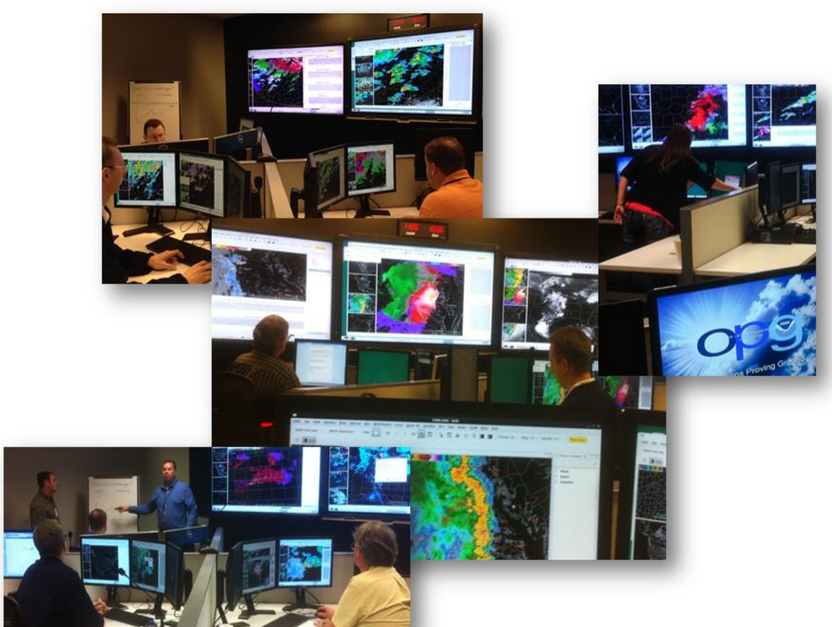
NWS Operations Proving Ground

Operational Readiness Evaluation Report

NASA SPoRT-MDL Tracking Meteogram



NWS Operations Proving Ground
7220 NW 101st Terrace
Kansas City, MO 64153

Executive Summary

The National Weather Service (NWS) Operations Proving Ground (OPG) conducted a formal Operational Readiness Evaluation (ORE) – the first of its kind – in May 2014. Evaluation sessions focused on a diagnostic tool known as the Tracking Meteogram (TM), developed for the NWS' Advanced Weather Information Processing System, Version 2 (AWIPS-2) by the NASA Short-Term Prediction and Transition (SPoRT) Center in collaboration with the NWS Meteorological Development Laboratory (MDL). The aim of this ORE was to assess the tool's practical usability, usefulness for decision making, and potential for workload impacts, from the perspective of NWS forecasters. Forecasters from four NWS regions were placed in a simulated WFO environment, and then led through a diverse array of warning and forecast scenarios, using both archived cases and live data sets. This report summarizes the TM's purpose and path to formal usability testing, describes the assessment process and participants, and reveals assessment results, including opinions regarding readiness for implementation.

Overall, NWS forecasters unanimously endorsed the TM as a useful tool that could add genuine, unique value to several practical forecaster tasks. A few key effective practices, along with specific findings and recommendations from the evaluation, are identified below.

A. Effective Practices

Effective Practice 1: The TM is very useful as a forecaster decision aid for tracking and displaying trends with meteorological features (e.g., mesocyclone, reflectivity core, cloud-top temperatures, and total lightning data).

Effective Practice 2: The TM is useful at interrogating gridded model output.

Effective Practice 3: The TM is useful for mesoanalysis applications such as monitoring trends in instability, convective inhibition, shear, and rainfall rates.

Effective Practice 4: The TM is useful at providing the forecaster an efficient way to extract and communicate meteorological information that is critical to NWS' core partners.

Effective Practice 5: The TM would be useful in post event analysis to evaluate how trends in data were related to warning decisions. (to improve warning process)

B. Findings and Recommendations

Finding 1: Forecasters found it difficult to assimilate and interpret the TM data when plotting four or more meteograms concurrently. Forecasters also observed that the depiction of numerous meteograms impacted system performance.

Recommendation 1: Limit the number of individual TMs displayed to minimize adverse impacts on the forecaster's interpretive capability and the system's processing efficiency.

Finding 2: To utilize the TM to monitor in-situ changes over a fixed location, it was necessary for the user to manually stack the TM's trace circles.

Recommendation 2: Create a TM stationary mode as a menu choice prior to AWIPS-2 field implementation.

Finding 3: Forecasters discovered that it is not possible to store a loaded TM location into an AWIPS-2 procedure.

Recommendation 3: Add the capability to store a stationary mode TM into an AWIPS-2 procedure prior to AWIPS-2 field implementation.

Finding 4: Functionality does not exist to set and plot a user-defined threshold value superimposed on the TM graph. Forecasters identified two specific applications of value for this feature: (1) as a visual reference on the trace to aid the forecaster's analysis and forecasting decision process, and (2) as an optional trigger point to activate a notification alert message when exceeded.

Recommendation 4: To optimize DSS effectiveness, developers should consider adding this user-defined threshold functionality to an upcoming TM version.

C. Lessons Learned about the ORE Process

Many topics and observations emerged from written surveys and oral discussions, but the most significant positive take-aways concerning the ORE process fell into four broad categories:

- Integration of the VLab Development Environment in the ORE Process
- Diversity and Realism of Cases Used for Evaluation Sessions
- Emphasis on Human Factors in Evaluating Usability and Usefulness
- Blend of Forecasters, Trainers, Developers, and Support Staff Who Participated

---

D. Recommendation for Field Implementation

All participating NWS forecasters unanimously endorsed the TM as a useful tool that will add unique value to several specific diagnostic and predictive tasks without posing adverse impacts on operational workflow or forecaster workload. While NWS forecasters believe it will be received well and readily adopted by forecasters in its current version, they recommend two enhancements prior to widespread field implementation. First, it is recommended that the stationary mode option, AWIPS-2 procedure capability, and tendency information be added to the TM's functionality. Second, a brief TM training video and one-page (front and back) informational handout should accompany the TM's implementation.

Based on the results of this ORE, the OPG recommends implementation of the Tracking Meteogram into the AWIPS-2 baseline without reservation.

