX (SJT)	6	Real-time training
18 May 2021	1950-2315	Amarillo, TX (AMA) Austin/San Antonio, TX (EWX) Corpus Christi, TX (CRP) Fort Worth-Dallas, TX (FWD) New Orleans/Baton Rouge, LA (LIX) Shreveport, LA (SHV)	6	Real time

Overall, the experiment was very successful, even with the limitations that come with not being able to conduct it in the traditional in-person format. In fact, all of the participating forecasters would either definitely or probably participate again and recommend coworkers to apply to participate if the experiment was required to be held virtually in the future (Fig. 2). In a post-COVID-19 future, it might be a cost-saving option to offer one of the experiment weeks as a virtual week to make it easier for forecasters located far away or have travel limitations, and get them a chance to join the experiment from where they live.

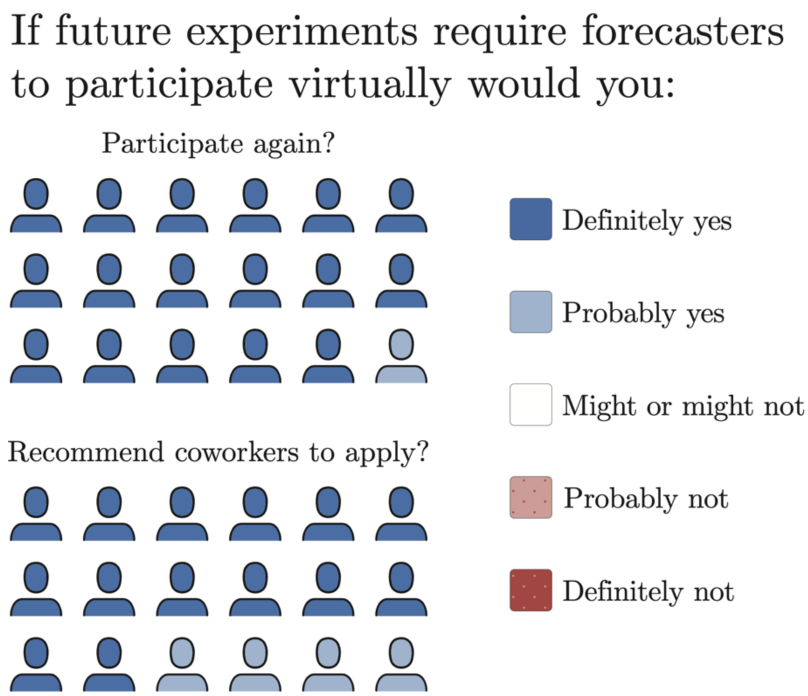


FIGURE 2. Survey responses about the forecasters' virtual experience. 17 out of 18 forecasters would definitely participate again in a virtual experiment, while one forecaster would probably participate again. 14 out of 18 forecasters would recommend their coworkers to apply to a virtual experiment, and 4 would probably recommend it.

Description of Evaluation Methods

As mentioned above, the three main evaluation methods were blog posts, surveys, and verbal discussions. The blog posts are a very insightful evaluation method, as they include the forecasters' description of situations and relevant screenshots so project PIs can go back and examine the situation they were depicting. While a good resource, the blog posts often focus on negative aspects of the algorithms, so the product reviews may appear more unfavorable than reality. However, the PIs communicated to the participants that pointing out flaws can be very helpful in order to make improvements to the products before any possible transition to operations occurs.

Surveys are used as a baseline feedback method that provides project PIs with results that can be aggregated and scored. After each evaluation case, whether real-time or archived, participants took a survey asking them questions regarding their use of the NTDA and NMDA algorithms for that specific event. At the end of the week after the experiment was finished, participants were provided a more detailed survey that asked their overall thoughts on the algorithms and how they changed throughout the week.

Verbal group discussions occurred after the completion of each evaluation period, with a final larger group discussion taking place at the end of the week on Friday. The discussions allowed participants and project PIs to openly discuss their thoughts and feelings with everyone regarding their use of the algorithms during the evaluations. Both positives and negatives would be brought up in this open format, allowing all parties to discuss possible solutions or next-steps on how the algorithms could grow or improve. Suggestions of future improvements or additions to the algorithms by the participants provided project PIs a wealth of ideas and information to further aid operational forecasters in the warning environment.

Feedback obtained from all of these evaluation methods forms the basis of the results for the 2021 HWT EWP RCA experiment, which are discussed in the following sections where blog post themes, surveys results, and discussion overviews concerning the NTDA and MDA will be outlined, including suggested action items for algorithm improvements to be conducted in FY22.

NTDA Results

Overall, forecaster reception of the NTDA was very positive. The participants found the product useful overall (Fig. 3), and the only forecaster that said that they were unsure if they would use it participated in the first week of the evaluation when the NTDA was less refined and the probability slider was not ready yet (Fig. 4). They said they would use the NTDA if there was a more organized readout to better utilize the information and additional filters for non-meteorological detections, both of which were partially adjusted for later weeks and are funded to be refined in FY22. The NTDA was judged as much better than the TDA, which is hardly ever used operationally and generally viewed as not very useful. The algorithm's main advantage was concluded to be its ability to provide good guidance for rapid decision making and boosting forecaster confidence in high-stress situations. Many forecasters express that they would use the NTDA as a storm-interrogation tool for warning decisions or for overall situational awareness.

There are a couple of things to improve upon before the NTDA would be ready for operations, such as a time matching issue, which will be worked on as part of the AWIPS-II visualization task, and noise detections. One other concern was raised by the USAF forecasters. They said that they would very much like to use the NTDA, but that they do not have AWIPS-II available to them in their office and would need someone to develop GR2/GR3 placefiles to be able to use the product.

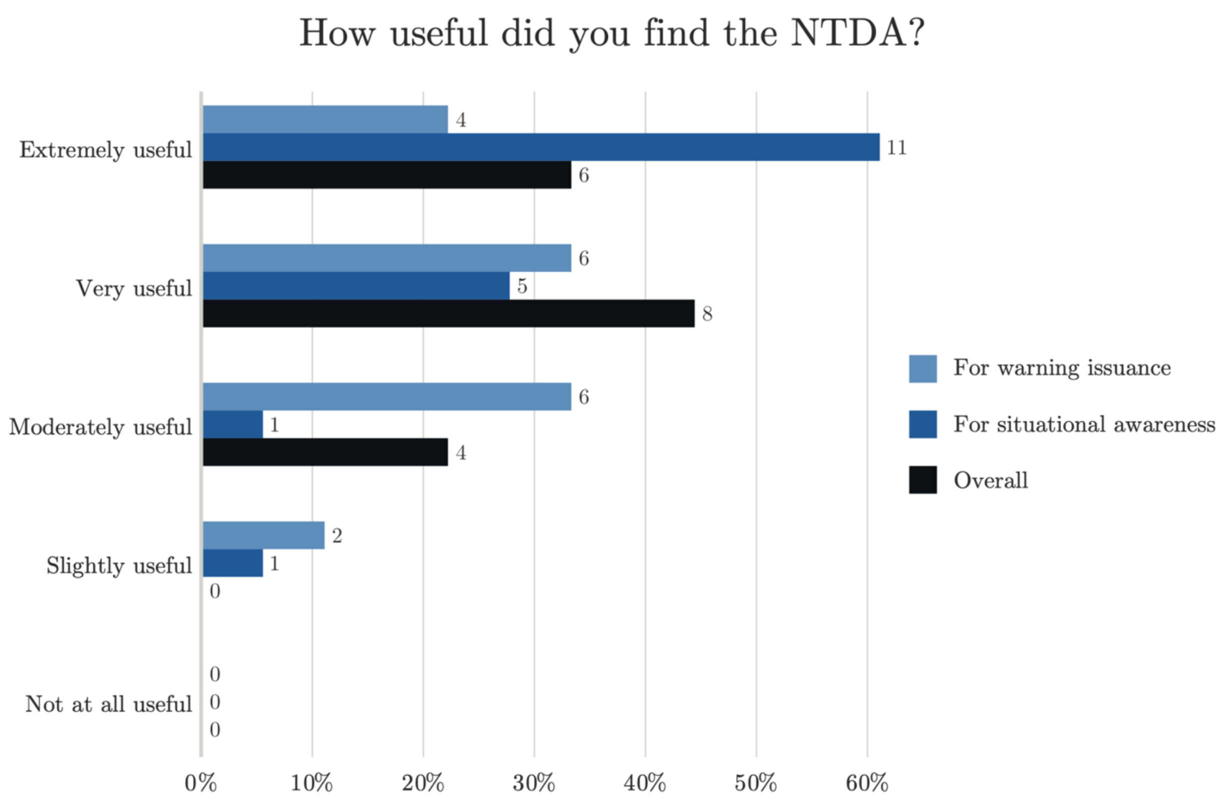


FIGURE 3. Survey responses about the usefulness of the NTDA.

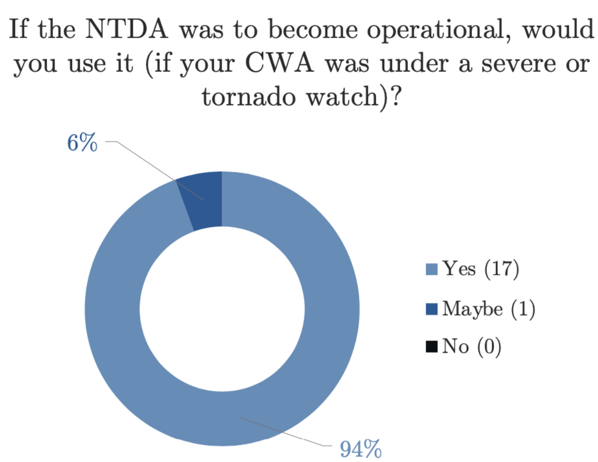


FIGURE 4. Survey responses about using the NTDA in operations.

There was agreement that the NTDA's ability to detect tornadoes was good for the majority of the time (Fig. 5), and below are some of the forecasters' quotes when answering why they would use the NTDA:

"I came into this week a bit skeptical as I've found the legacy version not very useful. Well, after a week of using the NTDA in various geographic regions and with different storm types, I am very

impressed with this new version and it has exceeded my expectations."

"I have no reservations about the NTDA becoming operational with only minor revisions."

"The probabilities generated through the random forest analysis were very useful, but the false alarm detections could be distracting in cases with several storms."

"On the merits of the algorithm itself, I am already comfortable saying that it would be useful in operations."

"I am all in on the NTDA."

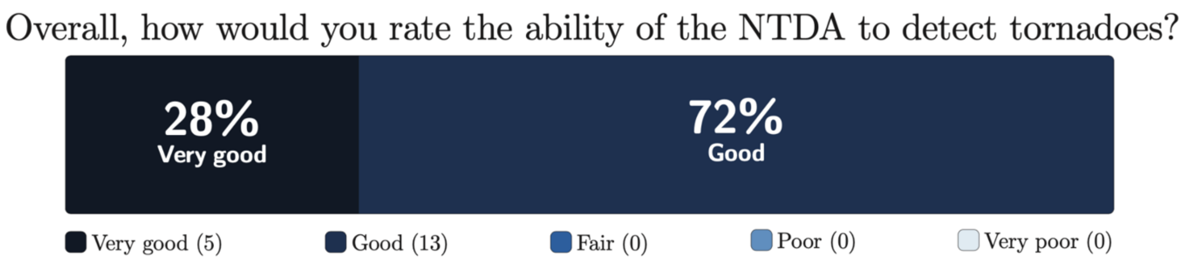


FIGURE 5. Survey response ratings of the NTDA's ability to detect tornadoes.

Comparison with TDA

After the 2019 HWT evaluation of the NMDA, it became clear that very few forecasters still use the MDA on a regular basis. This proved true for the TDA as well (Fig. 6), and 28% of the participating forecasters had, in fact, never used the TDA due to its reputation of being outdated (Fig. 7). The forecasters often mixed up the "TDA" and "NTDA" names during the discussions, and due to this, combined with the reputation of the TDA, we recommend a name change before the product becomes operational. The name used internally has been the tornado potential algorithm or TORP. When asked whether the forecasters would use the TDA, the NTDA, or neither, all of the forecasters that were asked replied that they would use the NTDA (Fig. 8), some citing the large screen area taken up by the TDA table and the number of false alarms with it. Below are some direct quotes from the participants when the forecasters that had used the TDA before were asked to compare it with the NTDA:

"The TDA is terrible. The NTDA is much, much better. Almost hard to compare because it's such a big difference."

"No comparison, the TDA has a ridiculous false alarm rate, and has become unusable."

"Light years better. Really hoping to see this and the NMDA in operations!"

"The current TDA is functionally useless as an algorithm for me to use. There are way too many false alarms and I find myself wasting precious time looking for the current TDA/TVS icon when

I actually had it on a display. I've since offloaded all of those from any procedures and do not recommend forecasters to use the current TDA/TVS feature."

"I am ready for an overhaul. I want to see this NTDA used."

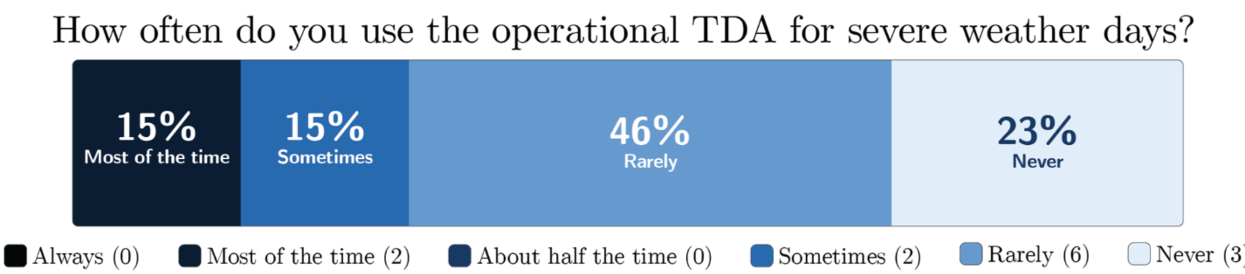


FIGURE 6. How often the forecasters (who answered that they had ever used the NTDA) use the current TDA in operations.

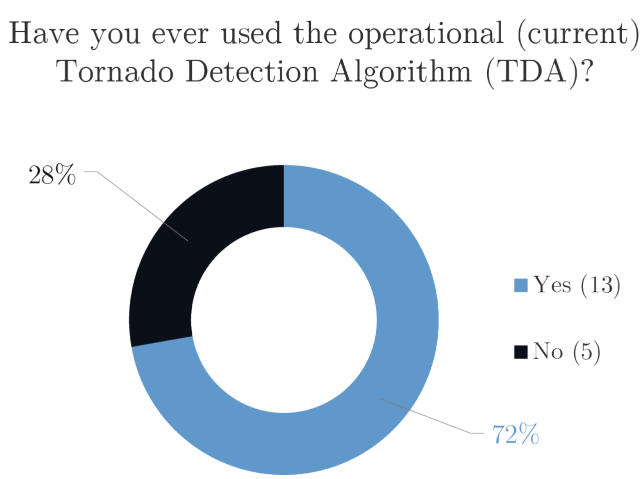


FIGURE 7. Survey of how many forecasters have ever used the TDA.

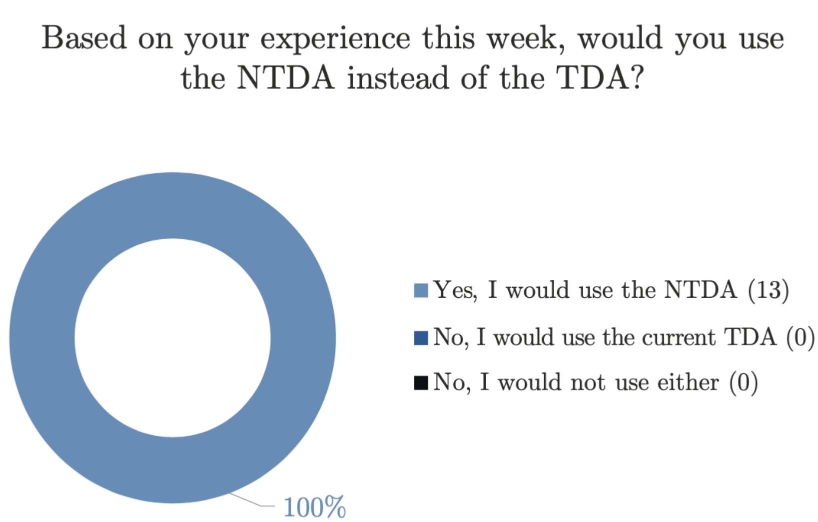


FIGURE 8. Weekly survey results about which tornado detection algorithm the forecasters would use if any.

