Earth Networks Total Lightning Data and Dangerous Thunderstorm Alerts Evaluation in the NOAA Hazardous Weather Testbed

Kristin M. Calhoun[1,2,*], Darrel M. Kingfield[1,2], Tiffany Meyer[1,2], Woody Roberts[3], Jim Ramer[3,4], Brian Motta[5], and Lans Rothfusz[2]

[1]University of Oklahoma / Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), Norman, OK

[2]NOAA/OAR/National Severe Storms Laboratory (NSSL), Norman, OK

[3]NOAA/OAR/ESRL/Global Systems Division (GSD), Boulder, CO

[4]Cooperative Institute for Research in the Atmosphere (CIRA), Boulder, CO

[4]NOAA/NWS/Forecast Decision Training Division, Boulder, CO

March 2016

---

[*]corresponding author address: kristin.kuhlman@noaa.gov

Executive Summary

Earth Networks, Inc., (ENI) has indicated the potential for their total lightning data and automated Dangerous Thunderstorm Alerts (DTA) system to help increase lead-times over current National Weather Service (NWS) severe weather and tornado warnings, while maintaining a similar probability of detection. In order to test the value of ENI total lightning data and algorithms within National Weather Service warning operations, both the Earth Networks Total Lightning Network (ENTLN) data and products from the DTA system were implemented within NWS operational software (Advanced Weather Interactive Processing System 2nd generation, AWIPS2) for evaluation in NOAAs Hazardous Weather Testbed (HWT) in Norman, OK during two experiments. Additionally, a verification study of both the DTAs and NWS severe and tornado storm-based warnings was undertaken using NWS verification methodology for the same period of data.

Following an initial period of development and testing, the first experiment ran 21 July - 29 August 2014. This experiment included 18 NWS forecasters in the HWT over a period of six weeks for a full product evaluation. The forecasters completed a series of six two-hour long weather-warning simulations in displaced real-time across a variety of convective regimes throughout the United States ranging from marginally severe to high-impact tornadic events. Utilizing a repeated measures design, each forecaster was randomly assigned one of three tiers of data to be used during the simulation. The data tiers were as follows:

- Tier 1: The full suite of WSR-88D radar products available in AWIPS2.
- Tier 2: ENTLN total lightning point data and all radar products available in tier 1.
- Tier 3: ENI total lightning cell tracking and flash rate products and associated alert polygons in addition to all products available in tier 2.

Forecaster feedback was collected through a series of online surveys, interviews, and group discussion. Results show that all three tiers of forecasters performed similarly with the Tier 2 group (total lightning + radar) performing slightly better in terms of overall false alarm ratio and probability of detection for all events. Based on skill scores (i.e., Probability of Detection and False Alarm Ratio), forecasters that were not already experts at radar interrogation and severe storm forecasting saw the most benefit from the inclusion of total lightning data and associated products during warning operations. Forecaster feedback from the 2014 experiment highlighted a desire to set dynamic

thresholds for the alerts, the ability to view the lightning flash rates as a time series plot, and for more in-depth training. Overall, the results from year one indicate that while the forecasters see the total lightning and derived products as useful in warning operations, the DTA polygons themselves have limited value.

A successive real-time evaluation in the HWT was completed 4 May - 12 June 2015. The domains of operations for this experiment were decided daily based on likelihood of severe weather for a region. Over the course of the experiment, we operated in 42 different NWS county warning areas with 31 forecasters participating. This real-time experiment provided additional insight on how the forecasters would use the data for warning operations when they had access to all currently available products and provided a stress test for the timeliness and usability of the ENTLN data and tools within operations. Similar to year one, forecasters gravitated to the total lightning data and storm-based derived flash rate trend information as well as the newly available time series display. Multiple forecasters noted in discussion and blog posts that the three layers of alerts (including the DTAs) often cluttered the screen and did not add much value to the warning process. Additionally, forecasters had difficulty using the alerts when the domain of operations crossed $104°$W longitude (e.g., Midland, TX), as ENI applies different thresholds for the alerts due to detection efficiency differences. On these days of operations, forecasters were confused on which set of products to use (i.e., East vs West) and often ceased using the derived products and alerts. However, throughout the experiment surveys and blog posts indicated that a majority of forecasters did find value in the additional DTA-system total lightning tools, such as the storm tracking and time series information. Additionally, some forecasters required these additional tools and derived products to feel comfortable utilizing the total lightning data in real-time operations.