I. Introduction

During the week of 12-16 May 2014, the National Weather Service (NWS) Operations Proving Ground (OPG) held its first formal Operational Readiness Evaluation (ORE). The ORE was completed on the Tracking Meteogram (TM), a tool developed for the NWS' Advanced Weather Information Processing System, Version 2 (AWIPS-2) by the NASA Short-Term Prediction and Transition (SPoRT) Center and the NWS Meteorological Development Laboratory (MDL). This report will summarize the TM's purpose, path to formal usability testing, the assessment process and participants, and overall findings and recommendations regarding readiness for implementation.

A. Background Information on Tracking Meteogram

The TM capability was developed within the NWS Operations and Services Improvement Process (OSIP) as Project 06-407. That project began in 2006 with a focus on observing trends in total lightning data. As the project matured and evolved, developers shifted their attention from a tool designed specifically for diagnosing total lightning to one that had broader application within AWIPS-2. Project 06-047 was approved through OSIP Gate 3 and then redirected to the Software Recommendation and Evaluation Committee (SREC) for prioritization.

Initial operational demonstrations of the TM centered on software maturity, performance characteristics, scientific integrity, and general value for impacting forecast decisions. These demonstrations were conducted in two settings. First, within a limited number of NWS Weather Forecast Offices (WFOs), chosen because of their proximity to a ground-based lightning mapping array (LMA). These WFOs installed the first version of the TM for use in real time applications (Bridenstine et al. 2005; Demetriades et al. 2008; Nadler et al. 2009; Darden et al. 2010; White et al. 2012). Second, the TM was one of several diagnostic applications at the Hazardous Weather Testbed (HWT) Spring Experiment in 2013 and 2014.

The focus on total lightning data initiated from two key considerations. First, the lightning data currently available to operational forecasters is limited to detection and depiction of cloud-to-ground strikes. This only comprises approximately 10% of a typical thunderstorm's electrical charge output. Access to, and visualization of, the entire spectrum of a lightning stroke will better prepare forecasters for data sets which will become available with the launch of GOES-R. Second, early operational tests suggested that certain trends in lightning activity, particularly the so-called "lightning jump" may provide several minutes lead time on thunderstorm intensification (Schultz et al. 2009; Gatlin and Goodman 2010; Schultz et al. 2011). These predictive indications are directly correlated to a thunderstorm's updraft strength in the mixed phase region of the cloud, and the attendant charge generation and charge separation.

Researchers, developers, and forecasters hypothesized that the physical relationship between updraft strength and lightning production might be exploited in the operational environment, especially given the fact that temporal updates of ground-based LMAs (and with the future Geostationary Lightning Mapper) are more frequent than those available via radar volume scans. While this reasoning was logical, and supported by a few anecdotal observations, a formal evaluation of the concept was needed to validate the theory. This would require verification that the TM adds genuine value to the decision-making process without having an adverse impact on workflow, data assimilation, and other human factors.

Accordingly, in response to the Announcement of Opportunity issued by the OPG through the NOAA Testbeds and Proving Grounds Coordinating Committee, NASA SPoRT and MDL submitted a proposal to participate in an ORE for the TM. The ORE would consist of a series of warning and forecast decision-making scenarios executed by NWS forecasters in a realistic operational setting.

B. Pathway to OPG

The OPG accepted the NASA SPoRT and MDL proposal in April 2013. However, the OPG staff was unable to complete its system installation and configuration in time for a 2013 session, due to severe budgetary restrictions that culminated in a government shutdown. This turned out to be fortuitous since it gave TM developers more time to refine the software and test its maturity in the context of the NOAA testbed experiment process.

TM evaluations from the HWT Spring Experiments provided valuable feedback, both in terms of identifying needed improvements to the TM's performance characteristics and in suggesting additional practical uses for the tool. Specifically, forecasters suggested expanding the fundamental functionality to include any gridded data in AWIPS-2. Forecasters felt the TM may gain broader acceptance if it was advertised as an interrogation tool capable of creating a time series trend plot of any gridded field in the data base. Thereafter, NASA SPoRT and MDL began collaborating on this expanded function. Once these enhancements were coded and tested, a new ORE was scheduled with the OPG staff. These code revisions, written to accommodate 64-bit capabilities of the AWIPS-2 Build 14.1 operating system, were provided to the OPG along with the NASA SPoRT visualization plug-in.

Forecasters participating at the HWT also felt there were a variety of limitations with the TM. Many of these limitations were features available within the TM tool menu but unknown to the participating forecasters. Developers concluded these misperceptions could be addressed with a focused up-front training session at the start of the ORE. It was at this time that SPoRT, MDL, and the OPG staff decided to include WDTB training staff responsible for developing the TM training for AWIPS-2.

Currently, the weather event simulator (WES) does not exist for AWIPS-2. If in place, this functionality would have allowed the OPG to develop high-quality simulations for each ORE scenario. NASA SPoRT programmer Jason Burks resolved this issue by writing a Python script to ingest and display archived radar, satellite, lightning data, and surface observations in displaced real time. While there were inherent limitations to this method (detailed in Part III, Section 2), Mr. Burks' solution was an adequate and effective means to accomplish the goals of the ORE. Jason deserves special recognition for this innovative work-around. Simply put, completing the ORE would not have been possible otherwise.

II. Evaluation Overview

A. Participants

ORE participants included four NWS forecasters, one from each of the four NWS Regions within the Continental United States (CONUS); two programmers/developers (one from NASA SPoRT and one from MDL); a subject matter expert (SME) from NASA SPoRT; a Training

Designer/Instructor from the NWS Warning Decision Training Branch (WDTB); and four OPG staff members.

Considerable thought was given to the number and background of the NWS forecasters. First, ORE facilitators decided it was essential to benefit from the knowledge and perspective of forecasters from a variety of geographic locations. It was felt this would minimize the chance that opinions would be biased toward the forecast challenges of a particular climate regime or subregion of the CONUS. Therefore, each of the four CONUS Regions were asked to provide a forecaster from their respective areas. Three conditions were recommended by the OPG staff.