Perceived Tornado-Warning Performance

One of the previously-mentioned suggested methods of evaluation for the algorithms was to simulate warning operations. Most forecasters expressed that the NTDA made them consider warning a storm that they otherwise would not have noticed, or at least help focus their attention to a region of interest in a linear storm system. Some pointed out that they probably would not have missed the storm or it did not directly influence their warning decision, but that the NTDA helped key them in on it for interrogation sooner than when pure base radar data would have grabbed their attention (Fig. 9 and 10). Another common theme was that the forecasters used probability guidance from the NTDA to nudge their decision to issue or hold off on a warning, or that they used it in conjunction with the three-ingredient method for QLCS tornado forecasting (Schaumann and Przybylinski 2012) to increase the confidence in their decisions.

The following quotes describe some examples of these situations that the participants experienced during the experiments:

"[F]or the DMX case, once I was able to re-calibrate the type of probabilities that were actually supporting tornadoes, I warned on a storm that I would otherwise never have thought to warn on given the radar fields. However, given the structure of that storm matched a previous storm, and had similar probabilities to a previous storm that did produce a tornado, I pulled the trigger on a tornado warning and soon after noticed that there was an LSR for it. Without the NTDA highlighting that storm, I would have missed it completely."

"In TSA's CWA [county warning area of NWS Tulsa], there were multiple notches in the line that weren't totally obvious if you were just looking at reflectivity, and sometimes not very obvious even looking at V/SRM [velocity/storm-relative motion]. I know that one such area that had increasing NTDA probabilities as it approached the AR border ended up producing a 31 mi long tornado track - from a QLCS!"

"Based on trends and radar presentation, I was leaning towards issuing a warning for a few storms throughout the week. When the next NTDA information came in, it helped move me towards issuing a warning. In most cases, it was the right thing to do although a few times it seems to have resulted in a false alarm but even without the algorithm the end result may have been the same with the FAR. So overall, I think the NTDA was more beneficial to me than harmful from a warning perspective."

"In at least one instance, I was able to use the NTDA to issue a TOR that I would not have otherwise"

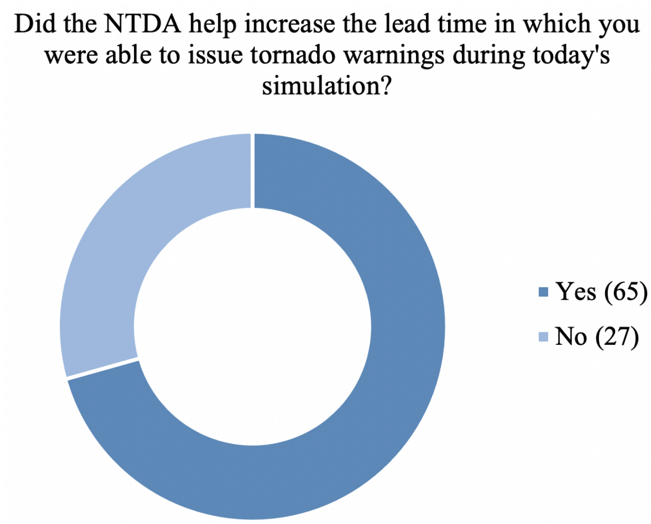


FIGURE 9. Case-by-case survey results for whether the forecasters thought NTDA helped in increasing lead time for a specific case.

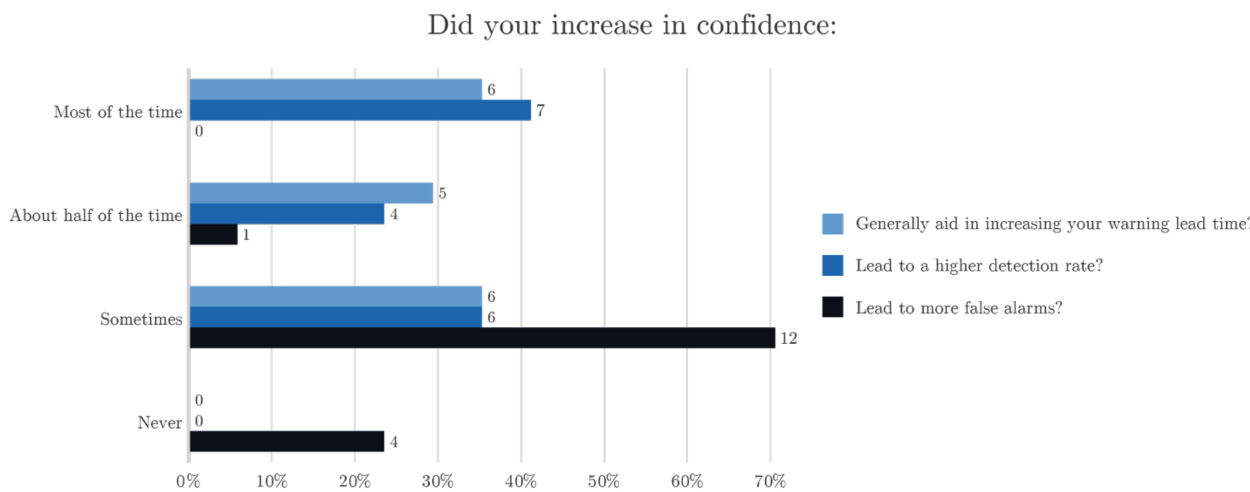


FIGURE 10. Weekly survey results by forecasters who answered that their confidence in warning decisions were enhanced by the NTDA, and their experience with that.

NMDA Results

With the NMDA having been previously evaluated in the 2019 HWT EWP Radar & Satellite Convective Applications Experiment, the 2021 HWT evaluation afforded the opportunity to obtain new feedback based on the significant improvements implemented to the NMDA, both in performance and visualization, that were based on the 2019 HWT findings. In the 2019 HWT, the NMDA was evaluated against the legacy MDA products with findings similar to those found between the NTDA and legacy TDA in this 2021 HWT experiment. The majority of participants had either never or rarely used the legacy MDA products due to their poor performance and difficult-to-use display. With this finding known, the NMDA was evaluated solely without the legacy MDA in the 2021 experiment.

Participants evaluated the NMDA on a number of different aspects that focused on the main areas of algorithm performance, impact to warning decisions, and the visualization package. The findings associated with these topics are described below.

NMDA - Performance Aspects

Focusing first on algorithm performance, the NMDA underwent several advancements between the 2019 and 2021 HWT evaluations that handled the detection building, tracking, and quality control elements within the algorithm. Comparing weekly survey results between the 2019 and 2021 HWT evaluations provides a quantifiable metric on which these algorithm improvements can be measured. When participants were asked to rate the NMDA's ability to detect individual mesocyclones, ~83% of 2021 participants provided a "Good" or better rating, a ~16% increase over the 2019 evaluation (~67%, Figure 11). Similarly, a ~14% increase between the 2019 (~58%) and 2021 (~72%) evaluations occurred for those participants that provided a "Good" or better rating when asked about the NMDA's ability to track an individual mesocyclone. While not large increases, it shows that the additional improvements to the NMDA are trending in the right direction. Findings from the 2021 evaluation will build upon those improvements as the NMDA is slated to be completed in 2022.

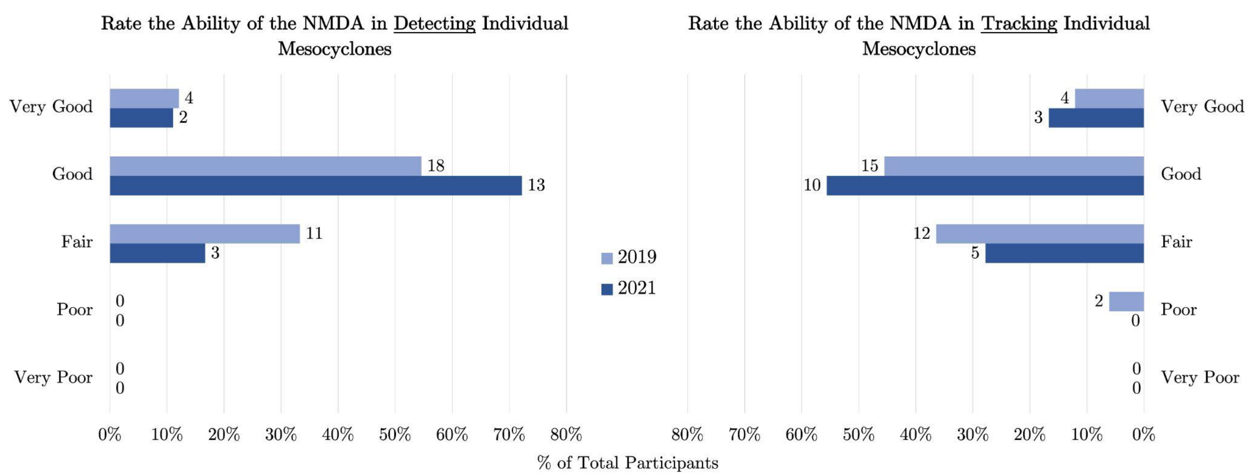


Figure 11: Weekly survey results from the 2019 and 2021 HWT evaluations comparing the participant's ratings of the NMDA's ability to detect and track an individual mesocyclone.

Along with their ratings, participants were also asked to provide a brief description outlining the reasoning behind them. Even though ~83% and ~72% of participants provided a "Good" or better rating for the NMDA's detecting and tracking abilities, respectively, only ~11% (detecting) and ~17% (tracking) of those were associated with a rating of "Very Good". Examining the participant's reasonings for ratings of "Good" or lower, many mentioned false alarm detections as the contributing factor to their specific rating. A user-adjustable detection thresholding tool was implemented in Weeks 2 & 3 of the experiment that helped participants reduce the amount of false alarm detections generated by the NMDA. While this thresholding tool worked, reducing false alarm detections internally in the algorithm before they are output will be a main point of emphasis in the final year of NMDA development.

NMDA - Impact on Warning Decisions

The impact that the NMDA had on a participant's warning issuance is another metric that was evaluated during the 2021 experiment. One way to quantify this is through asking each participant how their confidence changed over the course of the experiment week in using the NMDA during the warning decision process. Figure 12 shows that 15 of the 18 total participants had varying degrees of increased confidence, largely due to utilizing the algorithm over the multiple evaluations that covered different geographic locations and storm types. This helped them to understand the strengths and weaknesses of the algorithm, allowing them to mentally calibrate themselves and become familiar with how to use the NMDA in the warning process.

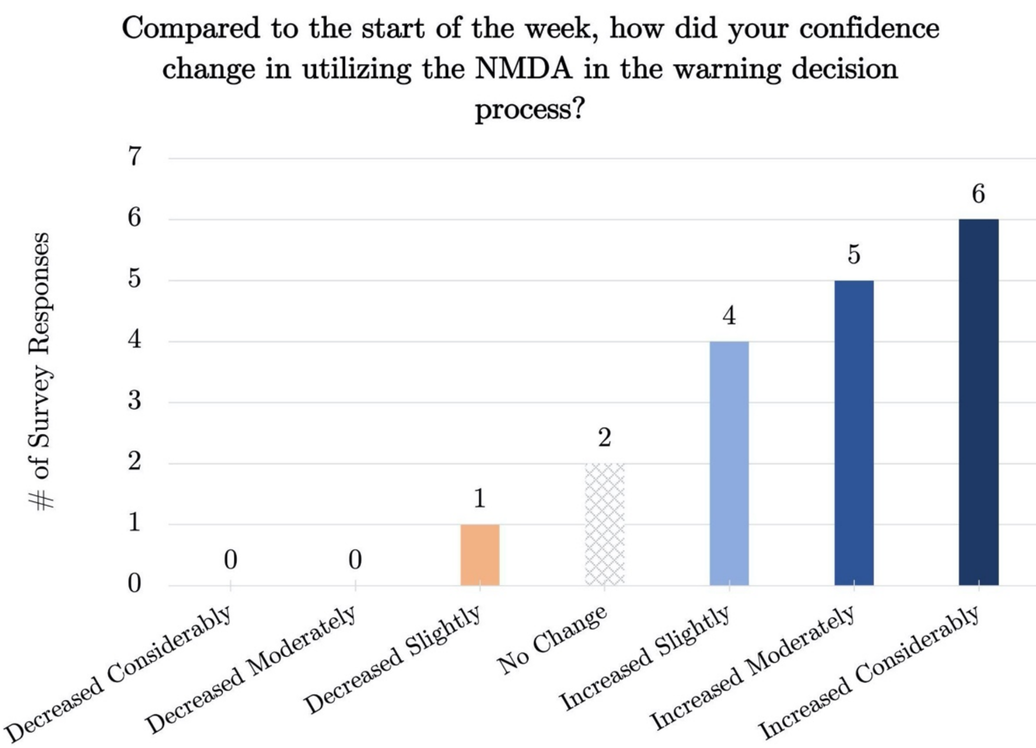


Figure 12: Weekly survey results describing how participant's confidence changed throughout each experiment week in using the NMDA in the warning decision process.

Further examining the participant's survey comments show that the NMDA was mainly used in the warning process in two ways: 1) as a situational awareness (SA) tool and 2) a confidence builder when already considering a warning. Especially in complex situations or for those participants who were playing the role of a mesoscale analysis, the NMDA helped as an SA tool to more quickly bring their attention to strengthening storms and/or decide which storms to further interrogate first. As a warning confidence builder, the degree of increased confidence during the warning issuance did vary depending on the situation. For more straight-froward cases, such as those containing isolated supercells, the NMDA provided less confidence since the participants normally were more sure with their warning decision from what they were visually seeing with the base radar data. The NMDA generally helped increase confidence for those less-certain situations such as QLCS circulations or marginally severe storms where the NMDA might highlight something the participant wasn't seeing or might have missed in the base data.

The following quotes further describe the participant's thoughts regarding the NMDA's use in the warning decision process:

"I really liked overlaying NMDA detections from several radars, as this helped my situational awareness, and helped me prioritize storms in the SRAD [Screen, Rank, Analyze, Decide] process."

"...there were a few instances (especially in some of the more marginal cases), where it drew me to features that I otherwise wouldn't have keyed in on."

"I never made a warning decision solely from the NMDA, rather confidence at the time of the warning decision or after. That extra confidence boost was extremely helpful."

"It was helpful in identifying areas of QLCSs where some mesocyclone development was beginning to occur or becoming mature, which can help increase confidence in a warning decision to some degree. For discrete supercells, there did not seem to be a significant increase in confidence as a result of the NMDA algorithm output."

NMDA - Operational Readiness and Additional Improvements

With the final NMDA version slated to be transitioned to the ROC in FY22, it was important to learn from the participants that, if it were available, would they operationally use the current version of the NMDA. Overwhelmingly, 17 of the 18 participants stated that in its current format, they would utilize the NMDA operationally, with many of those saying they would use it at least in a situational awareness framework. The ability for the NMDA to help draw their attention to areas of interest, especially in messy or mixed mode storm environments, is a strong suit of the NMDA that has an immediate operational impact to those serving as a mesoanalyst forecaster. However, as a warning operations tool, while a couple of participants said they would utilize it, many participants stopped short of this statement for a multitude of reasons. Specifically related to the algorithm, only a few stated that their reasoning was performance related, largely due to the false alarm issues that were mentioned in the previous section. The majority of reasoning centered on the fact that while it did increase overall confidence in the warning process, it didn't add any significant value in directly making warning decisions. Most veteran warning forecasters have an established tried and true method they use in warning operations that comes from their multiple years of experience. To that end, a warning forecaster wouldn't necessarily use the NMDA in directly issuing a warning, instead relying heavily on their own analysis of the base data to make those decisions.

To most appropriately leverage the remaining year of work on the NMDA, participants were also asked what they thought was the single most important aspect that could be improved with the NMDA. Responses largely boiled down into three main categories: 1) visualization updates, 2) reducing false alarms/weak detections, and 3) improved upper-level detection capability. Visualization updates include adding capabilities to the AWIPS-II plugin that improve the display of the trend information and allow a more concise way to view, locate, and sort all available detections. From a performance standpoint, reducing false alarms and weak detections will improve the overall usability of the product. Additional work in making sure the full vertical extent of mesocyclones is sampled by the detections will allow end-users a more complete picture of what is occurring aloft.

As a final note, during the post-HWT analysis process, the project PI noted that participants frequently stated the legacy MDA name when referring to the NMDA. Due to the poor reputation that the legacy MDA has with NWS forecasters, this creates a problem for the NMDA as that reputation from the legacy products might unknowingly become pre-associated before an end-user has had a chance to make their own opinion. In an effort to sever these ties with the legacy algorithm, the project PI suggests that a name change occurs for the NMDA before the product is deemed operational. Efforts are underway to develop a suitable name replacement that still encapsulates the algorithm's intended purpose.