Finally, an event-specific verification analysis was completed across the United States of the DTA polygons against NWS severe thunderstorm and tornado warnings using *NOAA StormData* over the period 1 January 2013 - 30 September 2015. For equity of comparison and verification with NWS warnings, all DTAs that did not overlap land area of the CONUS were removed (e.g., DTAs over the Gulf of Mexico). Over this time period for the CONUS, NWS severe (tornado) warnings had a Probability of Detection (POD) of 79% (63%) and False Alarm Ratio (FAR) of 50% (71%). The DTAs had a POD of 49% for severe events, 53% specifically for tornadoes, and 49% for any severe or tornado report. The FAR for the DTAs for all categories was significantly higher than NWS warnings: DTAs had a FAR of 80% when compared to severe reports, 98% for tornadoes alone, and 80%

for any severe or tornado report. Lead time for tornadoes by the DTAs was longer than NWS tornado warnings (14.82 min compared to 7.86 min), but this was at the expense of the substantially higher FAR. For the period of study, NWS severe warnings had a better lead time than DTAs for severe reports (17.00 min compared to 14.82 min). Based on these verification statistics, it is doubtful that the alerts alone could be of much help to the warning forecaster.

Based on the two years of experiments within the HWT and the verification study, we recommend four items to ensure beneficial use of total lightning data and associated derived products by NWS forecasters: (1) additional training on the relationship between storm kinematics, dynamics and microphysics with lightning and best practices for operational use, (2) high-resolution detection efficiency maps for total lightning data, (3) delivery of storm-based flash rate decision-assistance tools such as the DTA-system time series and trend information in addition to the raw total lightning data (though not any of the alert products, including the top-level DTAs), and (4) removal of boundaries that bisect forecast offices for derived products.

Table of Contents

List of Figures

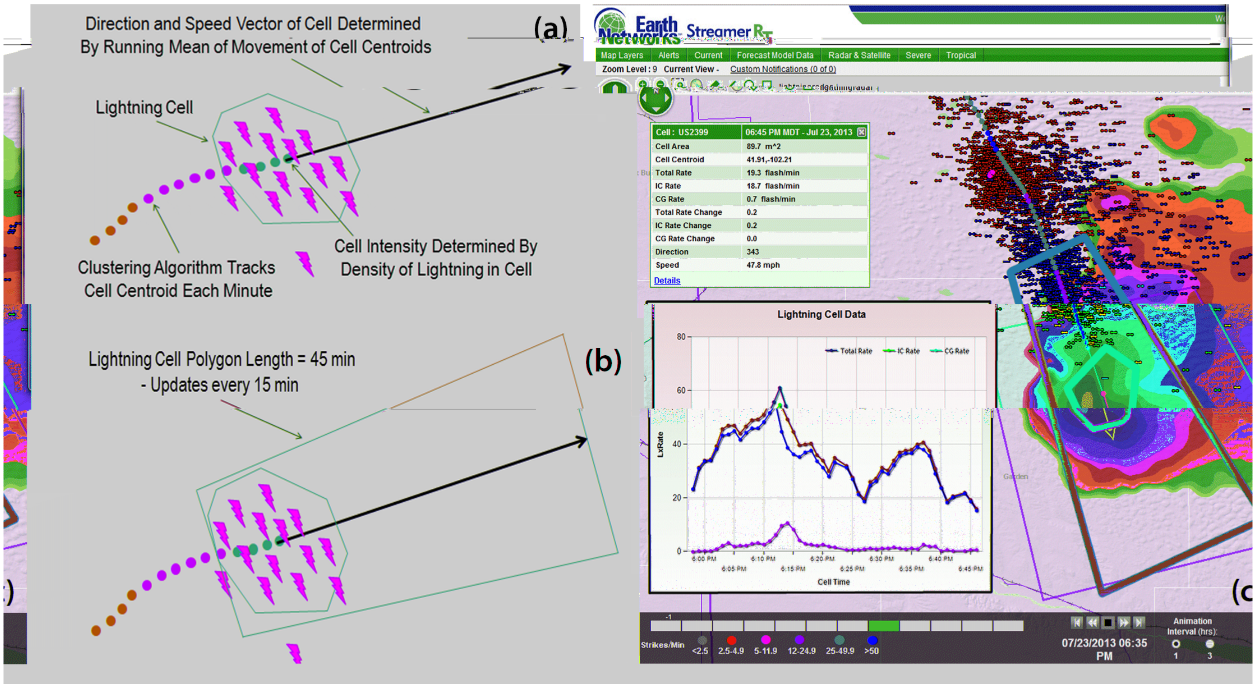
List of Tables

1 Introduction and Background

Previous studies incorporating total lightning data have shown that increases in activity may be a precursor or signal of severe weather potential (e.g., Goodman et al. 1988; Schultz et al. 2011) due to the inherent link between storm electrification and updraft size and strength (e.g., Saunders et al. 2006; Calhoun et al. 2013). In addition to the detection of total lightning (in-cloud and cloud-to-ground) flash rates, Earth Networks Incorporated (ENI) has developed the Dangerous Thunderstorm Alerts (DTAs) system based on a proprietary storm tracking and lightning flash rate threshold algorithm. The DTA system and products are meant to identify specific areas of increased potential for dangerous convective conditions and track significant storm cells over time (Bill Callahan, Earth Networks, personal communication).