The first condition was that each region select a meteorologist or a Science and Operations Officer (SOO) with sufficient breadth of experience such that he/she would not be intimidated by being placed into analysis and forecasting challenges from a variety of locations and situations. Second, all candidates were required to be reasonably proficient with the AWIPS-2 operating platform. At the time of this experiment, some offices were still operating on the legacy AWIPS-1 system. The ORE was to be conducted strictly on AWIPS-2 workstations and training would not be provided to participants on the unique features of that system. Finally, the request was made that all nominees come ready to be actively engaged, open-minded, and willing to share their opinions candidly. This was critical to obtaining an unbiased result because their opinions, findings, concerns, and recommendations would represent the field forecaster viewpoint concerning the TM's operational value.

Four excellent candidates were selected, each exceeding the requested qualifications. The group was comprised of two Lead Forecasters (Kent Prochazka, WFO Houston and Rob Radzanowski, WFO State College), one General Forecaster (Mark Wankowski, WFO Pueblo); and one SOO (Marc Singer, WFO Billings). The SME was Geoffrey Stano of NASA SPoRT and Tiffany Meyer of the WDTB represented the Training Division. The two programmer/developer contributors were Jason Burks of NASA SPoRT and Ken Sperow of MDL. OPG staff included Chad Gravelle, Chief Scientist; Jack Richardson, Systems Engineer; Andrew Ansorge, Case Development Meteorologist; and Kim Runk, OPG Director. Figure 1 shows the ORE participants during one of the evaluation scenarios.

Figure 1. Participants during one of the evaluation sessions for the NASA/SPoRT-MDL Tracking Meteogram ORE at the NWS Operations Proving Ground in May, 2014.

B. Evaluation Process

Participating NWS forecasters were asked to travel Monday morning and report for familiarization training Monday afternoon. This allowed the entire group to be prepared for active evaluation sessions scheduled to begin Tuesday morning. On Monday afternoon, Geoffrey Stano provided a tutorial that outlined the background, purpose, and design of the TM. This led into a hands-on demonstration on accessing the tool from the AWIPS menu and creating tracks and meteograms for various gridded parameters. Finally, Geoffrey guided the NWS forecasters through a few cases, describing the assorted parameter options a user can plot (maxima, minima, differences, etc.) and illustrating ways early testers have found those choices valuable for diagnosing trends. Following a brief Q&A session, ORE participants were given two archived cases with which to experiment and become familiar with the TM's basic operation. Initial impressions were captured in an afternoon debrief period. The goal of the training session was to ensure all participants were sufficiently comfortable and proficient with the TM in order to be prepared to engage in a series of weather scenarios during the course of the next three days.

On Tuesday, Wednesday, and Thursday, NWS forecasters were presented with multiple weather forecast challenges of varying complexity, mode (e.g., tornado outbreak, derecho, sub-severe convection) and geographical locations. In general, the weather scenarios became increasingly complex as the evaluation process progressed. Six cases were comprised of archived data sets, each selected for their value in assessing a specific warning or forecast problem. The two most complicated cases were used as TM stress tests. For the first stress test, NWS forecasters were instructed to approach the warning task as individuals. In the second, they were asked to perform as a team. For one scenario, in addition to using the TM to interrogate meteorological data, forecasters were asked to specifically focus on how the information extracted from the TM could be used to provide decision support services (DSS).

For one of the morning sessions, live weather data was utilized. During that period, the objective was to give each NWS forecaster the freedom to experiment and innovate with any gridded dataset in any part of the country. Variations in types of weather and event locations were deliberately selected to facilitate as comprehensive an assessment as possible for the TM's practical utility as a forecast decision aid. See Appendix 2 for a complete list of the scenarios used in the ORE.

At the conclusion of each scenario, immediate feedback was collected in 5-10 minute post-session assessment questionnaires. Once these questionnaires were completed the ORE participants spent between 15 and 30 minutes in candid, facilitated discussion about the TM experience during the scenario, focusing on usability, usefulness, specific diagnostic or predictive applications attempted, and any caveats or concerns they wished to share. These conversations were recorded to ensure comments, findings, and recommendations could be captured accurately for this report.

C. System Specifications

The ORE was conducted on an AWIPS-2 system, running Build 14.1. This is significant as it marks the first AWIPS software build that utilizes a 64-bit operating system. Each forecaster was assigned their own workstation for the duration of a scenario. Prior to beginning the next scenario each NWS forecaster rotated to a new workstation. Three of the four workstations were identical legacy workstations. The fourth was configured with a prototype monitor array, which was the focus of a separate but concurrent evaluation.

The prototype workstation, shown in Fig. 2 being used by Kent Prochazka from the NWS Forecast Office in Houston, TX, was an HP Z620 with one Intel Xeon E5-2630 v2 6-core CPU, 32GB RAM, and 500GB SATA HD. The video card was upgraded to an NVidia GeForce GTX 760 GPU with 4GB of 256-bit GDDR5 memory. The primary workspace, utilized for graphics, imagery, model output, etc., consisted of two (2) 27-inch LCD IPS panel monitors (Dell U2713HM), which were configured with spanning desktops. An additional 19-inch legacy monitor was provided as a standalone desktop, strictly for composing and displaying text products.