Visualization

The visualization plugin for the NTDA and NMDA was created by Jonny Madden as imagined by PIs Sandmael and Smith. There were a lot of new features compared to the legacy algorithms and many opportunities for the users to customize their NTDA and NMDA experience. The forecasters enjoyed the visualizations overall and provided many terrific suggestions in further improving them that would benefit their use in the operational warning environment. For the NTDA, a suggestion to the default algorithm readout based on feedback is shown in Fig. 13, which contains the old default list of variables (bottom) and a suggestion for a new default (top). Similarly, the NMDA used the same visualization plugin as the NTDA but displayed a slightly different list of variables for participants to use that were more applicable in identifying the attributes of a mesocyclone.

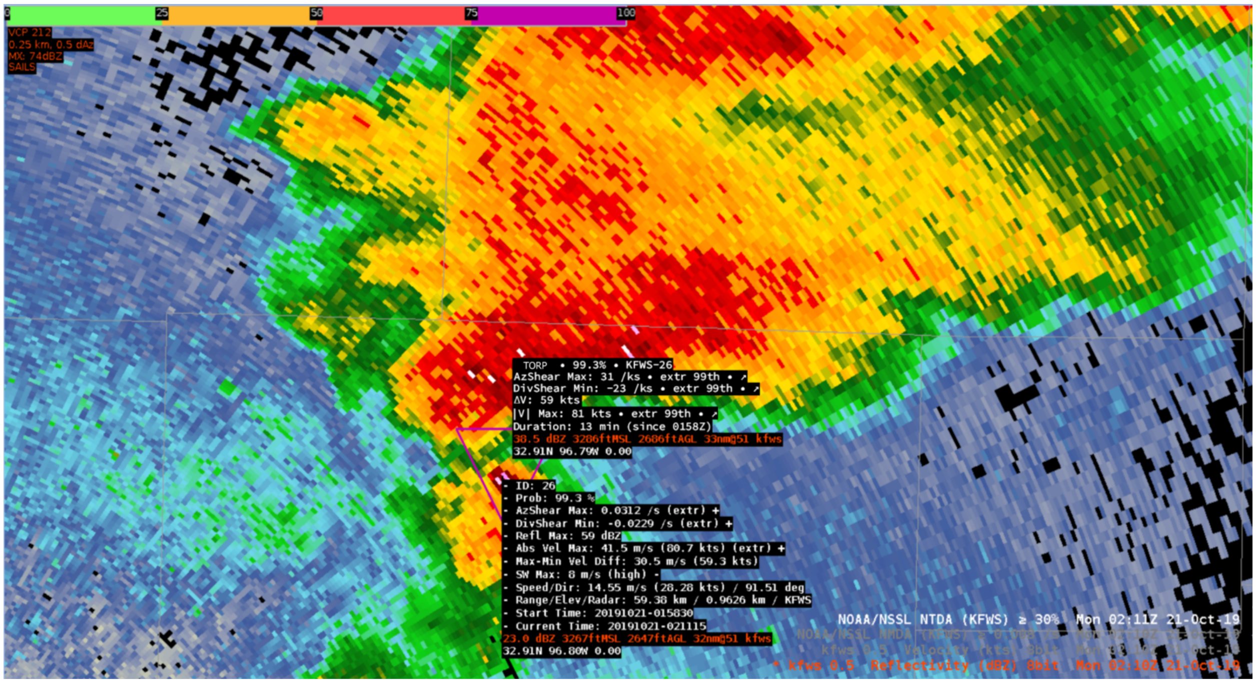


FIGURE 13. Old and suggested new default NTDA output of the hover-over variable list.

Future Plans

NTDA

A list of action items based on the HWT forecaster feedback was compiled, and work to complete updates was funded by the ROC for FY22. No new experiment is planned at this time, but one has been proposed as a new ROC task for FY23.

NMDA

HWT forecaster feedback was used to create a list of final improvements to be implemented in the algorithm, many of which were described in the previous sections. Once completed, the code will be finalized and transferred to the ROC in FY22. No new experiment is planned at this time.

Project Contacts

Thea Sandmael	CIWRO and NOAA/NSSL	PI / Developer
Brandon Smith	CIWRO and NOAA/NSSL	PI / Developer
Jonathan Madden	CIWRO and NOAA/NSSL	Tech support
Justin Monroe	CIWRO and NOAA/NSSL	Tech support
Patrick Hyland	CIWRO and NOAA/NSSL	Project observer
Benjamin Schenkel	CIWRO and NOAA/NSSL	Project observer

Probabilistic Hazard Information (PHI) Prototype Experiment

Summary by Kristin Calhoun, Thea Sandmael, and Clarice Satrio

Overview

The 2021 PHI experiment examined new algorithms for first-guess probabilities and tested different approaches to handling forecaster workload during severe weather. Of particular interest was determining the usability of the new PHI-tornado, an algorithm utilizing machine learning to derive tornado probabilities based on 8 years of tornado data developed by Sandmael et al. 2021. Additionally, multiple prototype tool aspects were developed to handle aspects better including expected deviant storm motion and to better separate between "steady state" and "evolving" storm properties.

The virtual experiment included three forecasters per week across three separate weeks using a combination of archive events and live data for testing. Some cases (archive and live weather) were constructed such that forecasters covered all three hazards (severe, tornado, lightning) for one or two storms. Other cases had multiple forecasters over the same domain where one forecaster was responsible for a single hazard over the entire domain, often about the size as a current NWS County Warning Area (CWA). With this design, forecasters were typically responsible for lightning only, tornado only, or severe hazards alone.

Forecasters were able to choose between multiple tracked objects for each hazard and test the manipulation speed and motion uncertainty relative to each PHI threat-in-motion. In addition to automated guidance available within the PHI-prototype system, including ProbSevere (Cintineo et al. 2020), ProbLightning (Calhoun et al. 2018), and new PHItor algorithms, forecasters had access to real-time and archived data within the AWIPS-II cloud platform for storm interrogation and analysis.

Experiment Details and Results

Each Monday of the experiment, forecasters worked with developers and subject matter experts to better familiarize themselves with the tools, algorithms, and process of issuing PHI. Tuesday, Wednesday, and Thursday consisted of a combination of an archive case to begin the day followed by live weather in the afternoon. If the live weather had little to no potential of severe weather in the afternoon, another archive case was presented to the forecasters. Generally, cases were designed to be initially easier in the week with increasing meteorological complexity as experience in the system grew.

Archive and Live Weather Case Descriptions

Tuesday Morning:
Archive Case, 28 May 2019: Discrete supercell and multicell storms ahead of a growing line of storms over eastern Indiana and western Ohio. Forecasters were all given the same domain and

were told to focus on a single hazard for all storms across the entire area (Tornado, Severe, or Lightning).

Tues Afternoon:

Week 1- Live weather, 27 April 2021: Scattered storms including supercell storms across the west Texas, Lubbock domain. Large hail, wind damage, and isolated tornadoes occurred over the region. All three forecasters covered the same domain with each forecaster focusing on a different hazard (severe, tornado, or lightning).

Week 2 - Live weather, 11 May 2021: Large domain covering eastern and south Texas including Ft Worth, San Angelo/Austin, and Houston, Texas County Warning Areas (CWA). Elevated storm clusters in the northwest section of the domain with isolated supercells farther south. Each forecaster covered all hazards for the 1-2 storms in their specific domain.

Week 3 - Live weather, 25 May 2021: Large hail from isolated storms along the dryline in west Texas and scattered storms and embedded supercells along and ahead of the cold in central Wisconsin. Each forecaster covered all hazards over their individual domain including La Crosse, WI, Minneapolis, MN, and Lubbock/Amarillo, TX CWAs.

Wed Morning:

Archive Case, 15 May 2018: Discrete supercells ahead of a QLCS within New York/New England region. Forecasters were each given a separate area and told to cover all threats for their area.

Wed Afternoon:

Week 1- Live weather, 28 April 2021: Scattered storms including supercells along the dryline and other boundaries across central Texas. Large hail, severe wind, and isolated tornadoes occurred over the San Angelo region. A single forecaster covered multiple hazards for the entire Dallas/Ft. Worth domain, while the other forecasters shared responsibility over the San Angelo region (one covered severe/lightning and the other tornado) .

Week 2 - Archive case, 21 Oct 2019: Supercells ahead of developing QLCS in Oklahoma. Event focused on evening to overnight hours where storms were enhanced by the development of the low-level jet. Due to the nature of the activity, two forecasters focused solely on the tornado hazard (one within Norman CWA and one within Tulsa), while the third covered both lightning and severe threats over the Tulsa domain.

Week 3 - Live weather, 26 May 2021: Nebraska, central/western Kansas. Steep mid-level lapse rates supported supercell storms along and ahead of the dryline and in central Kansas. Tornadoes, large hail, and damaging wind occurred across the region. Each forecaster covered all three hazards for their specific domain.

Thurs Morning:

Archive Case, 19 July 2018: Multiple tornadic showers and storms across Iowa of varying intensity and storm motion. Forecasters were each given a separate area and told to cover all threats for their area. (Fig. 1).

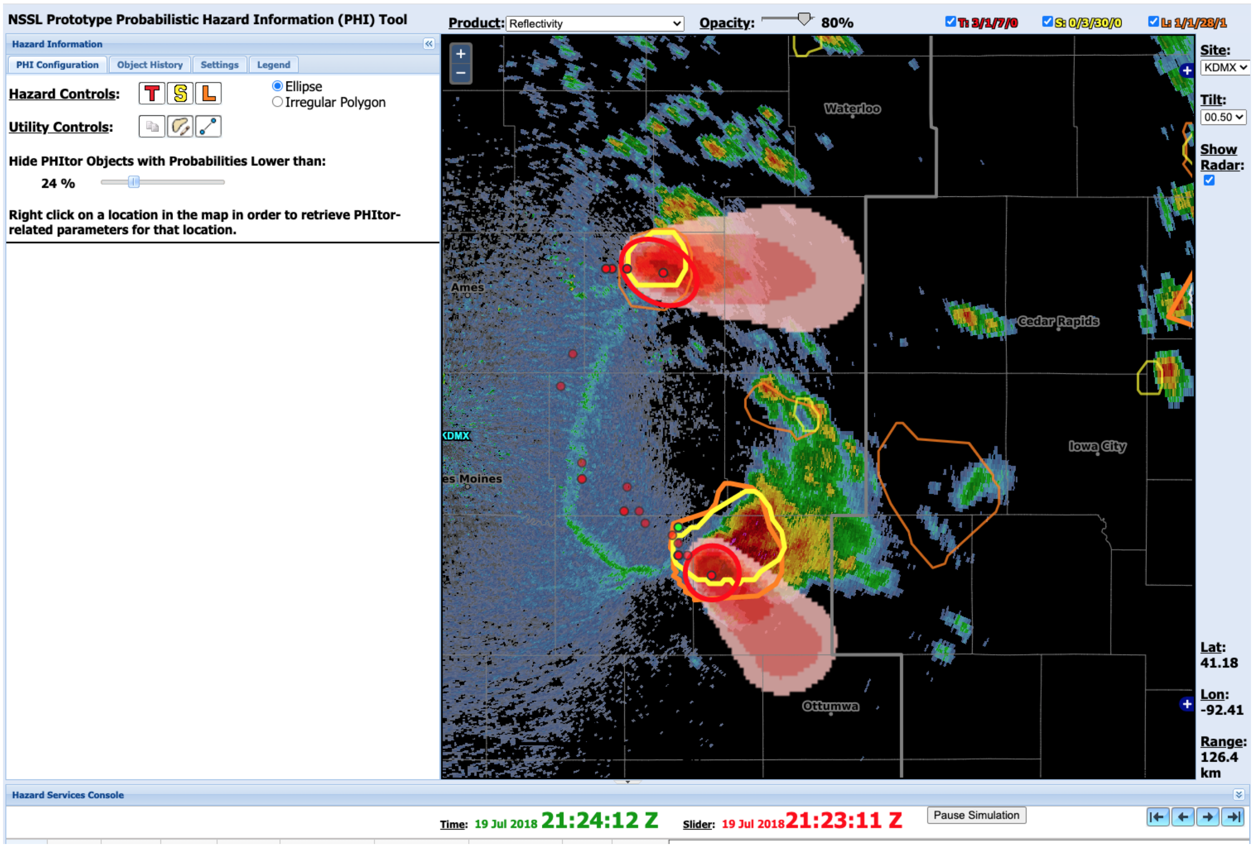


Fig. 1. Screenshot of the Prototype PHI tool during the 19 July 2018 Iowa archive case from week 1 of the experiment with 0.5 deg reflectivity from KDMX and tornado PHI plumes (red shading). Storm objects are outlined by hazard (red-tornado, yellow-severe, orange lightning) and storm reports are included as dots with most recent being brighter (red-tornado, green-hail, blue-wind).

Thurs Afternoon:

Week 1- Live weather, 28 April 2021: Central Texas, scattered storms including supercells along the dryline and other boundaries across central Texas. Large hail, severe wind, and isolated tornadoes occurred over the San Angelo region. A single forecaster covered multiple hazards for the entire Dallas/Ft. Worth domain, while the other forecasters shared responsibility over the San Angelo region (one covered severe/lightning and the other tornado) .

Week 2 - Archive case, 20-21 June 2019: Colorado, Wyoming, western Kansas. Isolated supercells (some elevated) with potential for large hail and isolated tornadoes, additional development during the event due to upslope flow. One forecaster covered the tornado hazard in Cheyenne, another the tornado hazard within the Boulder CWA, and the third covered both severe and lightning hazards over the Boulder CWA.

Week 3 - Live weather, 27 May 2021: Discrete supercells over central Oklahoma producing large hail with threat of strong winds and tornadoes. Forecasters were sectorized by hazard with one forecaster covering the severe threat, another tornado, and the third covered lightning.

Results

Summary Statistics

Forecasters were responsible for more individual storms when covering a single hazard (Table 1). Forecasters working this single hazard maintained more objects on average, typically with a lower number of updates per object times and more time between updates. In particular, lightning received significantly more attention when a forecaster was responsible for that as a single hazard. When maintaining multiple hazards forecasters typically triaged their responsibilities focusing first on the tornado threat, followed by severe, then updating lightning if time allowed. Most of the time, forecasters in this situation were comfortable allowing the automation to maintain the lightning coverage.

On average, forecasters had a lower initial probability to create tornado PHI objects compared to both severe and lightning (Fig. 2). While this specific probability varied by case, median initial tornado probabilities ranged from 30-60% whereas initial median severe probabilities were 80% and above (except for the Iowa archive case [DMX] which was lower due to lower likelihood of severe hail and wind). Median initial lightning probabilities were >90% across all the cases evaluated in the experiment.

Hazard type	Severe	Tornado	Lightning
No. of objects	8.67 (2.94)	7.75 (4.06)	10.33 (2.62)
No. of updates	14.67 (7.35)	15.75 (10.0)	22.67 (4.69)
Updates per object	1.69 (2.5)	2.03 (2.47)	2.19 (1.79)
Avg time per update (s)	196.64 (105.29)	139.22 (113.39)	105.96 (94.57)
Freq of update (min)	28.34 (18.09)	18.15 (10.81)	16.63 (19.12)

Table 1. Task analysis when working single (multiple) hazards for cases during the experiment. "No. of objects'' provides the average number of unique objects the forecasters controlled during a single case. "No. of updates'' is the average number of updates issued by the forecaster for both manually created or automatically created objects. "Updates per object" is the average number of updates for each unique object. "Avg time per update'' is the average duration (in seconds) it took a forecaster between clicking to update an object and clicking the ''issue'' button. "Freq of update" is the average time between updates for each unique object in minutes.

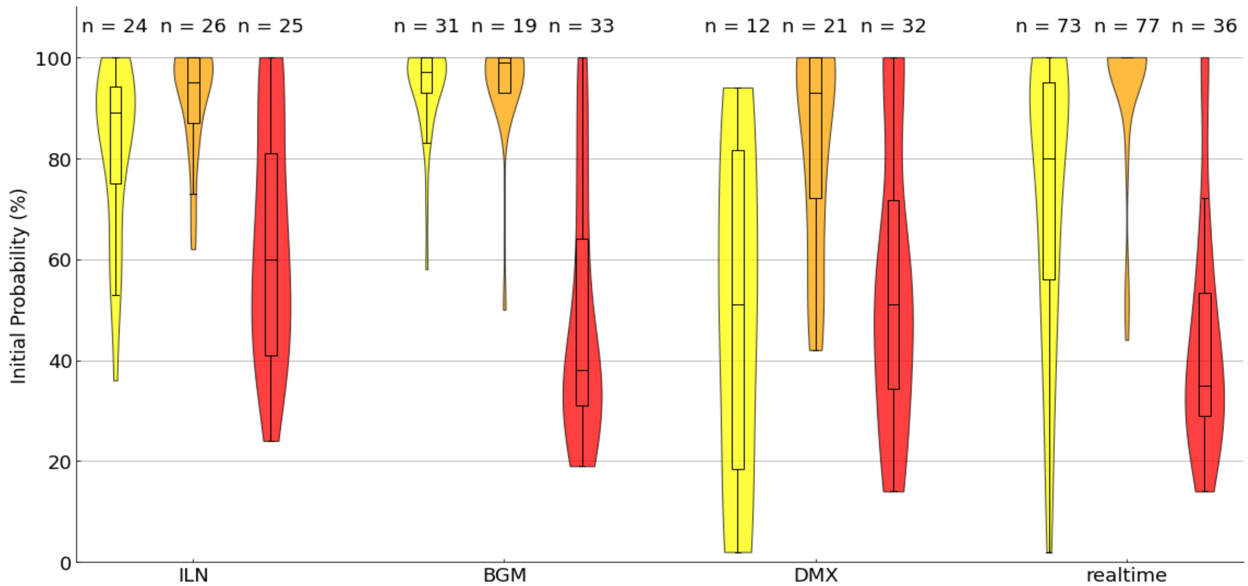


Fig. 2. Initial probability of objects issued by forecasters (both automated and manual) for each of the archive cases worked by all forecasters as well across the realtime (live) weather activities.