Currently, the ENI DTA system produces a portfolio of total lightning derivative products for identifying and tracking storms across the continuous United States and adjacent coastal and land areas as well as multiple international regions for commercial and government users. The DTAs are a combination of cell identification and tracking algorithms employing multiple thresholds on the total lightning flash rate (Fig. 1 and Table 1). For NWS forecasters, the DTA system is currently available in real time outside the NWS operational Advanced Weather Interactive Processing System (2nd generation, AWIPS2) software using the ENI StreamerRT web application. The StreamerRT display includes a polygon specifying the location and forecast threat area as well as the cell identification, past track, and time series of the cell flash rates (Fig. 1).

Year one of this project integrated the Earth Networks data and products into the NWS AWIPS2 operational software and tested both the feasibility of use in warning operations as well as the impact on warnings issued by NWS forecasters in NOAAs Hazardous Weather Testbed (HWT) in Norman, OK. Year two modified the products based on forecaster feedback from Year one and tested the updated products within real-time operations in the HWT Spring Experiment.



Cell : US2399	06:45 PM MDT - Jul 23, 2013
Cell Area	89.7 m^2
Cell Centroid	41.91,-102.21
Total Rate	19.3 flash/min
IC Rate	18.7 flash/min
CG Rate	0.7 flash/min
Total Rate Change	0.2
IC Rate Change	0.2
CG Rate Change	0.0
Direction	343
Speed	47.8 mph

Details

Figure 1. ENI DTA system product creation and web-based visualization. (a) Lightning cell identification process determined via grouping of lightning density. (b) Cell polygon or alert created determined using cell-based flash rate thresholds in Table 1. Alerts last 45 min. (c) DTA system as visualized in ENIs web-based StreamerRT platform.

2 Year One Experiment

2.1 AWIPS2 Implementation

Initial development and production of the ENI DTA system inside of the AWIPS2 platform was completed by OAR/Global Systems Division (GSD); development of the raw total lightning flash locations was completed locally at the National Severe Storms Laboratory (NSSL). The goal for the year 1 evaluation was simply to replicate the display capabilities of the ENI StreamerRT web display. The individual components of the DTA system as visualized in the AWIPS2 software are shown in Fig. 2. This included: three levels of alerts for East/West United States storm-cell polygon objects (Table 1), cell-based flash rates, past tracks (colored by flash rate threshold) and storm projection. Raw latitude and longitude locations for both in-cloud (IC) and cloud-to-ground (CG) from the ENTLN were also available to forecasters within the AWIPS2 software.

Alert Type	**Flash Rate Threshold (East)**	**Flash Rate Threshold (West)**
General Thunderstorm	3 flashes min^{-1}	3 flashes min^{-1}
Significant Thunderstorm	20 flashes min^{-1}	12 flashes min^{-1}
Dangerous Thunderstorm	20 flashes min^{-1}	25 flashes min^{-1}

Table 1. ENI thunderstorm alert levels and respective flash rate thresholds for eastern and western US. The boundary between the east and west locations is set at $104°$W longitude due to differences in flash detection efficiency between the eastern and western US.

2.2 Experiment Design

The first HWT experiment ran 21 July - 29 August 2014 and consisted of 18 forecasters, three per week, with each participating in the evaluation Monday-Friday. Each Monday, forecasters reviewed the training material with the principal investigators. Following the initial training review and discussion period, the forecasters participated in two weather event scenarios specifically for additional training, one was configured as an enhanced case review (forecasters had the ability to see all the data for the entire two hours at once) and the other was viewed in displaced real time. During both of these training simulations, forecasters had access to all the data types in AWIPS2 that they would see throughout the week: radar, lightning, DTA system. During this time, forecasters were encouraged to develop procedures for product loading and display that they would use later in the week. Each forecaster then completed six two-hour long scenarios Tuesday through Thursday using the operational WarnGen software within AWIPS2 to issue warnings in displaced real time. The scenarios ranged in intensity from high impact tornadic events to marginally severe events and included a variety of geographic locations (Table 2). The order of the scenarios was swapped each week, such that no one scenario was biased by always being the first one seen by the forecasters. For each event, one forecaster acted as the control or Tier 1, operating with currently available radar and radar-derived products only while a second forecaster completed the same event with the addition of ENI total lightning data (Tier 2), and the third forecaster had access to radar, total lightning data, and all the total lightning-derived products from the DTA system, including storm cell identification, past track, cell flash rates, and alerts (Tier 3, Fig. 2). Each forecaster rotated through the tiers for each of the scenarios, with every forecaster working each tier twice.