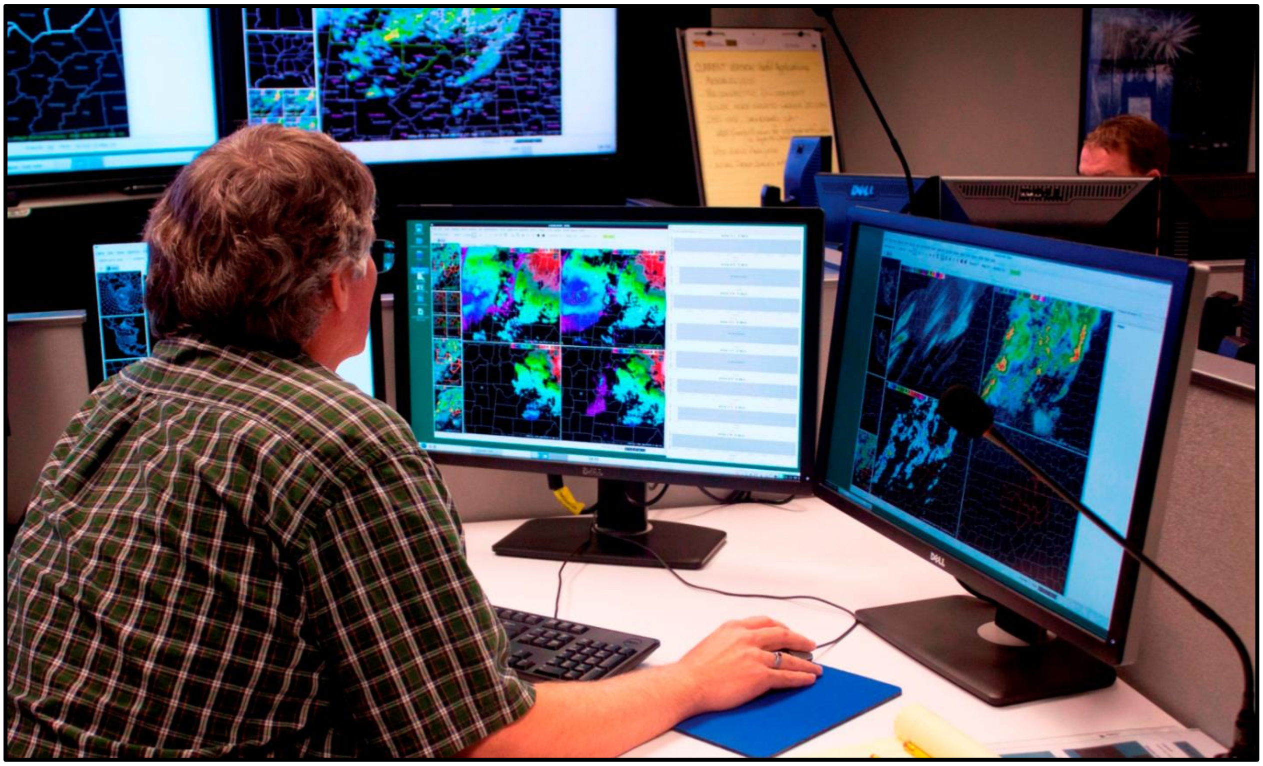


Figure 2. Kent Prochazka, Lead Forecaster from the NWS Forecast Office in Houston, TX, evaluates the tracking meteogram on the new prototype workstation.

This prototype configuration is being assessed as a potential upgrade for WFOs during the next hardware replacement cycle. Four areas of inquiry were explored to evaluate forecaster impressions and opinions on the design of the workstation layout, as well as its usability:

(1) Spanning desktops with continuous workspace, distributed as user sees fit
(2) One keyboard and mouse controlling real estate of both screens
(3) Swapping tabs
(4) Separate text option (19-inch monitor)

Thus far, opinions are unanimous that the improved performance and enhanced flexibility are superior to the legacy display set-up of three 19-inch monitors with fixed screen area.

III. Operational Readiness Evaluation Results

A. Tracking Meteogram - Forecaster Evaluations

After each scenario, NWS forecasters expressed their impressions of the TM by completing a post-scenario assessment questionnaire (some examples of the questionnaires are provided in Appendix C). These questions focused on the TM's ease of use, actual performance as compared to expectation, applicability to the scenario, quantitative and qualitative interpretive value, impact on warning decisions, and potential use for DSS applications. Unique questionnaires were developed for the DSS case and the live data experiment because of their focused objectives. All post-scenario assessment questionnaires were supplemented by a discussion that included all of the ORE participants that typically lasted between 15-30 minutes. During these discussions, NWS forecasters shared observations and ideas from their experience completing the scenario. In some cases, survey results were reinforced; in others, new information emerged, which had not been captured by the written accounts. With their permission, those discussions were recorded to preserve accuracy. A final questionnaire was completed at the end of the week to summarize forecaster findings, concerns, lessons learned, and recommendations (both positive and negative).

Overall, NWS forecasters unanimously endorsed the TM as a useful tool that could add genuine, unique value to several practical forecaster tasks. Listed below are key effective practices and specific findings and recommendations from the evaluation.

<u>*Effective Practices*</u>

Effective Practice 1: The TM is very useful as a forecaster decision aid for tracking and displaying trends with meteorological features (e.g., mesocyclone, reflectivity core, cloud-top temperatures, and total lightning data).

> *"The TM was very useful for monitoring the trend of intensification of a circulation at 3.4 degrees with the storm approaching KTLX. I didn't have to try to keep track of it in my head, this was very helpful."*

An example illustrating how a mesocyclone could be tracked and monitored in operations is shown in Fig. 3 (next page).

Effective Practice 2: The TM is useful at interrogating gridded model output.

"This would help to build confidence in forecast decisions and may even help to highlight areas of improvement needed in grid forecasts, particularly with respect to temperature changes associated with a frontal passage."

Effective Practice 3: The TM is useful for mesoanalysis applications such as monitoring trends in instability, convective inhibition, shear, and rainfall rates.

"I could see real value in using the tracking tool for trends in near storm environment across the CWA as well as for forecast trends for specific points (TAFs, DSS events across the CWA)."