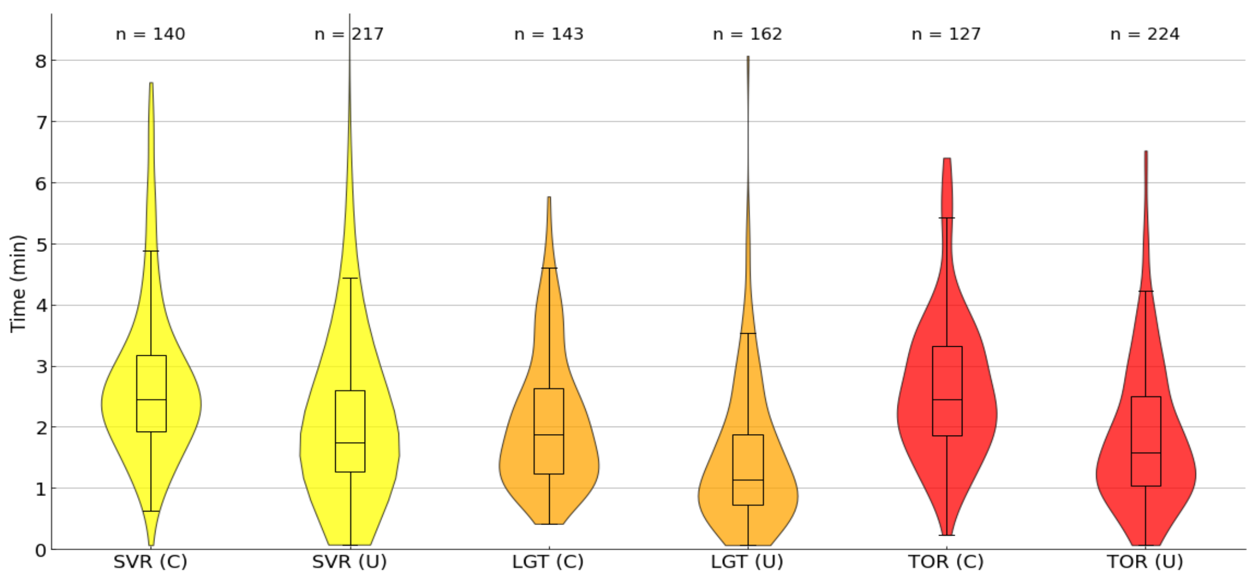


Fig 3. Amount of time taken to create (C) or update (U) a PHI object, according to hazard for all cases during the experiment. Counts for each are labeled at above the violin plot (n = count).

Forecaster creation and update times throughout the experiment (Fig. 3) were similar to previous experiments in 2016–2017 (Karstens et al. 2018) with creation taking on average a minute longer than an update. The distribution for creation and update durations were similar for severe and tornado hazards (median of ~2.5 min to create; 1.5 min to update). This was unlike past experiments where tornado creation took longer than severe, possibly due to the lack of automated tornado objects in previous experiments. Median times to create and update lightning objects were 30-45 sec < the tornado/severe objects, likely due to additional trust in the lightning algorithm performance and lower prioritization of lightning than severe or tornado hazards.

For each of the hazards, forecasters had potential automated objects to use to create PHI, though they could create their own object, if desired or deemed necessary. Forecasters were encouraged to add a buffer (additional coverage area) to both severe and lightning automated objects. For severe objects, the buffer size was typically between 0-10 km radius, with a median of 2 km. Future development may consider including a 2-5 km buffer on ProbSevere objects by default. Forecasters frequently added larger buffers to lightning objects to cover possible cloud-to-ground flashes outside of the storm cores and into stratiform rain areas with outliers extending to 40-60 km. Unlike severe, these are harder for an automated system to anticipate and may need to be developed within the storm tracking algorithm or added manually by forecasters.

Evaluation and use of the new PHItor Algorithm

The new PHI tornado guidance product (PHItor) provided automated tornado objects for the first time in HWT evaluation (Fig. 4). The algorithm is based on high rotation detected by single-radar azimuthal shear and uses a random forest machine learning model to summarize storm-based single-radar data for tornado potential. In addition to the object and associated probabilities, forecasters were able to examine trends in single-radar data associated with the object. This included AzShear, DivShear, Spectrum Width, and the Absolute Maximum of Velocity. In addition to the automated objects, forecasters could also right-click on any point on the PHI Tool map to evaluate that location for tornado probability (Fig 5). Based on this query, forecasters could then create a new tornado object at that location with these starting values.

Based on survey results and discussion, most forecasters found the addition of PHItor very or extremely useful within the tool. However, most forecasters used it primarily for situational awareness and determining which areas to interact with as opposed to using it primarily for PHI plume creation. In regard to PHI plume creation, forecasters often became frustrated with unstable storm motion and lack of location consistency, particularly with storm mergers, splits or occlusion processes. Suggestions for future use include either copying values from more stable Severe objects for storm motion or removing the default ability to use automated values for storm motion and location and require forecasters to set this manually. Forecasters also valued access to the trend graphs for radar data associated with the algorithm; 77% of responses noted these trends were "very" to "extremely useful." Overall, forecasters responded that the algorithm helped increase their confidence in decision-making and that this confidence sped up their decision-making process and led to a higher detection rate. However, most forecasters noted this sometimes could also lead to more false alarms.

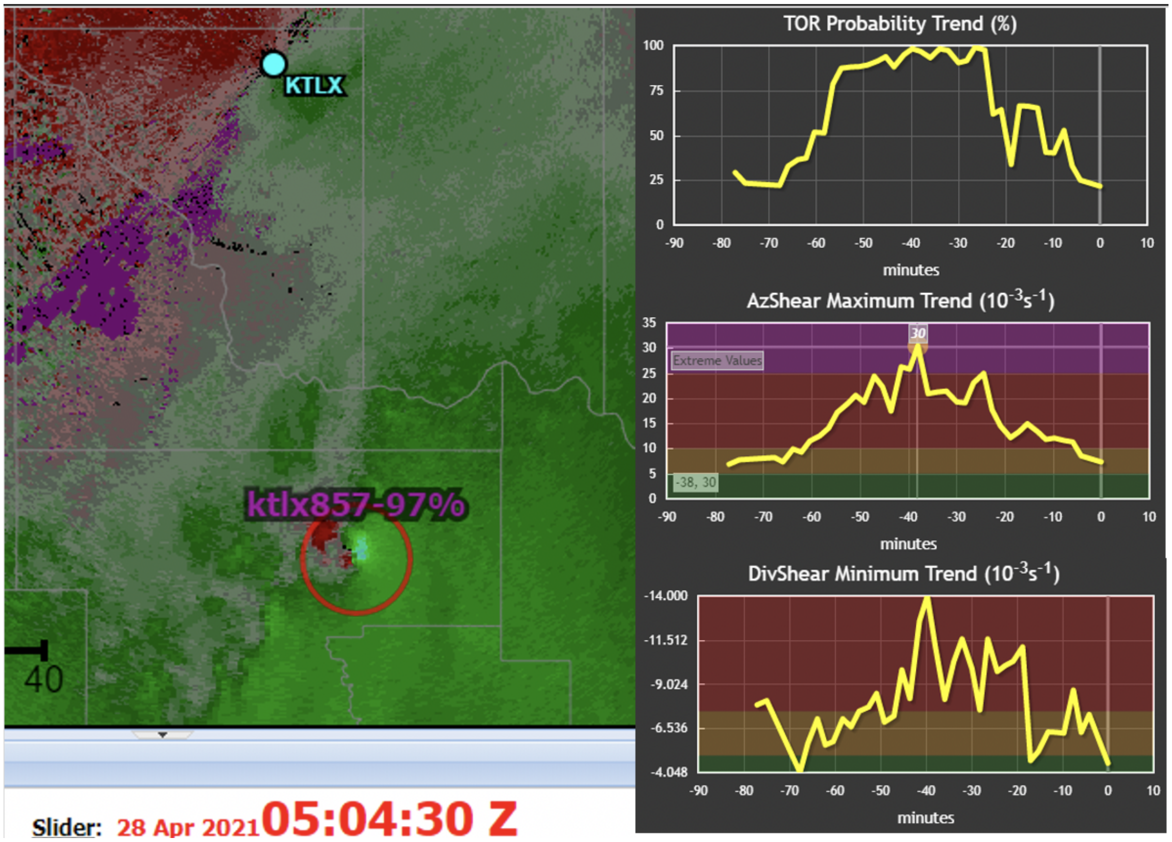


Fig. 4. Left: PHItor object (ktlx857) with 97% probability overlaid on KTLX velocity. Right: Associated trends in Tornado Probability (top), AzShear (middle), and DivShear (bottom). Background color provides climatological relative guidance for the forecaster evaluation.

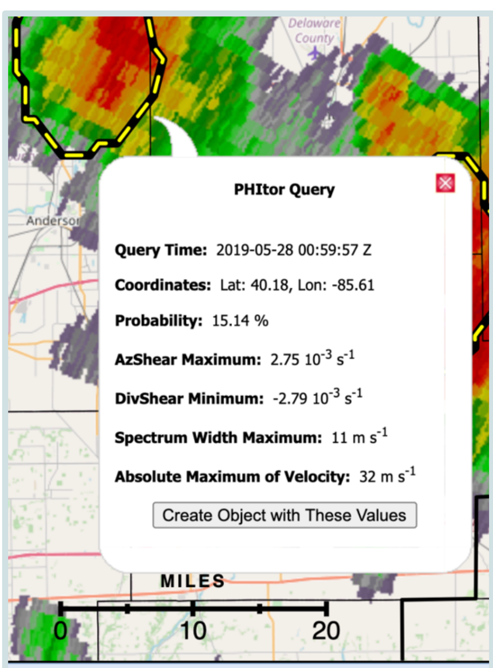


Fig. 5. Example of right-click query for a location that did not meet the threshold for automated PHItor objects. Query provides time, location, probability, and values for AzShear, DivShear, Spectrum Width, and Absolute Maximum of Velocity at that location. The forecaster can then create a new tornado object at that location, if desired.

Forecaster Workload

Forecasters completed NASA Task Load Index surveys following each archive case and live activity. When judged by individual hazard, forecasters focusing on the Tornado hazard rated their workload the highest, followed by severe, then lightning. The TLX scores for forecasters working all three hazards scored an average closer to Severe forecaster workload.

From discussions and survey questions, most forecasters believed that the overall workload for the cases depended more upon the complexity of the meteorology of a given event than how the hazards were separated (i.e., multiple hazards for 1 to 2 storms or a single hazard across an entire CWA). However, specific forecaster comments noted that managing all three hazards across a small area (1-2 storms) felt like a higher workload. This higher workload appears to also be measured in statistics across the cases (Table 1). When forecasters managed all three hazards for one to storms simultaneously, they were updating the severe and tornado objects more often (median value of 8-10 min less between updates) with more updates per object. It is unclear if these more frequent updates were made to correct issues that didn't occur when forecasters maintained a single hazard or if forecasters typically felt the need to update when rotating between hazards. Future experiments should try to better understand what an ideal timing is for updating PHI and if that varies by hazard. Once established we may be able to better understand situational workload by hazard or by area better.

Summary and Future Plans

Overall, the machine learning algorithms (ProbSevere, ProbLightning, and new PHItor) provided a necessary probability baseline and first guess guidance. This helped forecasters summarize data, prioritize hazards, and sped up decision making. With the addition of the new PHItor algorithm, creation time of tornado PHI was roughly the same as severe PHI. This is unlike previous experiments where tornado PHI creation took on average 50% longer than severe.

While updating objects took about a minute less than creating objects, after observing forecasters throughout the experiment, we believe this could be done faster and with less burden if the update system were redesigned. Currently, forecasters follow a similar pattern to update as when they create an object. However, these updates are often to adjust a single aspect of the storm such as the storm motion or the probability. Future development may consider a way for forecasters to update this information independently without requiring forecasters to update all elements. This would likely further reduce the time necessary to complete an update from a median value of 1.5 min to under 30 sec or less.

Project Contacts

Kristin Calhoun	NOAA/NSSL	Experiment Lead, Project PI
Thea Sandmael	CIWRO and NOAA/NSSL	PHItor PI and Experiment Scientist
Clarice Satrio	CIWRO and NOAA/NSSL	Experiment Scientist
Patrick Hyland	CIWRO and NOAA/NSSL	Experiment Scientist
Adrian Campbell	CIWRO and NOAA/NSSL	PHI Tool Development and Support
Rebecca Steeves	CIWRO and NOAA/NSSL	PHI Tool Development and Support
Justin Monroe	CIWRO and NOAA/NSSL	AWIPS Support
Jonny Madden	CIWRO and NOAA/NSSL	AWIPS Support

References:

Calhoun, K. M., and Coauthors, 2018: Cloud-to-ground lightning probabilities and warnings within an integrated warning team. AMS Annual Meeting. *Special Symp. on Impact-Based Decision Support Service*s, Austin, TX, Amer. Meteor. Soc., 4.4, https://ams.confex.com/ams/98Annual/meetingapp.cgi/Paper/329888

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020: NOAA ProbSevere v2.0—ProbHail, ProbWind, and ProbTor. *Wea. and Forecasting*, 35, 1523-1543

Karstens, C. D., J. Correia Jr., D. S. LaDue, J. Wolfe, T. C. Meyer, D. R. Harrison, J. L. Cintineo, K. M. Calhoun, T. M. Smith, A. E. Gerard, and L. P. Rothfusz, 2018: Development of a Human–Machine Mix for Forecasting Severe Convective Events, *Wea. and Forecasting*, 33(3), 715-737.

Sandmael, T., R. B. Steeves, P. A. Campbell, K. M. Calhoun, Z. A. Cooper, 2021: The Probabilistic Hazard Information Tornado (PHItor) Product: Using Machine Learning to Provide Probability Guidance for Tornado Nowcasting. *11th Conference on Transition of Research to Operations. AMS Annual Meeting*. Virtual. https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/380062

Satellite Convective Applications Experiment

Summary by Kevin Thiel and Kristin Calhoun

Overview

Satellite Proving Ground demonstrations in the HWT have provided users with a glimpse into the capabilities, products and algorithms that are and will be available with new updates, technology, and products available on both geostationary and polar-orbiting satellites. The education and training received by participants in the HWT fosters interest and excitement for new satellite data and helps to promote readiness for the use of satellite 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 the developer to understand how forecasters use the product and what improvements might increase the product utility in NWS operations. Feedback received from participants in the HWT has proven invaluable to the continued development and refinement of GOES-R and JPSS algorithms. Furthermore, the EWP facilitates the testing of satellite-based products in the AWIPS-II data processing and visualization system currently used at NWS Weather Forecast Offices (WFOs).

Experiment Details and Results

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

Typical feedback included suggestions for improving the algorithms, ideas for making the displays more effective for information transfer to forecasters, best practices for product use, suggestions for training, and situations in which the tools worked well and not so well. Most of the products evaluated in 2021 were advancements of previous product iterations from the 2019 GOES-R/JPSS Proving Ground. This included data from the Geostationary Lightning Mapper

(GLM), the Probability of Severe (ProbSevere) model – Version 3, and the NOAA Unique Combined Atmospheric Processing Systems (NUCAPS). The Optical Flow Winds product, created from the GOES-R series ABI, was evaluated in the HWT for the first time. Forecasters viewed the GLM, NUCAPS, and ProbSevere data in the cloud-based instances of AWIPS, and the Optical Flow Winds product was available in a web-based interface.

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

Based upon the evaluation of GLM, NUCAPS, ProvSevere, and Optical Flow Winds data in the 2021 Satellite Convective Applications Experiment, the following recommendations have been made for each of the demonstrated products:

GLM

- It is recommended that the NWS continue to emphasize training forecasters on GLM-related products, often through subject matter experts within their local NWS offices.
- It is strongly recommended that the gridded GLM Flash Extent Density product stay in the baseline (Level2) product, and that the Minimum Flash Area product be added.
- It is highly recommended that a storm-based time series display of GLM flash rates, and potentially flash rates from other ground-based lightning networks, be developed for use in AWIPS.
- It is recommended that the utility of the GLM flash point product and its associated meta-data continue to be explored.