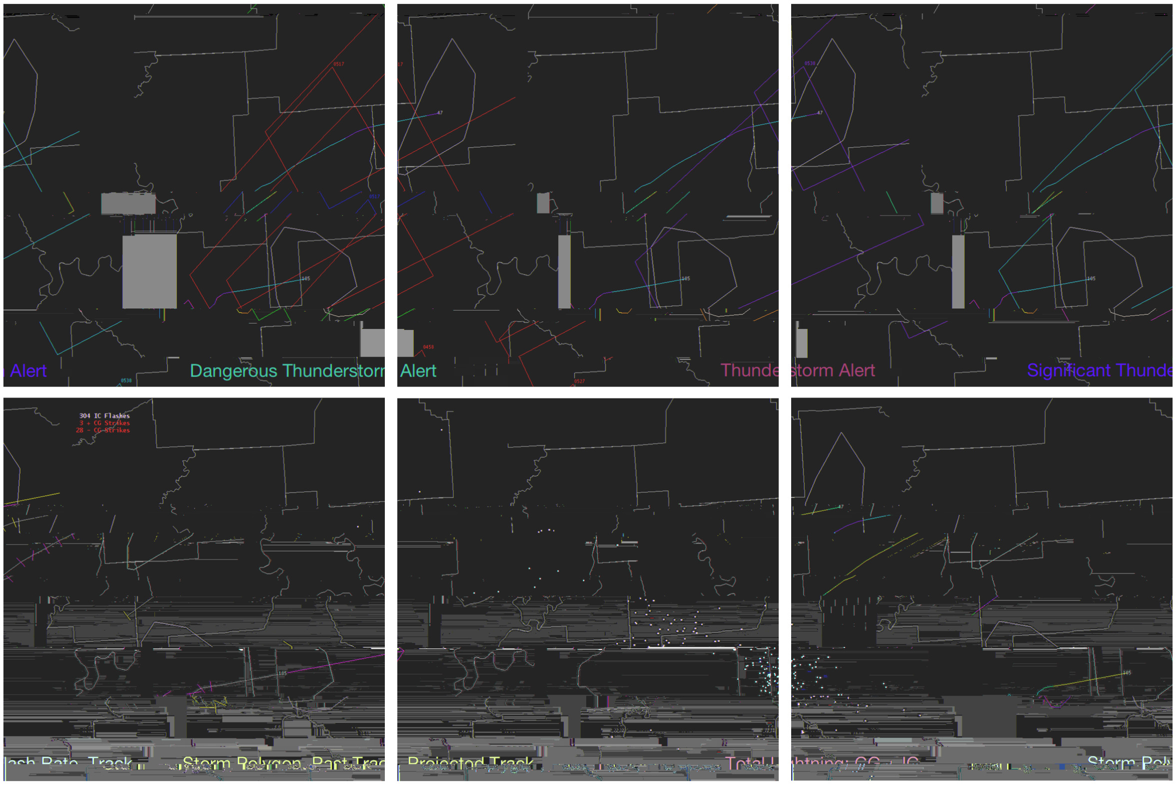


Figure 2. Earth Networks DTA system as visualized in AWIPS2. Top row: general thunderstorm alerts (green), significant thunderstorm alerts (orange) and dangerous thunderstorm alerts (purple). Bottom row: Storm polygon objects, storm-based flash rates, past track (colored by flash rate level) and projected track (blue). Total lightning points and count (IC=gray, CG= green).

Feedback was gathered through multiple instruments. In addition to warning statistics (e.g., Probability of Detection) from each forecaster for each event, a perceived workload (i.e., NASA task load index; Hart and Staveland 1988; Hart 2006) survey was completed after each scenario. Additionally, forecasters were asked to document his/her warning confidence for each warning that was issued. Meanwhile, an observer documented what products each forecaster was using throughout the simulation and walked through each warning decision or “key-judgment point” following the simulation to determine how and why the forecaster integrated the ENI total lightning data and DTA system products into his or her warning decision process if it was available during that event. Two separate group discussion periods were used to understand forecaster thoughts on overall utility, training, and best practices for NWS of both total lightning and the DTA system. Forecasters also completed pre-week and post-week surveys on the perceived knowledge and overall utility of lightning data.