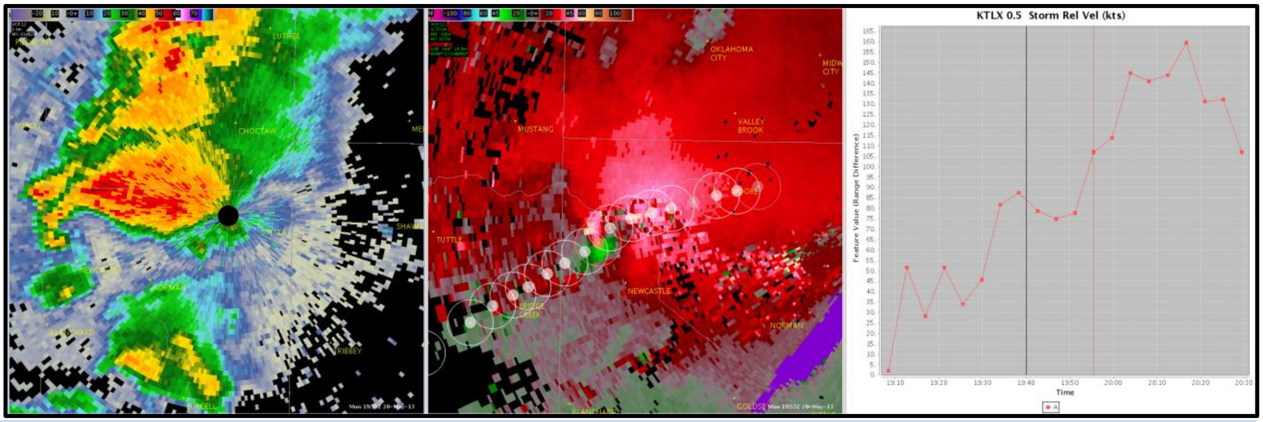


Figure 3. KTLX WSR-88D base reflectivity at 1955 UTC 20 May 2013 (left), KTLX WSR-88D storm-relative motion at 1955 UTC 20 May 2013 and tracking meteogram range circles (middle), and the resulting storm-relative motion tracking meteogram trace between 1909 UTC and 2029 UTC 20 May 2013 (right, vertical black line indicates the time when the tornado warning was issued and vertical red line indicates the time when the tornado occurred).

Effective Practice 4: The TM is useful at providing the forecaster an efficient way to extract and communicate meteorological information that is critical to NWS' core partners.

"It could be a different way to display ensemble model data to our customers, and it could be focused specifically for what they are looking for."

Figure 4 below provides an example of using the TM with PGLM total lightning flash extent density observations. The TM allows the forecaster to quickly assess quantitative output which allows them to provide fast and accurate information to NWS core partners on weather threats.

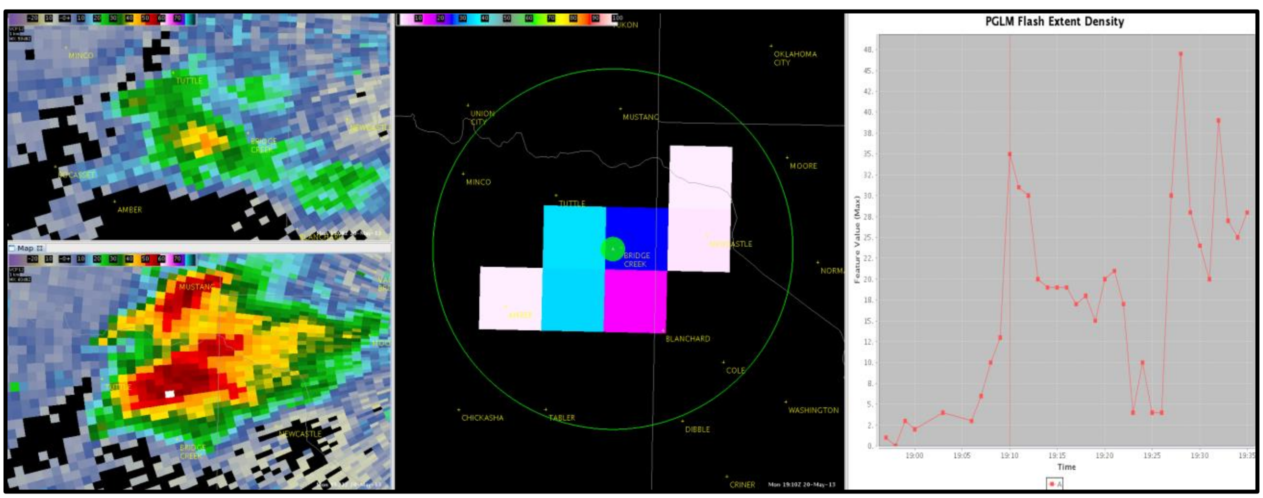


Figure 4. KTLX WSR-88D base reflectivity at 1908 UTC 20 May 2013 (left top), KTLX WSR-88D base reflectivity at 1921 UTC 20 May 2013 (left bottom), PGLM flash extent density at 1910 UTC 20 May 2013 and tracking meteogram range circle (middle), and the resulting PGLM flash extent density tracking meteogram trace between 1857 UTC and 1935 UTC 20 May 2013 (right).

Effective Practice 5: The TM would be useful in post event analysis to evaluate how trends in data were related to warning decisions. (to improve warning process)

2 . Forecaster Findings and Recommendations

Finding 1: Forecasters found it difficult to assimilate and interpret the TM data when plotting four or more meteograms concurrently. Forecasters also observed that the depiction of numerous meteograms impacted system performance.

Recommendation 1: Limit the number of individual TMs displayed to minimize adverse impacts on the forecaster's interpretive capability and the system's processing efficiency.

Finding 2: To utilize the TM to monitor in-situ changes over a fixed location, it was necessary for the user to manually stack the TM's trace circles.

During the live data experiment using model output, forecasters recognized that there would be value in creating a TM stationary mode. A TM stationary mode would function as a plume diagram. As meteorological variables advect through the TM's radius of influence, a trace would be created. NWS forecasters suggested this functionality could be used for aviation forecasting, flash flooding within wildfire burn scar areas, monitoring fog development (Fig. 5), and situational awareness for DSS.

"After trying a variety of ways to look at the model data, I eventually seemed to settle into mainly using it as a stationary location tool and letting the model data run through it."

Recommendation 2: Create a TM stationary mode as a menu choice prior to AWIPS-2 field implementation.