NUCAPS

- It is strongly recommended that the use of NUCAPS in SHARPpy continue to be explored, along with training for increased use in the HWT.
- It is recommended that the creation of a merged-Gridded NUCAPS product, which combines data from overpasses in close temporal proximity, be explored to potentially increase product utility.
- It is recommended that future forecaster training efforts regarding NUCAPS focus on best display practices, especially with the Gridded NUCAPS product, so users can fully leverage the available temperature and moisture profiles.

- It is strongly recommended that relevant overpass times be clearly stated for each NWS CWA, such that forecasters know when data are being collected from each satellite and made available through NUCAPS.

<u>ProbSevere</u>

- It is highly recommended that development and implementation of the AWIPS ProbSevere time series tool into Version 3 continue, so forecasters can readily diagnose convective trends when making warning decisions.
- It is strongly recommended that ProbSevere Version 3 be implemented into AWIPS, for increased performance in both severe and sub-severe environments when compared with Version 2.
- It is recommended that future training of forecasters with ProbSevere Version 3 involve recalibration for new probabilities, the variables included for each hazard model, performance metrics for each WFO's CWA, and case studies from a variety of convective intensities and modes.

<u>Optical Flow Winds</u>

- It is strongly recommended that future research and development efforts of the optical flow winds product involve derived fields such as horizontal divergence and vertical wind shear, to increase the applicability of high resolution, satellite-derived wind fields.
- It is strongly recommended that additional training be provided to NWS forecasters regarding the use of satellite-derived winds when interrogating pre-convective environments and cloud top characteristics of convection.
- It is recommended that integration of the Optical Flow Winds product into AWIPS be explored, along with effective visualization techniques when used in concert with ABI imagery.

One new addition this year was the creation of simulated DSS events within operations, with forecasters providing useful and timely information for public safety. As a growing mission within the NWS, the creation of these events created a new evaluation opportunity that challenges forecasters to use the data beyond warning decisions. Interpretation and communication of experimental products within this framework further revealed forecaster understanding, which directly impacts a product's ability to be integrated into operations and its associated training. This year the experiment only featured two simulated DSS events, however expansion of DSS applications within the evaluation process may provide additional insight into direct applications of experimental products within the testbed. Additionally, a few forecasters did comment that pre-built procedures may have been helpful for increasing how soon forecasters could begin looking at products in a condensed operations schedule this year, along with following best display practices to fully leverage the experimental products. These recommendations have been documented and will be used in future Satellite Convective Applications Experiments.

Like all other HWT activities during the 2021 Spring Experiment, the Satellite Proving Ground was held entirely in a virtual format due to the COVID-19 Pandemic. When asked about their experiences in a virtual testbed, both developers and forecasters generally felt that the experiment still provided a valuable opportunity to test new and developing products. However, those with experience in previous iterations of the satellite testbed consistently said that there were some limitations to this year's entirely virtual experiment that impacted the quality and quantity of feedback received. One of the most obvious limitations was the compressed operations schedule with static start and end times. This was done to reduce forecaster screen time and fatigue throughout the week, with still a few forecasters commenting that mild physical and mental fatigue was noticeable by the end of the week. Time in operations was further reduced by various technical difficulties associated with each participant's unique computing setup, and often required constant technical support from CIWRO IT staff involved in the experiment. Adapting to the cloud-based AWIPS-II instances also took time on the first day but was often a minimal disruption, as the instances were reliable throughout nearly all the experiment. Some participants noted the convenience of not having to arrange the logistics of traveling to Norman, Oklahoma for the Spring Experiment. Additionally, a few forecasters noted that they typically were unable to participate in an in-person experiment due to personal commitments that limited the time they could spend away from their area of residence. Under these scenarios, a testbed with more remote elements may allow for a more diverse pool of applicants across the NWS or other sectors than previously observed.

Future Plans

Based upon the successes and lessons learned from the 2021 Satellite Convective Applications Experiment, the 2022 experiment will retain a similar format from the previous year with minor additions and adjustments. Three full weeks of virtual evaluations are planned to take place between 1 June and 18 June of 2022. For 2022, we have increased the number of available participants to 24 (8 per week) from 18 (6 per week). The expectation for this increase is greater survey feedback and collaborative blog posts, along with more discussions amongst forecasters while in their simulated WFOs. One concern expressed was adding too many forecasters, such that some volunteers may be unable to provide adequate feedback if the weekly forecaster groups were larger than 10, especially during the daily debrief periods. Previous methods of feedback such as daily and weekly surveys, blog posts, debrief sessions, and open discussions during operations provided sufficient insight into key forecaster recommendations, and will also be employed again. During operations, the PIs plan to provide pre-built procedures in AWIPS so the forecasters can more quickly view data and leverage the best display practices recommended by the product developers. Additionally, more DSS events planned to be held while in operations to provide another perspective of product applications and limitations.

Web Presence

- 2021 GOES-R/JPSS Satellite Proving Ground Final Report
- GOES-R Proving Ground
- EWP Blog Posts
- NWS Satellite Book Club Presentation

Project Contacts

Kristin Calhoun	NOAA/NSSL	Project Co-PI
Kevin Thiel	CIWRO and NOAA/SPC	Project Co-PI
Justin Monroe	CIWRO and NOAA/NSSL	AWIPS Support
Jonny Madden	CIWRO and NOAA/NSSL	AWIPS Support

Hazard Services - Threats-in-Motion (TIM) Virtual Experiment

Summary by Greg Stumpf and Chen Ling

Overview

Current severe weather warnings (tornadoes, wind, and hail) require the forecaster to issue multiple sequential warnings for long-tracked storms because the current policy prohibits extending a warning's area and time during updates. This situation frequently results in non-uniform lead times for users on the downstream border of a warning polygon. For example, nearly adjacent locations can have dramatically different lead times if one location is just outside the upstream warning.

Threats-In-Motion (TIM) is a proposed warning decision and dissemination approach that would enable the NWS to upgrade severe thunderstorm and tornado warnings from the current static polygon system to continuously-updating warning polygons that move with the storm. TIM is a proposed first evolutionary step of Forecasting A Continuum of Environmental Threats (FACETs) for the convective weather warning scale. With TIM, the forecaster would only need to issue a single warning, updated regularly as workload permits, embodying a "one storm-one story" concept. This approach would reduce forecaster workload because downstream warning issuance would be replaced by a less time-consuming warning update. In addition, TIM provides a continuous history for each storm, which would lead to simplified and consistent messaging for key partners and improving event verification.

The most significant benefits of TIM are from improvements in hazard communication. If implemented, TIM can provide more equitable (uniform) lead times for those in the path of long-tracked severe storms because these storms remain continually tracked and warned. As such, TIM mitigates gaps in warning coverage and improves the handling of storm motion changes. In addition, warnings could be automatically cleared from locations where the threat has passed. This change would result in greater average lead times and decreased average time spent in a warning relative to today's warnings, with little impact on average false alarm time. This impact is particularly noteworthy for storms expected to live longer than the average warning duration (30 – 45 minutes), such as the long-tracked supercells seen during violent severe weather outbreaks. A robust statistical analysis of TIM's scientific benefits is available in Stumpf and Gerard (2021).

Efforts have been underway since 2019 to develop the software capability to issue TIM for convective weather warnings (tornadoes, wind, and hail) within the Advanced Weather Interactive Processing System (AWIPS) Hazard Services. This new software, known as HS-TIM, was tested in the HWT in a limited sense in 2019. NWS/MDL, in partnership with NSSL, GSL, WDTD, and U. Akron, carried out another HWT experiment in the summer of 2021. This was the first experiment that focused 100% on TIM, allowing us to explore several ideas to represent realistic challenges currently faced in warning operations in order to focus on workload differences with the current NWS warning system known as AWIPS – WarnGen. Only forecasters participated (no end users), so the resulting feedback is from the operational NWS forecaster perspective. The major differences between WarnGen and HS-TIM that forecasters had to learn and get used to are:

- The Hazard Services screen layout (Spatial Display, Console, Hazard Information Dialog). This will be the layout for the upcoming WarnGen replacement (HS-Convective), and is already available in other HS perspectives (e.g., HS-Hydro), so some of the forecasters already had some experience.
- Using 2D storm "objects" (versus points and lines) to define and track current threat areas, and to project the future threat areas at 1-minute intervals to create the warning polygon swaths. This includes new 2D drawing tools (polygon, freeform, ellipse) that are currently unavailable in WarnGen.
- The "Persist" feature, which, when enabled, places a warning "in motion", updating the location every minute. For persisted warnings, during subsequent warning updates, the 2D objects can be quickly repositioned and reshaped to continue the warnings indefinitely or until the storm dissipates. This process consumes far less workload than re-issuing brand new warnings each time a warning expires.
- A Warning Decision Discussion (WDD) which allows the warning forecaster to add their warning decision thoughts about why they issued or modified the storm object (e.g., "mid-level rotation is strengthening"). These are thoughts that are typically not included in the actual text of a warning product, but may be relayed via NWSChat or similar end-user communication software. Because TIM warnings use the same ID throughout the lifetime of the storm, the WDD history for the storm is linked from the first time the warning is issued, giving a "story about the storm".

The HS-TIM experiment was carried out for four weeks in the summer of 2021, with the first week being a "shakedown" of the system using two "test" forecasters. The remaining operational weeks included a pair of participants from several WFOs nationwide. Each week saw forecasters learning how to use the software via a guided training exercise on the first day, and displaced real-time (DRT) severe weather scenarios on Days 2 through 4. The final day was spent conducting a 2-hour guided interview of the forecasters on their experience during the experiment week.

Because of the continuing COVID-19 pandemic restrictions, the experiment was conducted virtually, using a version of the AWIPS software hosted within the Amazon cloud services (see Fig. 1). There were some pros and cons to this approach:

- Pros: Developers had quick and convenient browser access to the cloud systems from anywhere, so that quick software tests could be performed; forecasters who typically cannot travel to Norman for a variety of reasons can now participate in an HWT experiment; developers could quickly diagnose problems, without having to ask forecasters to leave their workstations; each participant had a close up, high-resolution view of the workstations (rather than looking over shoulders or on mounted TV screens).

- Cons: There was no in-person social interaction outside working hours; we could not split instances to two monitors; workdays had to be shortened (due to “Zoom fatigue”), resulting in less time for forecasters to participate in hands-on scenarios

Archived data events were used to perform the evaluations; there were four cases from different locations nationwide; each had unique adjacent CWA domains and represented a variety of different severe storm types (e.g., squall lines, long-tracked tornadoes, etc.).

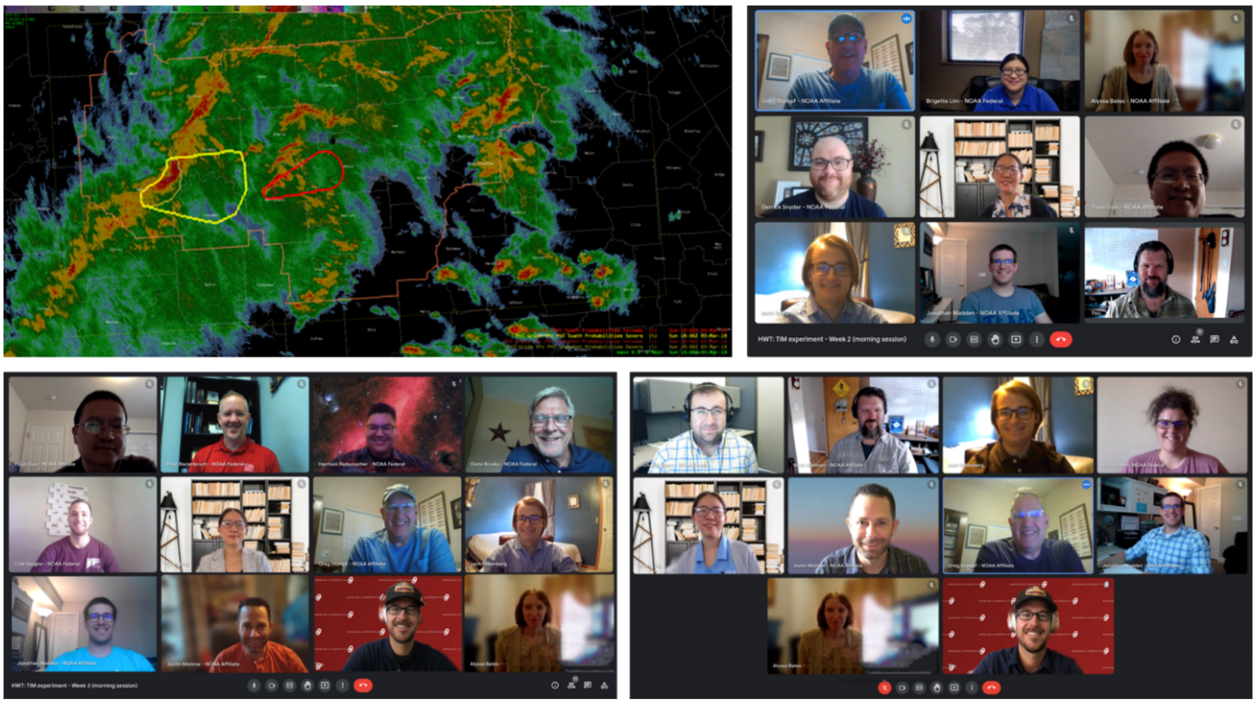


Figure 1. *Images from the HS-TIM experiment. Top left: HS-TIM output from a scenario in southeast Alabama - the yellow (red) polygons show severe thunderstorm (tornado) warnings; the large orange polygon is the county warning area (CWA). The remaining three images are group photos of the virtual meetings during each of the three weeks of the experiment.*

Since its initial limited test in 2019, the HS-TIM software had been robustly hardened, and thus remained stable during the experiment, causing little to no impact on the evaluation of forecaster workload. A number of major new functionalities were added to HS-TIM and evaluated during this experiment:

- "Drag Me To Storm" (a.k.a. "Reset Motion Vector") mode is now on for each update (saves some mouse clicks).
- Added a "Select Hazard Type" label to a new object before the hazard type is selected to prompt the user to take the correct action (saves some mental workload).
- Added a “Latest TIM Frame” button, because sometimes the most-current radar frame is more recent, and the TIM object is not visible.

- Uncertainty values can now be manually set to zero to account for steady-state storm motion situations without greatly fanning out the warning polygons. The default uncertainty values are only used on brand new objects.
- The motion vector is not re-calculated when the object shape is changed. This prevents drastic changes to the motion vector based on the calculation of a new object centroid.
- Changed the duration behavior for non-Persisting warnings (the End Time remains constant and the duration ages off), and Persisting warnings (the duration will reset to the previous manually-set duration; the duration will not age off and the End Time will increment one minute with each automatic update).
- Improved warning product formatting, with updated IBW tags for Severe Thunderstorm Warnings based on recent NWS directives (Aug 2021).

Several suggestions for improved software functionality were offered by the forecasters. Some of these suggestions will be incorporated into a future version (as funding and budgets permit) in order to make the software more robust. These suggestions are listed in Appendix A. In addition, the forecasters made some suggestions to improve future experiment logistics. These are listed in Appendix B.

As with previous experiments, this evaluation included human factor experts who recorded video and audio, and administered surveys and interviews to measure mental workload, confidence, and software usability. More information about the human factors analysis is included in Appendix C.

Key Takeaways

- Using 2D objects with Persist to create Threats-In-Motion is a huge workload saver for subsequent warnings
    - The warning updates/follow-ups take much less time with TIM versus WarnGen.
    - At times, forecasters can update warnings more frequently using HS-TIM, which is especially important for high-impact events.
    - The warning updates are more precise.
    - One forecaster's testimonial, "handling these busy cases using WarnGen would have been pretty poor. We would be load shedding in order to keep up with the workload."
- "Persist" should be allowed for *any* long-tracked storm, not just those with observed significant hazards. Should be considered for Tiny TIM.
- Along with the many benefits of TIM from a statistical verification standpoint, the savings in forecaster workload make this a must-need for the NWS.

Future Plans

- Another virtual HWT experiment during Summer 2022.
    - Include new cases

  - The addition of a 3rd NWS forecaster, to act as the mesoanalyst and the conduit to end users.
- Continue software refinement and development of new functionality, including:
  - Hardening code (debugging, refactoring).
  - EDEX Recommender (opening up the possibility to use WES2 for displaced real-time scenarios).
  - Integration of HS-TIM into the operational version of Hazard Services, to take advantage of the many new functionalities in the newest versions of HS.
  - Windshield Wiper Effect (WWE) mitigation, such as a 5-10 minute "cool down" period added to locations removed from the warning.
  - Automated county clipping.
  - Advanced motion uncertainty features (e.g., splitting right and left direction uncertainty).
  - Separate front and back end of the warnings in motion.
  - Modify the Hazard Information Dialog to better match HS-Convective and Tiny TIM content.
  - Ability to merge, split, and copy threat objects.
  - WFO Collaboration domain permission to work with WFO localization.
  - Any of the items suggested by forecasters in the Appendices.