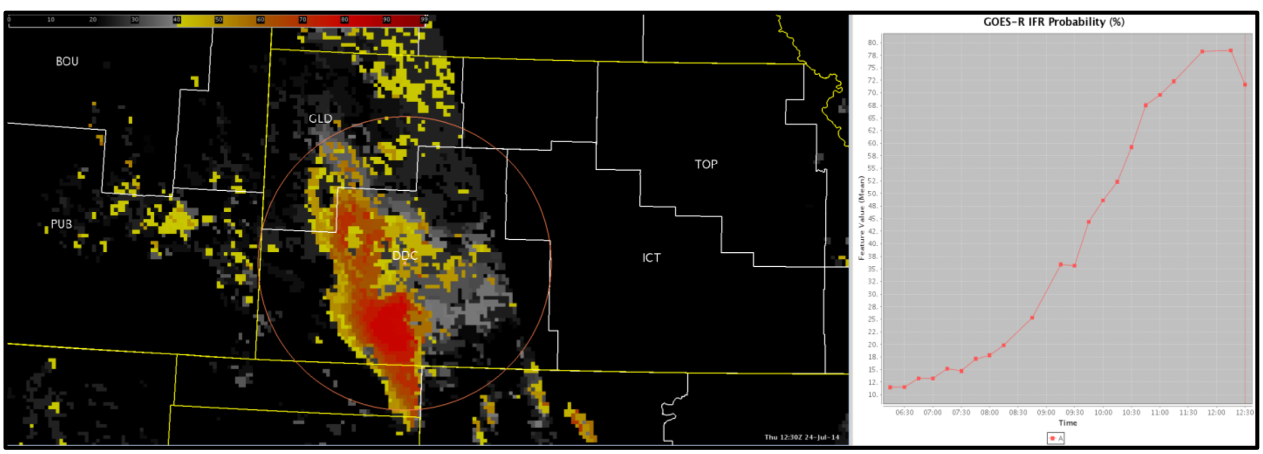


Figure 5. GOES-R fog and low stratus instrument flight rules probability product at 1230 UTC 24 July 2014 (left) and the resulting GOES-R fog and low stratus instrument flight rules mean probability tracking meteogram trace between 0615 UTC and 1230 UTC 24 July 2014 (right).

Finding 3: Forecasters discovered that it is not possible to store a loaded TM location into an AWIPS-2 procedure.

This capability will be important to use with the TM stationary mode (Finding 2) as the user can set up predetermined locations within the procedure. For example, if the user would like to monitor how meteorological variables within 20 km of a TAF site or DSS location are changing, they could create an AWIPS-2 procedure which an instance of the TM centered on that location. Currently, when a procedure is created with the TM loaded, the TM location is not saved as one of the procedure's elements.

Recommendation 3: Add the capability to store a stationary mode TM into an AWIPS-2 procedure prior to AWIPS-2 field implementation.

Finding 4: Functionality does not exist to set and plot a user-defined threshold value superimposed on the TM graph. Forecasters identified two specific applications of value for this feature: (1) as a visual reference on the trace to aid the forecaster's analysis and forecasting decision process, and (2) as an optional trigger point to activate a notification alert message when exceeded.

Recommendation 4: To optimize DSS effectiveness, developers should consider adding this user-defined threshold functionality to an upcoming TM version.

Many TM concerns NWS forecasters identified during the ORE were remedied by developers on the fly. In fact, more than 30 issues were resolved over the course of the week and the TM went through six revisions from Monday to Thursday. The TM version in use on the system by the end of the week was enthusiastically and unanimously supported by the evaluating forecasters for operational implementation. Typical of this opinion was the NWS forecaster comment:

"*If you could get this installed on our AWIPS tomorrow, I guarantee I would use it and promote it to my fellow forecasters.*"

B. The ORE Process - Participant Observations and Recommendations

While many topics and observations were mentioned through the written surveys and oral discussions, the most significant take-aways concerning the ORE process fell into four broad categories:

- Integration of the VLab Development Environment in the ORE Process
- Diversity and Realism of Cases Used for Evaluation Sessions
- Emphasis on Human Factors in Evaluating Usability and Usefulness
- Blend of Forecasters, Trainers, Developers, and Support Staff Who Participated

1. VLab Development Environment

Prior to the ORE, a TM development community had been in place within VLab as a clearinghouse for documenting code updates, maintaining version control, and tracking the project's progress toward AWIPS-2 implementation. During Monday's introductory training session, Ken Sperow enrolled the ORE participants into the TM's VLab community as members. This allowed participants to enter observations, software issues, and feature requests directly into the community database. When software issues such as minor bugs or new feature requests were documented in the database, they became a natural part of the ORE process and developers could investigate. As alluded to previously, in many cases these issues were resolved on site, and an updated version of the TM code was installed on the operational system in time for ORE forecasters to test the enhancement in the next scenario. This comment made by one of the NWS forecasters is perhaps the best portrayal of the group's view:

"The VLab environment really streamlined the process of reporting issues. Developers were very responsive to suggestions for adding new features, and to fixing minor bugs in time for the next case."

2. Diversity and Realism of Cases

In order to derive a meaningful conclusion about the operational readiness of any new tool or capability, it is important to evaluate its use in multiple contextual circumstances. To the fullest extent possible, consideration should be given to building OREs that encompass a variety of weather conditions, climate regimes, seasons, geographic locations, and service challenges. The capability to accomplish all those goals for the TM was limited by the need for all cases to be AWIPS-2 based. For example, the TM was not assessed with a winter-weather scenario. However, TM ORE sessions were conducted in five different geographic locations, each featuring a different warm season weather challenge. Additionally, for one entire morning a live data feed was used, and forecasters were encouraged to choose their own locations of interest to test applications of the tool. In general, forecasters felt the variety of cases was sufficient to give the TM a robust evaluation. In addition to the assigned analysis and forecasting tasks, unplanned distractions were injected into the team stress test scenarios in an effort to simulate the interruptions and chaos of an actual warning event. Two NWS forecasters suggested it would appropriate to inject even more diversionary activity into future ORE scenarios.

Nevertheless, the overall opinion was positive. One of the participants stated,

"I was impressed with the diversity of scenarios we worked and the overall set-up of the OPG. It simulated a real operational environment fairly well, and allowed us to explore a variety of uses for the tool."

The lack of an efficient archive case playback capability built into the AWIPS-2 created a minor complication with the execution of scenarios. For archived data sets, the system's Local Data Monitor (LDM) feed was shut down, the real-time database purged (using a script developed by Jason Burks), and the scenario case data could then be initiated for pseudo real-time streaming. Purging the data was necessary because the archived scenario data contained its native time-stamp information and would therefore be overwritten by real-time information immediately.

This restriction caused OPG staff to turn the LDM feedback on and let the system run overnight to build up enough data into the system to use for the live data scenario. A major finding concerning the ORE process is that *the OPG needs greater versatility in transitioning between live data and archived cases.*

The OPG staff has identified this as a priority issue to resolve and has already begun discussions to design and build a prototype solution in FY15.

3. Emphasis on Human Factors

One characteristic that distinguishes OPG evaluations from traditional testbed experiments is the deliberate attention paid to human factors. In particular, the OPG is concerned about how tools and capabilities may have adverse impacts on forecaster workload, impedance of workflow, human-computer interface, product design, and the interplay between physical sciences and social sciences with respect to the process of translating scientific expertise into effective, persuasive messaging. This emphasis was integrated into the fabric of every scenario and discussed explicitly during each post-scenario discussion with NWS forecasters.

The prominence placed on human factors was appreciated by all the NWS forecasters who participated in the ORE. During a post-scenario discussion, one of the NWS forecasters noted,

"Evaluating whether a tool provides some unique value compared to existing tools is really important."

Another immediately added,

"So is ensuring the tool doesn't create a workload problem. You break trust if you promise something cool, and then forecasters discover it's just another way to do something they already do, with no clear benefit."

During this ORE, forecasters had absolutely no concerns regarding the potential for the TM to create additional workload or any other negative human factor impacts that might compete against TM acceptance and use when implemented operationally.

4. Combination of Participants

As briefly mentioned earlier, a good deal of forethought was given to the composition of the ORE participants. The need for competent, forward-thinking NWS forecasters is self-evident. These individuals must be capable of learning quickly, adapting that learning to operations, able to consider alternatives to the proposed utility, and willing to offer candid and constructive critique about any potential issues and/or concerns. NWS forecasters at this ORE unanimously expressed appreciation for the presence and contributions of the SMEs and training representative. The SMEs explained and demonstrated how to use the TM, remained available to answer questions and troubleshoot problems, and in turn learned useful information from the NWS forecasters about the type, format, and length of familiarization training that should be developed and distributed to field forecasters with the rollout of the TM. Developers and OPG technical staff also contributed to the flow, content, and effectiveness of the ORE. One forecaster commented about this unique ORE demographic,

"The cross-section of people involved was fantastic. Having forecasters interact with leaders, developers, researchers, and trainers – who together demonstrated genuine collaboration – that was powerful. Our organization preaches collaborative innovation. This experience nailed it. Every NWS forecaster needs to experience something like this."

III. OPG Recommendation

All participating NWS forecasters unanimously endorsed the TM as a useful tool that will add unique value to several specific diagnostic and predictive tasks without posing adverse impacts on operational workflow or forecaster workload. While NWS forecasters believe it will be received well and readily adopted by forecasters in its current version, they recommend two enhancements prior to widespread field implementation. First, it is recommended that the stationary mode option, AWIPS-2 procedure capability, and tendency information be added to the TM's functionality. Second, a brief TM training video and one-page (front and back) informational handout should accompany the TM's implementation.

Based on the results of this ORE, the OPG recommends implementation of the Tracking Meteogram into the AWIPS-2 baseline without reservation.

APPENDIX 1 - Background on the NWS Operations Proving Ground

The OPG was created as a part of the Weather-Ready Nation initiative to improve the transition of new and emerging capabilities from research-to-operations (R2O); and the reciprocal operations-to-research (O2R) feedback loop essential to maintain relevance, usefulness, and developmental focus. The concept is motivated by, and founded upon, the following premises:

1. Investment in efficient and effective R2O/O2R is a priority for the NWS.

- Operationalizing weather-related research advancements is key to a science-based service organization's ability to remain progressive and effective.
- Improving transition of new capabilities in science and technology to operations is a stated goal of Weather Ready Nation Roadmap. It is also explicitly recommended in both the 2012 National Research Council Report ("*Weather Services for the Nation: Becoming Second to None*") and the 2013 National Academy of Public Administrators Report ("*Forecast for the Future: Assuring the Capacity of the National Weather Service*").

2. Opportunities exist to enhance, streamline, and optimize R2O/O2R to meet the emerging needs of the Weather Ready Nation paradigm more effectively and efficiently.
Components crucial to effective R2O/O2R processes include:

- Identification and communication of operational requirements to the research community, driving the prioritization of applied research projects.
- Mechanisms to provide feedback from operations to labs and universities, ensuring research investment connects to practical application and operational needs.

3. Increased engagement and collaboration enhance value financially and functionally.

- Improved resource alignment strengthens results and makes better use of available staff and funding in a constrained budget climate.
- Working together maximizes the likelihood of achieving mutually-shared goals.
- A collegial model affords opportunities to create a development, training, and implementation framework that integrates diverse science disciplines and skill sets; is characterized by greater transparency, accountability, and sustainability; and incorporates input from both internal and external stakeholders.

The NWS operates in a dynamic environment and keeping pace with technology, science, and stakeholder expectations are key challenges. Furthermore, the needs of core partners are evolving, especially with respect to the increasing role the NWS serves within the National Response Framework with FEMA and other emergency management decision makers. Therefore, the NWS must maintain an ongoing capacity for development and testing of its incremental technology upgrades, service enhancements, and infusion of new science into operations. Testbeds, Proving Grounds, and Pilot Projects aim to facilitate orderly transition of research to operations through developmental testing, pre-deployment testing, operational readiness evaluations, as well as usability and suitability checks. Given the likelihood of a protracted, austere budget climate, there is a marked urgency for the NWS to consider shifts in how resources are allocated and aligned to support and extend these valuable activities. By creating a broader coalition among the various testbed operations, the strengths of multiple parties can be coordinated and leveraged such that capabilities are improved with minimal cost and disruption to the organization.

In the OPG, testing of advanced operations, services, and/or new science and technology capabilities are facilitated via Operational Readiness Evaluations. The origin of this practice is rooted in the need for some tools to be exposed to more rigorous testing, beyond scientific validity and practical utility, prior to being recommended for implementation in field offices nationwide. History has shown that, in some cases, it is important to evaluate a new capability in conditions which simulate an actual production environment more closely than that which exists in a typical testbed experiment.