Web Presence (each subsequent item goes into more depth on TIM than the previous item)

- NSSL Bite-Sized Science 3-minute video on Threats-In-Motion
- A Blog summary about TIM (use NOAA credentials to log in)
- The TIM Weather and Forecasting journal article (Stumpf and Gerard, 2021).

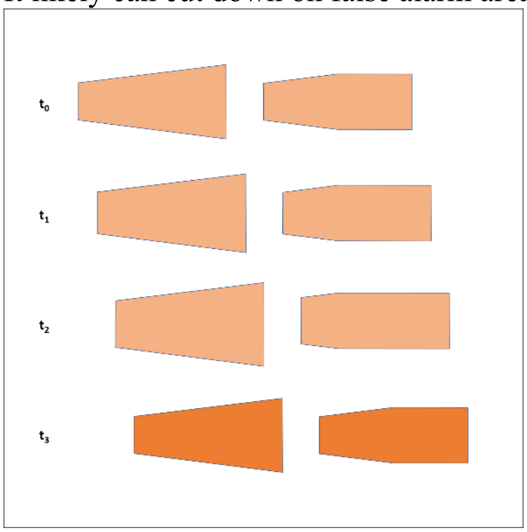
References

Stumpf, G. J., and A. E. Gerard, 2021: National Weather Service severe weather warnings as Threats-in-Motion (TIM). *Wea. Forecasting*, **36**, 627-643. https://doi.org/10.1175/WAF-D-20-0159.1

Software Functionality Suggestions

Warning Output:

- Add the option to have the uncertainty extend to only to 50% of the length of the track, and then go straight. Because the warning will likely be modified by the time we get 50% through the duration. Perhaps set the time at which the uncertainty stops expanding. It likely can cut down on false alarm area/time as well.

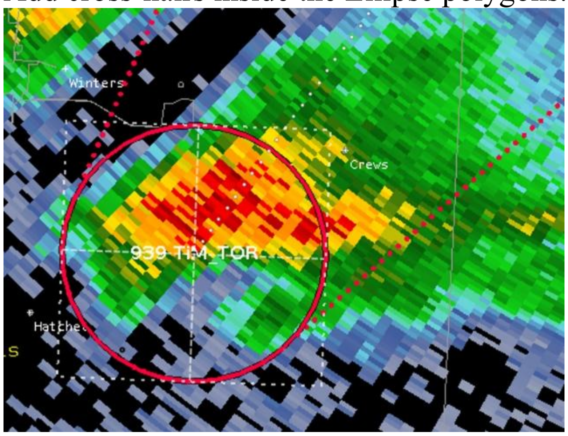


Object Editing:

- For ellipses, use the Ctrl key to make perfect circles. Also add the ability to do a single click for a 5 km wide circle centered on the mouse click.
- Change the workflow to be more like WarnGen. Draw the object, set the motion vector, select the hazard type. Or, to initiate a new object, select New SVR or New TOR, either from a button (see above), or a right-/long-/double- click on the spatial display. Less eye- and mouse-movement are best. Maybe 5 seconds of eye- or mouse-movement is saved, but saves some mental stress that stacks up over a 3-4-hour event.
- New ways to take action on objects to help with workflow:
    - Add Modify and End Object buttons in the object right-click menu.
    - Selecting an object via the spatial display should be with a double click (or the center or right mouse button) to prevent accidental selection when panning and zooming, or when toggling the legend.
- Add a way for forecasters to see the outline of the previous warning swath while they are modifying the current warning. So that they know what areas they are removing. This could be triggered on the "Preview Warning" button, and be shown as a different color.

Spatial Display (visual features):

- Bottom right rotation should be the two arrows around a circle (like PPT). Or, move the rotation to a short stem, not on the box (like Photoshop).
- Add cross-hairs inside the Ellipse polygons.



- Use thicker polygons to denote either 1) persisting warnings, or 2) significant warnings (based on IBW tag).
- New motion controls on swath decoration. Dragging the final point will change speed and direction. And ways to fan the swath wider to right and left to change direction uncertainty. These could be controls right on the swath, like the controls on the bounding box. All changes would reflect back on changes to initial conditions, and the upstream polygons and centroids would respond to these changes accordingly.

Console:

- Remove seconds on the console times to save space.
- The draw controls (which are only 5 buttons: radio:[Polygon, Freeform, Ellipse], radio:[New, Replace]) should be a separate breakaway in a narrow vertical stripe immediately next to the HID, instead of on the Console. That way, the Console could be broken off and put on another monitor as a SA display.
    - Alternatively, grey out Console components that are not used, or make these controls accessible by a right-/long-/double- click on the spatial display.
- Add “triangles” on the timeline to show the persisting portion too. Square would change to a triangle pointing right. If we freeze the back end, something that shows that too, like a back-facing triangle.

HID (General):

- On the HID Details box, have Modify, End Object, Persist, Preview, Reset, Latest TiM Frame, “pinned” into a separate Megawidget at the top (it would add more border boxes).
- Persist button should change to green when depressed. And change the label based on state (“Persist OFF” “Persist ON”).

HID (IBW tags and warning attributes):

- For SVR warning (3 rows, 2 columns [make sure the "Max" are vertically-aligned]):
    - Tornado
    - Hail Source, Max Hail
    - Wind Source, Max Wind
- For TOR warning (1 row; flip the current order):
    - Tornado Source, Tornado Severity

HID (Warning Decision Discussion):

- Add additional fields to WDD output:
    - WFO (so we know when object switched owners), storm motion values, Persist state, tag if geometry changed
- The initial issuance of a warning is the most time-sensitive and urgent, so we shouldn't require a WDD for a new warning. Instead, there should be a way to add a WDD at any time without updating the entire object.
    - This could be done by anyone in ops, including the mesoanalyst.
    - When the new warning is issued, unless the WDD is edited by the forecasters, it should include the phrase, "New warning - More details to come".
    - There should be a pop-up reminder if this isn't done within 5 minutes of new warning issuance.
- Add common WDD phrases (e.g., "no change in status") to a drop-down menu.
    - Include the option to repeat the previous discussion.
- Limit WDD to 280 characters, with a character countdown as you type.

Alerts and Pop-Ups:

- Need alert for expiring warning too like WarnGen. Options include, 1) pop-up, 2) blinking objects, 3) blinking rows in the console, 4) thick versus thin polygons.

Experiment logistics suggestions

Do an inventory of WarnGen functionality and layout to determine differences, and start converging.

Use a Slackbot for our WCM script. Could it be designed to use relative times instead of absolute times, since our reports are from the past?

Add background noise to scenarios: add other people talking not related to weather, phones ringing. Noise and ringing phones ramp up as storms move into the metro area.

Allow the forecasters to continue to “play in the sandbox” for about an additional hour or so after the Scenario 1 training event (keep the instances running, unsupervised).

Future TIM in-person experiment suggestion: when comparing WarnGen to HS-TIM, set up parameters of warnings for a controlled experiment: start at a specific time, have a standard duration (30 min), issue SVRs or updates at a specific time (every 10 minutes), direct them to which storm to warn on (they can decide no warn after analysis).

Procedures:

- Would have liked more time to set up procedures - perhaps on the training day, spend 30-60 min to create all procedures before starting the experiment.
- Or pick 1-2 core analysis procedures from each forecaster. We could also build a handful of pre-made procedures like the RAC procedures. We can alter them for our cases.
    - For example: clockwise from top left: Reflectivity At Lowest Altitude (RALA), Maximum Estimated Size of Hail (MESH) and MESH Tracks (60 min. accum.) image combination, Low-Level (0-2 km AGL) Rotation Tracks (60 min. accum.), and Vertically Integrated Ice (VII).

Should consider having some "pre-made" objects already available in all future scenarios.

Add a 4th DRT scenario to Friday morning (perhaps a back-building case), and have the afternoon for our 2-hour interview?

Human Factors Analysis of Survey Results

Participants in the HWT

Six NWS forecasters participated in the 2021 HS-TIM HWT, with two forecasters every week. The NWS work experience ranges from 2.75 years to 32 years, and the average was 9.8 years (standard deviation 11.83). The average NWS warning experience ranges from 2 years to 23 years, with an average of 7.4 years (standard deviation of 8.79).

Mental Workload (NASA TLX) Survey

The NASA-TLX (Hart & Staveland, 1988; Hart, 2006) workload index is a questionnaire based workload rating tool. The tool encompasses 6 aspects of workload: mental demand, physical demand, temporal demand, performance, effort and frustration. The raw scores of the mental workload ranges from 1 to 100, with 1 stands for extremely low workload and 100 stands for extremely high. The ratings were averaged from all the sessions for each of the 6 aspects of workload.

Table 1 shows the average ratings for the six sub-dimensions and the overall workload for three archived hazardous weather scenarios. The average workload for 2021 Hazard Services TIM HWT across all testing scenarios was 48.5 (out of 100, standard deviation 10.89). Each of these scenarios was chosen with an increasing level of difficulty (more storm coverage) throughout the course of the week.

Table 1. HS-TIM 2021 Testbed NASA-TLX Mental Workload Rating for Three Test Scenarios. Scenario 1 was a training scenario and was not included here.

	Scenario 2 Mean (std)	Scenario 3 Mean (std)	Scenario 4 Mean (std)
Mental Demand	64.3 (13.81)	63.0 (22.52)	60.5 (19.78)
Physical Demand	29.1 (18.07)	34.0 (18.71)	34.2 (20.25)
Temporal Demand	70.0 (11.40)	51.7 (25.92)	55.3 (25.50)
Effort	70.0 (12.25)	56.7 (23.89)	59.2 (21.45)
Frustration	45.2(22.09)	33.8 (9.34)	37.5 (27.02)
Performance	22.8 (12.02)	26.0 (12.94)	27.2 (13.49)
Overall Workload	50.19 (5.84)	44.11 (11.69)	45.6 (14.07)

PSSUQ Usability Questionnaire

The Post Study System Usability Questionnaire (PSSUQ; Lewis, 2002) is a survey tool designed to evaluate usability of a computer system. The tool is designed with 19 usability questions to assess 4 different areas of System Usefulness (Questions 1-8), Information Quality (Questions 9-15), Interface Quality (Questions 16-18) and Overall Usability (Questions 1-19). The rating ranges from 1 to 7, with 1 corresponds to low level of usability, 7 to high level of usability, and 4 corresponds to neutral level of usability.

The PSSUQ questionnaire was filled out by the participants after they completed all test scenarios in the testbed. Table 2 shows the average responses for each of the 4 categories: Overall Usability, System Usability, Information Quality and Interface Quality.

Table 2. Usability Ratings based on the PSSUQ for 2016 – 2021 HWT (7-point scale)

	2016	2017	2018	2019	2021HS-TIM Mean(std)
Overall Usability	4.62	5.39	4.95	4.92	**5.97 (0.39)**
System Usability	4.96	5.56	5.21	4.88	**6.25 (0.25)**
Information Quality	4.37	5.00	4.37	4.81	**5.71 (0.62)**
Interface Quality	4.72	5.72	5.50	5.18	**5.78 (0.91)**

The overall usability was assessed at 5.97 (on a 7-point scale) for the 2021 HS-TIM testbed, with system usability rated at 6.25, information quality at 5.71, and interface quality at 5.78. It is worth noting that the usability rating has improved from the past years. There are several factors accounting for this. First, the previous four years included a probabilistic component to the software, which increased workload. Second, the software has become more stable throughout the years. Finally, suggestions made by forecasters to improve the workflow have been implemented over the years.

Project Contacts

Greg Stumpf	CIRA and NWS/MDL	HWT Experiment Coordinator / Project Co-PI
Kevin Manross	CIRA and NOAA/GSL	Lead Developer / Project Co-PI
Alyssa Bates	CIWRO and NWS/WDTD	Project Co-PI
Chen Ling	U. Akron	Project Co-PI
Yujun Guo	CIRA and NOAA/GSL	Developer
Pat Hyland	CIWRO and NOAA/NSSL	Project Scientist
Zach Falkenberg	U. Akron	Project Scientist
Justin Monroe	CIWRO and NOAA/NSSL	AWIPS Support
Jonny Madden	CIWRO and NOAA/NSSL	AWIPS Support
Kodi Berry	NOAA/NSSL	NSSL FACETs Program Lead

Threats-in-Motion Tabletop Exercise with End Users

Summary by Kodi Berry

Overview

A new warning technology is currently under development called Threats-in-Motion (TIM), which essentially takes modern deterministic severe and tornado warnings and updates them more frequently, following storms in a more continuous way. The primary benefit of TIM is more equitable lead time downstream of an initial severe thunderstorm or tornado warning. This technology has been envisioned as a first transition step toward a full FACETs paradigm in future years, where more continuous forecast information is envisioned to also include probabilistic hazard guidance.

The technology received significant endorsement from user communities, including weather warning dissemination partners, private sector entities, broadcasters, emergency managers, and NWS forecasters during an introductory stakeholder engagement workshop in 2019. That meeting concluded with a key recommendation to bring stakeholders and NWS entities together again in the future to iron out specifics for how the technology could work, particularly in dissemination techniques and policies. This workshop was held in the spring of 2021, and several key themes emerged, including user preferences for the way TIM should function and how NWS could effectively support the transition.

To further investigate user preferences for TIM, five tabletop exercises were conducted with broadcast meteorologists and emergency managers in the fall of 2021. Participants worked through surveys in which they were shown different TIM prototypes, and were then asked for feedback. Focus group discussion followed the surveys. Discussion topics included initial impressions, preferred update frequency, and possible concerns. Most of our participants voiced a desire for a product like TIM, but expressed some concern over update frequency. While most of the broadcasters felt they could keep up with 1-minute updates in their on-air coverage, they might struggle to keep other platforms, such as crawls and social media, up to speed. Emergency managers liked the removal of areas on the back end of the TIM warnings but expressed concerns about how more rapid updates could impact siren operations. Both groups strongly favored some type of software intervention to avoid areas being taken in and out of the same warning polygon over relatively short periods of time.

Future Plans

Thematic analyses of the emergency manager focus groups, broadcast meteorologist focus groups, and large group discussions are currently underway. Results from these thematic analyses were presented at the American Meteorological Society's Sixth Conference on Weather Warnings and Communication in June 2022.

In the summer of 2023, we will field a public-facing experiment to test the effects of TIM errors, as presented in the end-user tabletop, on public decision-making. This effort will assess which improves trust and the quality of sheltering decisions more: smoothed forecasts (less up-to-date, but with less error) or more continually-updating forecasts that present different kinds of spatial errors.

KPHI blog https://inside.nssl.noaa.gov/kphi/

Project Contacts

Holly Obermeier	CIWRO/NSSL	Broadcast Lead
Kim Klockow-McClain	CIWRO/NSSL	Emergency Management Lead
Kodi Berry	NSSL	Broadcast Researcher
Taylor DeWinter	CIWRO/NSSL	Emergency Management Researcher
Joseph Trujillo	CIWRO/NSSL	Graduate Research Assistant

Brief Vulnerability Overview Tool Experiment

Summary by Jack R. Friedman, Michelle E. Saunders, and Daphne S. LaDue

Figure 1. BVOT research team convened to conduct a virtual HWT experiment.

Overview

This HWT project applies and integrates relevant social and behavioral science methodologies to assesses WFO forecasters' and end-users' abilities to assess, understand, and respond effectively to 1) experimental, short-term (through Day 1) forecasts for convective weather hazards and 2) a tool that will enhance their awareness of vulnerabilities within their County Warning Area (CWA). Specifically, we will assess how 1) new Storm Prediction Center's ability to *identify enhanced threat corridors* within Day 1 Probabilistic Convective Outlooks and *communicate those in event-driven Outlook updates in temporally disaggregated Day 1 Outlooks* and 2) Friedman/LaDue's *Brief Vulnerability Overview Tool (BVOT) will impact WFO forecaster product issuance and messaging* with deep core partners (e.g., Emergency Managers) as well as how EMs interpret these products. This project is testing the impact of increased vulnerability knowledge and awareness of CWA-based vulnerabilities on how 1) WFO products are issued and 2) EMs and other deep core partners interact with WFOs and (independently) interpret new temporally fine-grained SPC Outlooks in a simulated operational environment. This project will assess impacts on WFO forecaster and EM behavior by simulating end-to-end severe weather communication — SPC to WFO to EMs — through realistic experimental scenarios involving SPC and WFO forecasters and EMs.