Thus, every effort is made to ensure formal ORE sessions are conducted in a realistic NWS operations-like setting. Candidate capabilities are identified from proposals submitted by participating NOAA testbed facilities or supplied to the OPG through NWS Regions, NWS HQ, or major acquisition programs, as appropriate. Prior to any ORE, a test plan and associated performance metrics will be defined for each candidate capability. Measurement categories will include, as appropriate:

- objective performance (e.g., accuracy/skill)
- subjective evaluations of utility (e.g., user feedback on balance positive)
- production/engineering readiness (e.g., systems and communications reliability/security/backup, data retention)
- workflow/workforce impacts

OREs are designed to ensure that promising new tools and decision aids emerging from testbeds are not only endorsed by forecasters as useful to the forecast process, but that they also present no adverse impacts on human factors, such as workflow, workload, cognitive assimilation, situational awareness, communication of hazards and impacts, and other forecast decisions.

If selected, candidate capabilities undergo validation activities, placing equal weight on three aspects of the forecast process:

1. usefulness of the tool, data set, or capability for improving forecaster decision making; enhancing forecast accuracy or confidence;
2. impact on workload, work flow, processing efficiency, value added in comparison with other known tools or practices, and other human factors; and
3. effective communication of weather information and associated risks/impacts to partners, including national-regional-local interactions in end-to-end service delivery.

All three factors must be endorsed as having met test plan objectives successfully without contributing any appreciable negative impact on existing systems and practices.

Training to Accompany New Tools

It is well known that data overload can create numerous challenges to the forecast decision process. Inherent confusion can arise simply because the forecaster has difficulty sorting through enormous amounts of information available to them in order to select the right information for the decision at hand. Therefore, coincident with the infusion of new science and technology into operations, it is crucial to develop appropriate effective practices and training to use the data being introduced. The OPG is committed to this end-to-end responsibility, as are other sister operations affiliated with the NOAA Testbeds and Proving Grounds. Since the OPG is co-located with a training center, a national center, and a regional headquarters for three federal agencies who act in partnership for disaster response activities (i.e., NWS, FEMA, EPA), it is able to take advantage of a networked community of partners who can collaborate on the identification of professional training needs associated with new tools and capabilities and, in some cases, in the development and delivery of that training.

APPENDIX 2 - Weather Scenarios and ORE Participants

Case 1 - 19 June 2013 (2200 - 2330 UTC) West Texas

Case 2 - 9/10 May 2013 (2300 - 0030 UTC) West Texas

Case 3 - 20 May 2013 (1800 - 2115 UTC) Central Oklahoma (Individual Stress Test)

Case 4 - 27 April 2013 (2030 - 2345 UTC) Northern AL (DSS Case)

Case 5 - 30 June 2012 (0145 - 0315 UTC) Washington D.C.

Case 6 - 27 April 2011 (1915 - 2230 UTC) Northern AL (Team Stress Test)

Supplementary Case - 28/29 June 2013 (0130 – 0300 UTC) Eastern Colorado

(Wednesday morning was devoted to Real Time Weather Evaluation)

ORE Participants

Marc Singer, WFO Billings, MT
Kent Prochazka, WFO Houston/Galveston, TX
Rob Radzanowski, WFO State College, PA
Mark Wankowski, WFO Pueblo, CO

Geoffrey Stano, NASA SPoRT, Huntsville, AL
Jason Burks, NASA SPoRT, Huntsville, AL
Ken Sperow, NWS MDL, Silver Spring, MD
Tiffany Meyer, WDTB, Norman, OK

Chad Gravelle, NWS OPG, Kansas City, MO
Andrew Ansorge, NWS OPG, Kansas City, MO
Jack Richardson, NWS OPG, Kansas City, mo
Kim Runk, NWS OPG, Kansas City, MO

APPENDIX 3 – Summary of Findings and Recommendations

Finding 1: Forecasters found it difficult to assimilate and interpret the TM data when plotting four or more meteograms concurrently. Forecasters also observed numerous meteograms impacted system performance.

Finding 2: To utilize the TM to monitor in-situ changes over a fixed location, it was necessary for the user to manually stack the TM's trace circles.

Finding 3: Forecasters discovered that it is not possible to store a loaded TM location into an AWIPS-2 procedure.

Finding 4: Functionality does not exist to set and plot a user-defined threshold value superimposed on the TM graph. Forecasters identified two specific applications of value for this feature: (1) as a visual reference on the trace to aid the forecaster's analysis and forecasting decision process, and (2) as an optional trigger point to activate a notification alert message when exceeded.

Recommendation 1: Limit the number of individual TMs displayed to minimize adverse impacts on the forecaster's interpretive capability and the system's processing efficiency.

Recommendation 2: Create a TM stationary mode as a menu choice prior to AWIPS-2 field implementation.

Recommendation 3: Add the capability to store a stationary mode TM into an AWIPS-2 procedure prior to AWIPS-2 field implementation.

Recommendation 4: To optimize DSS effectiveness, developers should consider adding this user-defined threshold functionality to an upcoming TM version.

ORE Process Positive Take-Aways

1. Incorporating the VLab development environment in the evaluation process streamlined procedures for reporting and documenting issues, prioritizing development tasks, and guiding updates or revisions.

2. The set-up of the OPG and the diversity of cases served to simulate the operational environment realistically, and allowed for a robust, comprehensive evaluation. Adding more injects and distractions to warning event simulations would add to the realism for future evaluations.

3. Maintain the emphasis on human factors. Evaluating whether a tool offers unique value without adversely impacting the workload is a crucial element which has heretofore been largely ignored.

4. Gathering the right combination of participants (SMEs, trainers, developers, forecasters, technical staff, and leadership) created an environment that encouraged collaboration, learning, sharing, and honest evaluation.

ORE Process Findings and Recommendations

Finding 1: The OPG needs greater versatility in transitioning between live data and archived cases.

Recommendation 1: Develop a system that utilizes a secondary AWIPS back end, which can be accessed by ORE forecasters for ingest, data storage, processing and data basing during archived payback cases. Meanwhile the main AWIPS system can continue ingesting and processing to accommodate quick conversion to live data feeds for ensuing experiment sessions. If eventually fielded at WFOs, such a system would permit a breakthrough in that it permits whole office simulation training (vs. the existing single workstation WES model).