We hypothesize: 1) that there will be a measurable difference in the content, timing, and frequency of messaging to deep core partners (e.g., EMs) due to consideration of spatially explicit vulnerability data (via the BVOT); 2) that there will be a difference in the content of forecasts, messaging to the public, and formal briefings provided to deep core partners associated with including new Storm Prediction Center Severe Timing Guidance products; 3) that the impact of both tools being used simultaneously will have an additive effect on changes in messaging and formal product issuance; and 4) that these changes in messaging and briefings will be positively perceived by Emergency Managers and will result in improved decision making.

The SPC Severe Timing Guidance Product

Our HWT experiment included testing the Storm Prediction Center's (SPC) new prototype graphics to provide automated timing information. These graphics were created using the HREF and SREF models and provide hourly updated guidance using a rolling 4-hour window of time. The time frame of these graphics is valid for the traditional SPC Day 1 severe weather outlook. We tested seven graphics in total, five static images and two animations (Fig. 2 & Fig. 3). The SPCs motivation is to create a nationally consistent database for severe weather timing information that can be used both internally at the WFO and to aid in messaging their core partners. These timing graphics begin to help communicate the timing of severe weather threats while maintaining consistency across spatial boundaries.

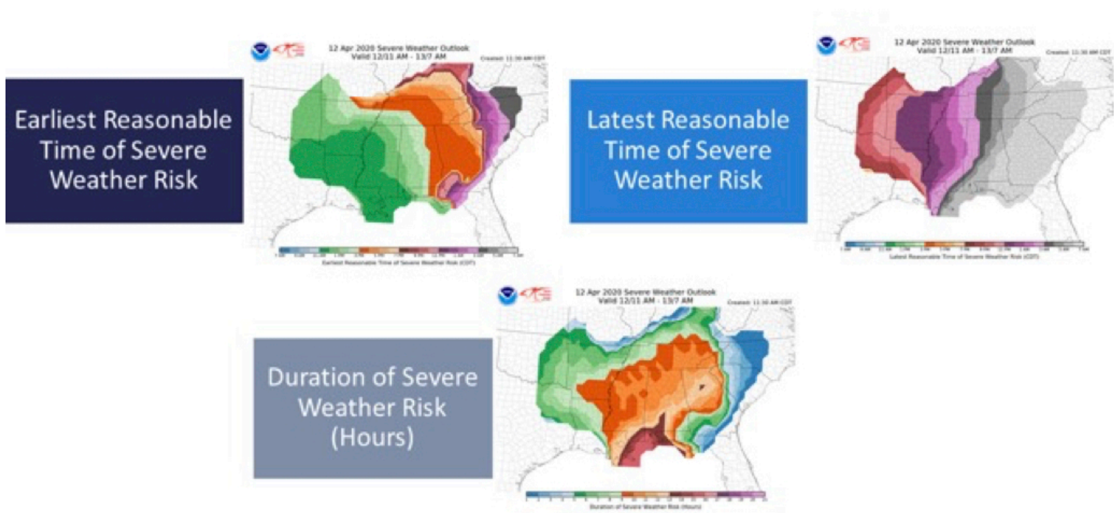


Figure 2. Three of seven experimental SPC timing graphics tested in our HWT experiment.

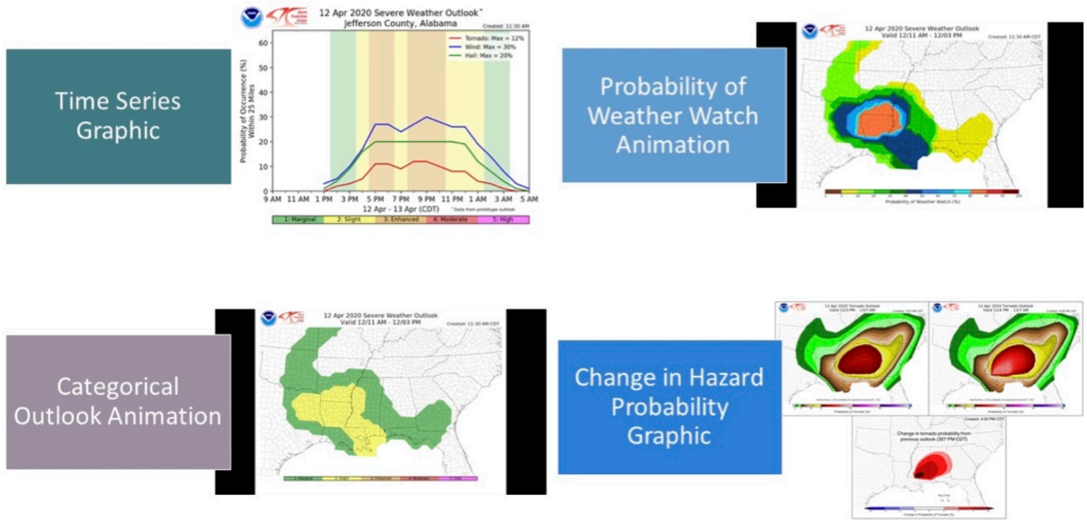


Figure 3. The remaining four of seven experimental SPC timing graphics tested in our HWT experiment.

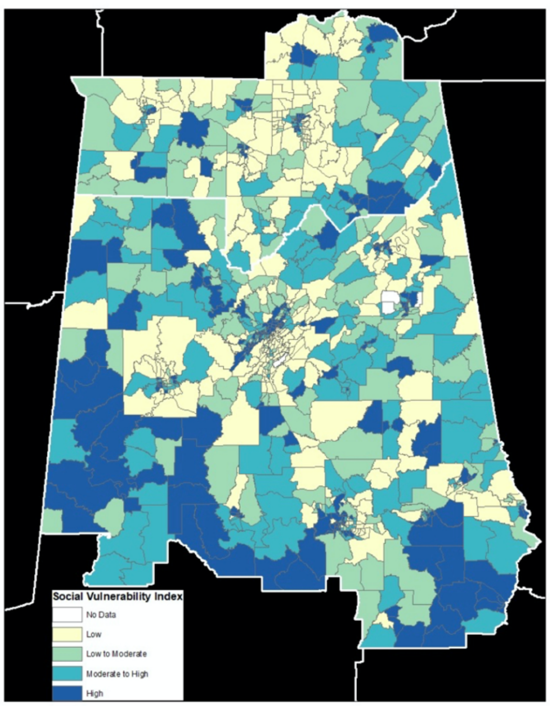
Our experimental design uses several methods and data collection techniques including participant observation, post case debriefings, and end of day and week whole group discussions. We utilize two surveys, the System Usability Scale (SUS) survey to measure how the usability of the graphics changes from the beginning of the experiment week to the end of the week. Finally, we use a post-experiment survey to gather any final information forecaster participants want to share about each of the seven graphics.

The Brief Vulnerability Overview Tool (BVOT)

The Brief Vulnerability Overview Tool (BVOT) is a GIS-based, graphical map overlay that can display specific, place-based, known vulnerabilities across a National Weather Service (NWS) Weather Forecasting Office's (WFO) County Warning Areas (CWA). The goal of the development of the BVOT is that it will provide additional spatial situational awareness to NWS WFO meteorologists by allowing them to quickly assess whether a weather hazard will be directly threatening a specific vulnerability. This will provide NWS meteorologists with a visual display that will allow them to quickly assess whether they will need to provide enhanced or specifically-tailored messaging to their partners (e.g., emergency managers, law enforcement, etc.).

The NWS has supported research to create a more generalizable and standardizable way of operationalizing vulnerability information for NWS meteorologists (Fig. 4). Many of these efforts focus on converting vulnerability data that was originally collected for non-meteorological purposes — various Social Vulnerability Indexes — into a format that can be used to inform NWS meteorologists' IDSS efforts. Many of these efforts take existing databases like the CDC's SVI and develop ways of mapping and visualizing census-derived vulnerability data — like households in poverty, population without a high school education, or percentage of residences that are mobile or manufactured housing — in order to provide spatial awareness of these vulnerabilities to operational meteorologists. However, because of the aggregated nature of these vulnerability data — aggregated spatially across census tracts and thematically by bringing together multiple vulnerabilities to develop social vulnerability "themes" — it is more difficult to translate these hypothesis-derived vulnerability data (HDVD) into actionable decision support for WFO-level, operational meteorologists who need to communicate Decision Support Service-oriented messaging to their core partners.

CDC Social Vulnerability Index (SVI) map for the northern 2/3 of Alabama and three counties in Tennessee (the CWA for NWS HUN and BMX). This is an example of vulnerability mapping drawing on Hypothesis-Derived Vulnerability Data (HDVD).



The Brief Vulnerability Overview Tool (BVOT) for the same area. This is an example of vulnerability mapping drawing on Knowledge-Derived Vulnerability Data (KDVD).

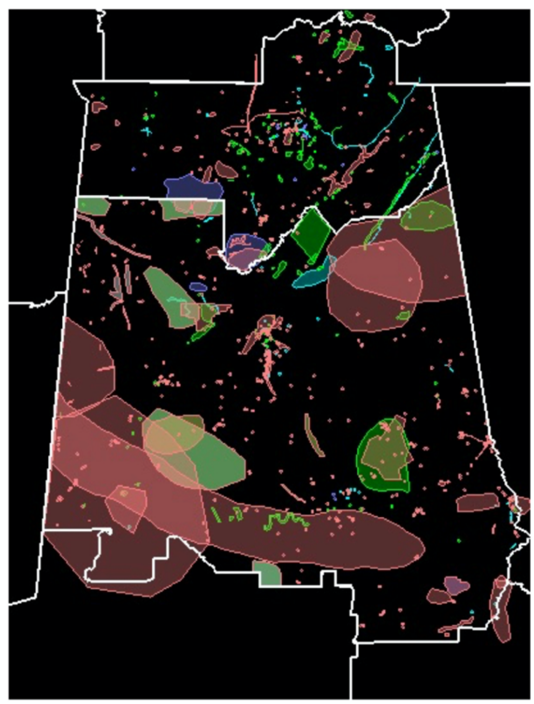


Figure 4. Comparison of the types of visualizations of different types of vulnerability data, in this case, comparing the CDC SVI with the Brief Vulnerability Overview Tool (BVOT).

The Brief Vulnerability Overview Tool, or BVOT, was developed as a way of balancing the strengths of the HDVD-approaches (like those that derive vulnerabilities from the CDC's SVI) with the specificity of something like the NWS's Impacts Catalog, while addressing weaknesses that have been identified by NWS meteorologists. The BVOT is composed of discrete, knowledge-derived vulnerability data (KDVD). This can be contrasted with general, hypothesis-derived vulnerability data (HDVD) that, as noted above, aggregates census tract-level data. In other words, what distinguishes KDVD from HDVD is that KDVD is more spatially fine-grained and the vulnerabilities that have been identified are known by experts to be at risk of being impacted by specific weather hazards.

In our HWT experiments, we have sought to understand how, when, and if added vulnerability awareness changes the 1) decision-making, 2) messaging, and/or 3) product issuance of NWS WFO-level meteorologists. We provide details on the experiment and results from 2021's experiments in the next section.

Experiment Details and Results

The experiment involves the participation of both NWS WFO-level meteorologists and emergency managers from a range of different backgrounds (e.g., county, municipal, higher education, military/federal, etc.). Each week of the experiment (3 in 2021, 5 planned for 2022) involves 6 NWS meteorologists (18 total in 2021) and 6-12 EMs recruited (30 total in 2021) from around the country. Each week is divided into 3 sections: 1) the first half of Day 1 involves orientation to the project, the SPC tool, and the BVOT; 2) the second half of Day 1 through the first half of Day 5 involves participants in decisions making, messaging, and product issuance based on 8 recorded (WES) severe weather cases (more details below); and 3) the second half of Day 5 is used to conduct focus groups to collect data on the overall impressions, usability, value, and concerns about the SPC and BVOT tools.

Experimentally, the study is designed around the NWS meteorologists who are divided into 3 teams of two meteorologists, with each team rotating through different experimental conditions through the week (Fig. 5):

- Condition A: BOTH the SPC Experimental Severe Timing Graphics AND the BVOT
- Condition B: Only the BVOT
- Condition C: Only the SPC Experimental Severe Timing Graphics

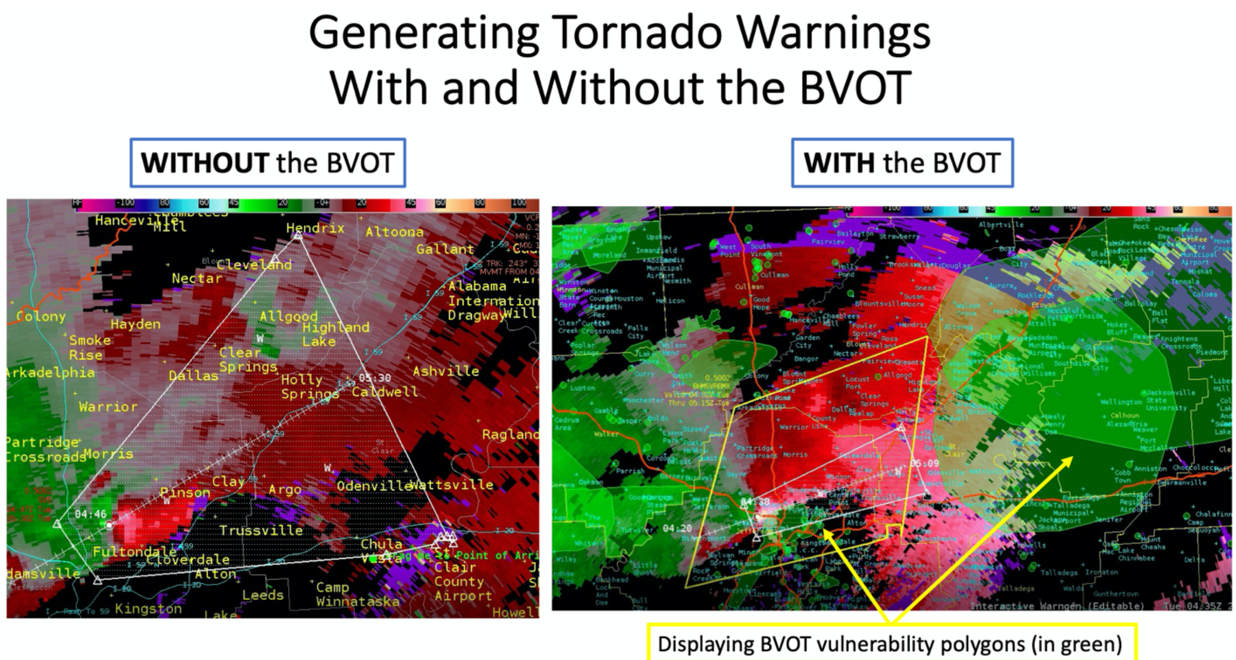


Figure 5. Example screenshots of the difference between a participants' AWIPS screens during the storm-on-the-ground period without (left) and with (right) the BVOT display.

Because we are assessing when these tools might be used or of value to NWS meteorologists in providing added or more focused messaging to their partners, each of the teams of meteorologists is tasked with making decisions based on different time periods in the evolution of a severe weather threat. These time periods are:

- 24-48 hours before the event
- 4-12 hours before the event
- Storm-on-the-ground (a 45-minute period during which active severe weather is on-the-ground, requiring critical engagement with AWIPS-in-the-cloud instances that are hosted on Amazon Web Services (AWS) and managed through the HWT)

Specifically, in each of the three time periods during each of the 8 cases, each team is given 45 minutes to review briefing material and then provide briefing packets, provide other forms of messaging to partners/publics (the project uses Slack to simulate both NWSChat and social media issuances), and, when relevant, issue formal NWS warning products. At the start of each period, teams are provided a briefing packet that contains SPC outlooks, SPC discussions, as well as the local NWS WFO's area forecast discussions (AFDs), and, during "storm-on-the-ground" periods, a radar loop showing the evolution of the event in order to provide the NWS meteorologists (and the EMs) realistic situational awareness on each event as it unfolds. While all EMs receive the most detailed briefing packets for each of the events, those NWS teams that are in Condition B will NOT receive the SPC Experimental Severe Timing Graphics, while those NWS teams that are in Condition C will NOT be able to access the BVOT vulnerability map layers. In this way, we are able to record any differences that we are seeing in the decision making, messaging, and/or product issuance correlated with different Conditions/access to these experimental tools.

Data collection includes:

- Recordings of all meteorologist decision-making through interactions (via recorded Google Meets) between meteorologists
- Recording all AWIPS instances (via screencapture) for each meteorologist throughout all case during all periods
- Recording all EM discussions and decision-making (via Google Meets) at each stage of the experiment
- Recording all interaction between meteorologists and EMs (including the briefing material/packets provided, any live webinar-style briefings (via Google Meets), and any text-based messaging or interaction (via Slack, representing both NWSChat and social media)
- The Post Study System Usability Questionnaire (start and end of the week)
- The NASA Task Load Index Survey (after each of the 8 cases)
- The Confidence Continuum Survey (after each of the 8 cases)
- The Secondary Traumatic Stress Survey (modified) (start and end of the week)
- Debriefing interviews with each team of meteorologists after each case period (24 total)
- Post-case debriefings at the end of each day involving both the meteorologists and the EMs
- Post-study survey gathering additional information about the usability and preferences regarding the SPC Severe Timing Guidance Products (for the NWS meteorologists)

SPC Severe Timing Graphics — Early Results

After preliminary analyses, we found that the time series graphics (see Fig. 6) are valued the most out of the seven graphics. The time series graphics display the probability of occurrence (%) within 25 miles for tornadoes, wind, and hail over time, along with the categorical outlook. Participants stated that they would feel comfortable sharing the time series graphics with core partners (Fig. 7).

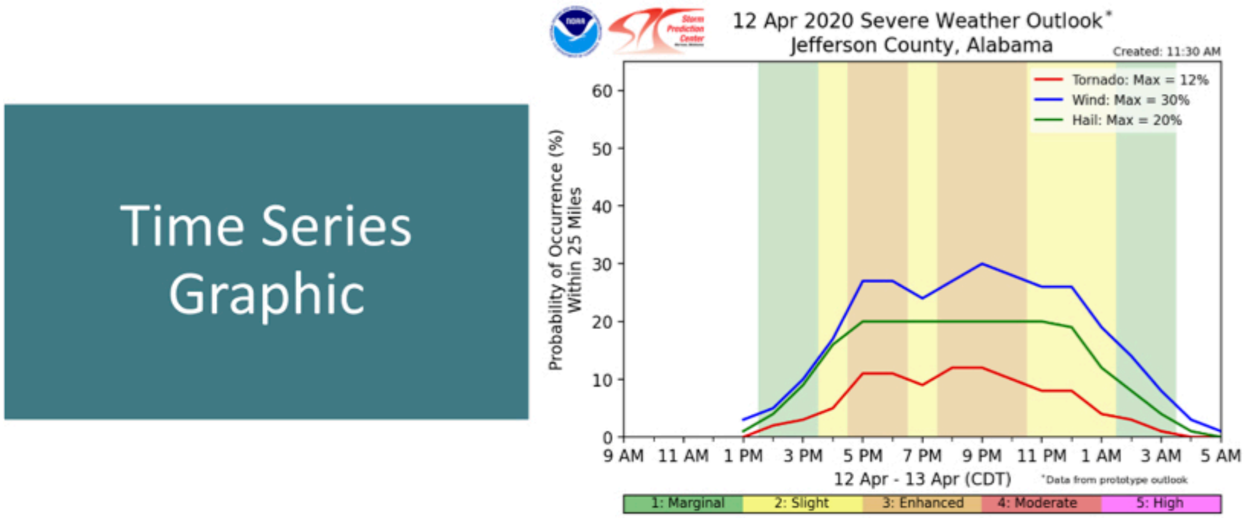


Figure 6. Example of a point-based time series graphic from the SPC's Severe Timing Graphic tool.

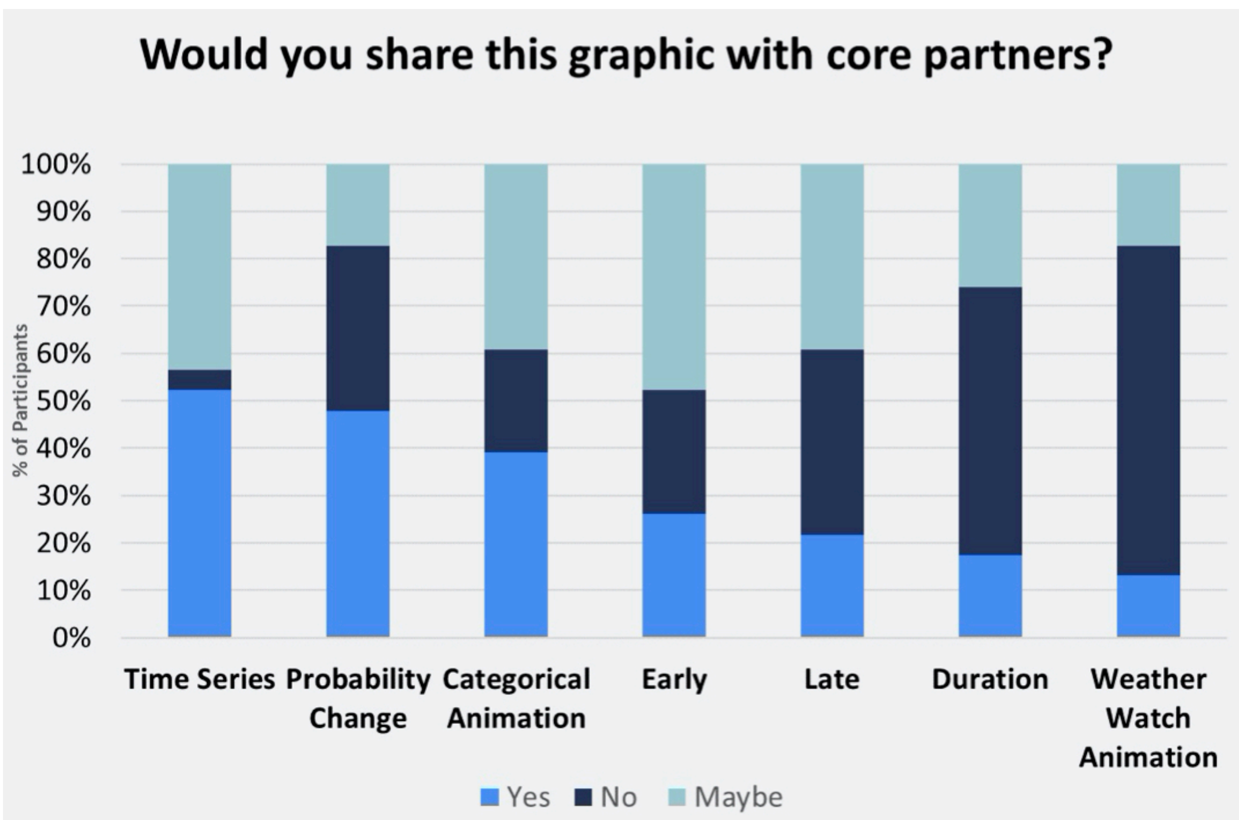


Figure 7. Post-experiment survey data from forecaster participants on whether they would share each experimental graphic with their core partners.

The time series graphics were also shared the most with EMs during our experimental cases. The earliest reasonable timing of severe weather risk graphic was the next most valued graphic by both forecaster and EM participants. Forecasters liked the graphical representation of severe weather timing and EMs stated that this graphic could be important for scheduling staffing and preparing internally before a severe weather event. So far, most forecaster and EM participants do not find the probability of a weather watch graphic (Fig. 3) to be as useful, especially compared to the other timing graphics being tested. One forecaster concluded their thoughts on this graphic, stating that, “It’s good information to have, but we can draw conclusions based on everything else.”

Participants described wanting to be able to customize the SPC timing graphics before sharing them with their partners and provided examples for how they would do this. Each group of forecasters have expressed interest in wanting to see the timing guidance for the maximum risk category for a weather event. At present, several graphics were designed to show the timing for the slight risk category.

Finally, the SUS scale (Fig. 8) indicated that there was an increase in inconsistency when using the timing graphics from the beginning to the end of the week. This inconsistency is most likely due to the peak categorical risk not appearing in some of the timing graphics, which was highlighted by participants during debriefings, group discussions, and in the post-experiment survey.

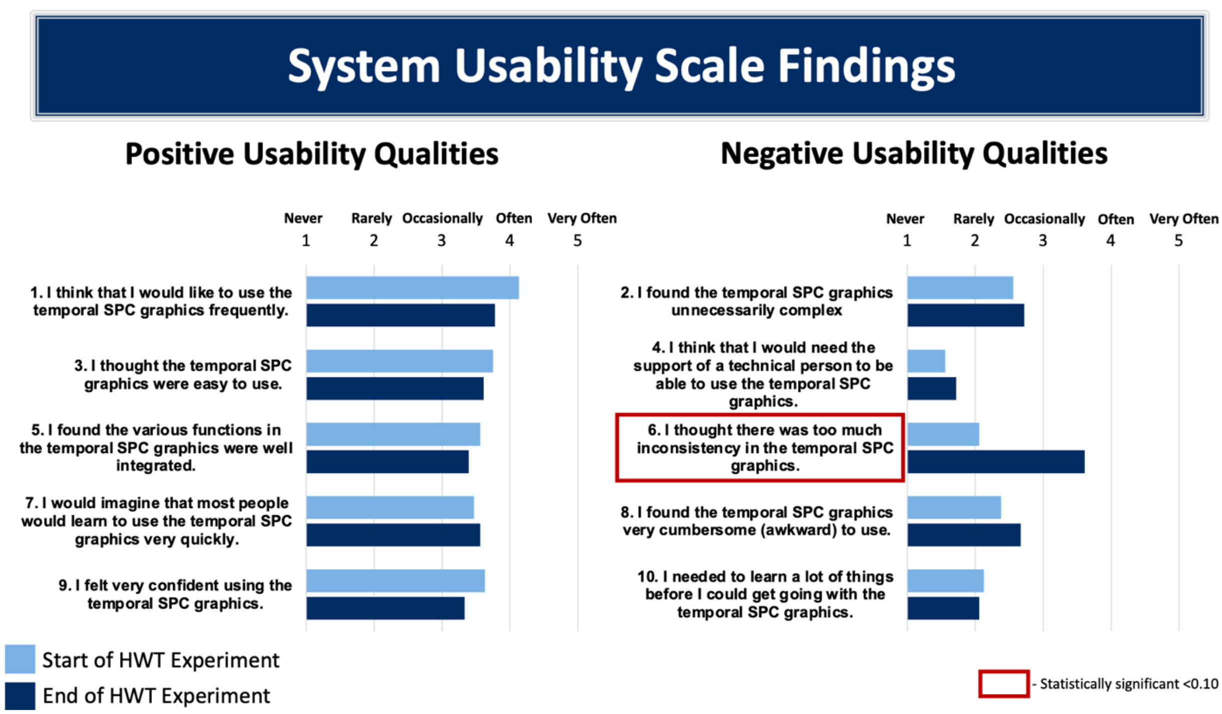


Figure 8. Summary of positive and negative usability qualities reported by NWS meteorologists regarding SPC Severe Timing Graphics.

Brief Vulnerability Overview Tool (BVOT) — Early Results

We have been evaluating (ongoing) the first three weeks of HWT experiments, focusing on four critical questions regarding the impact of increased awareness of vulnerabilities on NWS meteorologist operations:

- What impact does the BVOT have on the nature of messaging to/with core partners?
- What impact does the BVOT have on NWS meteorologists' understandings of the impacts of a severe weather event?
- What impact does the BVOT have on the practices that occur before, during, and after severe weather operations?
- What impact does the BVOT have on the issuance of formal NWS products (watches, warnings, etc.)?

We are in the middle of analyzing these data, so, while we have significant anecdotal evidence to begin to answer some of these questions, we will focus on discussing the last of these questions — What impact does the BVOT have on the issuance of formal NWS products (watches, warnings, etc.)? — since answering this question is critical for a number of reasons. First, we did not design the BVOT to strongly impact the decision to issue or not to issue a formal NWS product. Rather, the BVOT is designed to provide added awareness regarding the potential impacts of weather hazards on discrete, vulnerable people, places, or things. As such, we did not expect to see a significant impact on the decision to or the timing of formal product issuance. In general, this is what we have found, so far, in this experiment (Fig. 9). For instance, with our Case 4, a recording of the data from a 13-14 April 2019, overnight severe weather event that impacted NWS HUN, and spawned multiple verified EF-0 and EF-1 tornadoes, one can see that all of the NWS meteorologist teams, regardless of experimental condition, issued tornado warnings at or near the exact same times.

Timing of Issuance of Tornado Warnings: Example from Experimental Case 4, 13-14 April 2019, HUN WFO, Multiple EF-0 and EF-1 Verified

	Condition A (w/BVOT & SPC Timing Guidance)	Condition B (w/BVOT only)	Condition C (w/SPC Timing Only, NO BVOT DATA)
Week 1	11:58	11:57	11:58 12:20 (reissue)
Week 2	11:56 12:21 (reissue)	11:56	11:58
Week 3	11:53 12:10 (reissue)	11:54 12:22 (reissue)	11:58 12:01 (reissue)

Figure 9. Timing of all tornado warnings issued for Case 4 in the HWT experiment. All warnings were issued within 4-5 minutes of each other across the first three weeks of the experiment.

While the decision to issue/not to issue a formal warning on a storm was not impacted by increased vulnerability awareness (i.e., those NWS meteorologists under Condition A or Condition B, who had access to the BVOT), we *did* find that increased awareness of vulnerabilities could impact the *tags* that were associated with tornado warnings. We found that, in several of the severe weather cases, the added awareness of specific vulnerabilities *combined with* the radar-indicated severity of the storm (e.g., a clear and significant TDS) prompted meteorologists in Condition A and Condition (those with access to the BVOT) to issue Severe Weather Statements along with their tornado warnings that indicated the Particularly Dangerous Storm (PDS) tag. Event Case 8 was a particularly telling example (Fig. 10) because it was a sudden and quick-spin-up tornado that was in an area that had few AWIPS base-map locations but had several BVOT-indicated vulnerabilities. In this case, while all of the meteorologists all of the weeks (regardless of experimental Condition) issued tornado warnings, *only* those meteorologists with access to the BVOT added the PDS tag, recognizing the potential for significant risk to life and property due to their awareness of on-the-ground vulnerabilities.

Timing of Issuance of Tornado Warnings:
Example from Experimental Case 8, 25 January 2021 (Fultondale, AL), HUN BMX, EF-3 Verified

	Condition A (w/BVOT & SPC Timing Guidance)	Condition B (w/BVOT only)	Condition C (w/SPC Timing Only, NO BVOT DATA)
Week 1	10:32 Svr TS 10:40 TOR 10:46 SWS (PDS) 10:55 SWS (PDS)	10:33 Svr TS 10:36 TOR 10:40 SWS 10:46 SWS (PDS) 10:59 TOR (reissue)	10:39 TOR 10:47 SWS
Week 2	10:32 TOR 10:38 SWS (PDS) 10:46 SWS (PDS) 10:52 SWS (PDS)	10:37 Svr TS 10:39 TOR 10:44 SWS 10:51 SWS (PDS)	10:36 Svr TS 10:43 TOR 10:49 SWS
Week 3	10:33 TOR 10:38 TOR 10:50 SWS (PDS) 10:52 SWS	10:32 SWS (Svr TS) 10:37 TOR 10:42 SWS (BVOT-points) 10:47 SWS (PDS)	10:33 TOR 10:37 SWS 10:43 SWS 10:45 TOR (reissue) 10:47 SWS 10:51 SWS 10:52 SWS

Figure 9. Showing the timing of a number of products issued for experimental Case 8 during the first three weeks of the HWT study. While all teams of NWS meteorologists — regardless of experimental Conditions — issued tornado warnings, *only* the teams with access to the BVOT issued PDS Severe Weather Statements (SWS). In addition, in one case, the SWS contained a specific reference to a BVOT-only point, meant to convey additional impact-based messaging.

Future Plans

The PIs expect to complete the final 5 weeks of HWT-based experiments during 2022 (the next three have been scheduled in January, February, and March of 2022; the final 2 we hope to schedule early in the Fall of 2022).

Project Contacts

Jack R. Friedman	Univ. of Oklahoma, CASR	Project PI
Daphne LaDue	Univ. of Oklahoma, CAPS	Project Co-PI
Michelle Saunders	Univ. of Oklahoma, CASR	Postdoctoral Research Scientist
Alex Marmo	Univ. of Oklahoma, CAPS	Research Associate
Justin Monroe	CIWRO and NOAA/NSSL	EWP Technical Lead

4. PERSONNEL

Alan Gerard
Acting HWT Co-Executive Officer alan.e.gerard@noaa.gov

Kodi Berry
Acting HWT Co-Executive Officer kodi.berry@noaa.gov

Kristin Calhoun
Acting HWT Co-Executive Officer kristin.calhoun@noaa.gov

Justin Monroe
EWP Technical Lead justin.monroe@noaa.gov

Jonathan Madden
EWP Technical Advisor jonathan.madden@noaa.gov

5. ACKNOWLEDGMENTS

We would like to acknowledge the contributions of the following individuals: Karen Cooper, Michael Taylor, Vicki Farmer, and Alan Gerard.

HWT Infrastructure Support was provided, in part, by the U.S. Weather Research Program and Joint Technology Transfer Initiative. Research was funded by the U.S. Weather Research Program, Joint Technology Transfer Initiative, National Severe Storms Laboratory, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreements #NA16OAR4320115 and #NA21OAR4320204, U.S. Department of Commerce.