County Warning Area	**Date and Time of Event**	**Synopsis**
Birmingham, AL (BMX)	17 May 2013 2145-2345 UTC	Scattered strong to near severe thunderstorms, many with weak rotation. All storms remained below severe threshold; no reports were received.
Sterling, VA (LWX)	13 June 2013 1800-2000 UTC	Severe mesoscale convective system, multiple reports of damaging winds followed by an embedded tornado near the Washington, D.C. metropolitan area.
Grand Junction, CO (GJT)	22 Sept 2013 1830-2030 UTC	Monsoonal moisture combined with a cold front moving across mountainous terrain resulted in numerous severe and sub-severe storms producing large hail and heavy rain in a region of poor radar coverage.
Paducah, KY (PAH)	17 Nov 2013 1830-2030 UTC	Strong low-level shear ahead of a pre-frontal trough of low pressure led to the development of a long-lived cyclic tornadic supercell storm within the warm sector of southern Kentucky.
Grand Rapids, MI (GRR)	12 April 2014 2000-2200 UTC	Two separate severe thunderstorms moved onshore from Lake Michigan. Both linear systems produced damaging winds and large hail.
Ft. Worth/Dallas, TX (FWD)	27 April 2014 2110-2310 UTC	Isolated supercell with extreme damaging hail (multiple reports $\geq 2\ 75$ in).
Birmingham, AL (BMX)*	28-29 April 2014 2345-0045 UTC	Multiple tornadic supercells in the warm sector ahead of a cold front. (scenario used for forecaster training)
Boulder, CO (BOU)*	21 May 2014 1800-2330 UTC	Multiple severe thunderstorms in and around the Denver metropolitan area. Storms produced large hail and five short-lived tornadoes. (scenario used for training only via enhanced case review)

Table 2. Local NWS county warning areas, date and time window, description of event, and reports for each scenario the forecasters viewed during the year 1 experiment. The BOU* (enhanced case review) and 2014 BMX* (displaced real-time) cases were used only for training on Monday each week.

2.3 Experiment Results and Feedback

Based on the pre-week surveys, a wide range of forecaster experience, expertise, and locations were included for the study. NWS Intern, Journeyman (or General), and Lead forecasters were all included in the experiment. At the time of the experiment, five forecasters had five years or less of experience in the NWS, five had between six and 10 years, three had 11 to 15 years of experience and five had 20 or more years of experience as an NWS forecaster. Forecasters were from all regions of the NWS, with all forecasters making at least one warning decision in the past year. All forecasters noted previous experience with cloud-to-ground (CG) lightning data from the National Lightning Detection Network (NLDN), with 83% frequently using it for any purpose, and 55% frequently using it in warning decisions. Most forecasters (13/18 or 72.2%) were familiar with the ENTLN, however, only 3/18 (or 16.6%) noted frequently using it for any purpose or within warning decisions prior to the experiment.[1]

Across all six scenarios, forecasters verification statistics fall in line with current NWS standards; overall skill scores had a probability of detection (POD) between 0.74 - 0.87 and a false alarm ratio (FAR) between 0.12 - 0.55 (Fig. 3). Initial results show that all three tiers of forecasters performed similarly (resampling provides a high-level of overlap between all three tiers), with forecasters in Tier 2 (total lightning + radar) performing slightly better in terms of overall FAR and POD for all events (Fig. 4). As shown in the bottom panels of Fig. 4, the additional data had the biggest impact on the poorest performing forecasters. Whereas all three data tiers of the top six forecasters perform with roughly the same skill (CSI scores between 0.47 - 0.50), the bottom six show more range with best performing group having access to both total lightning and radar data (radar only CSI = 0.32; radar and total lightning CSI = 0.46). When forecasters in the lowest group had access to total lightning data in addition to radar, the skill scores became similar to that of top six forecasters.

The Hart and Staveland (1988) NASA Task Load Index (NASA-TLX) was applied to the

---

[1] ENTLN data has been more recently introduced to the NWS operational environment. At the time of the experiment in 2014, the ENTLN data was not yet available within the NWS AWIPS2 operational system though it was available to forecasters through the StreamerRT web-based application. This likely contributed to the lack of familiarity and use compared to other lightning networks.

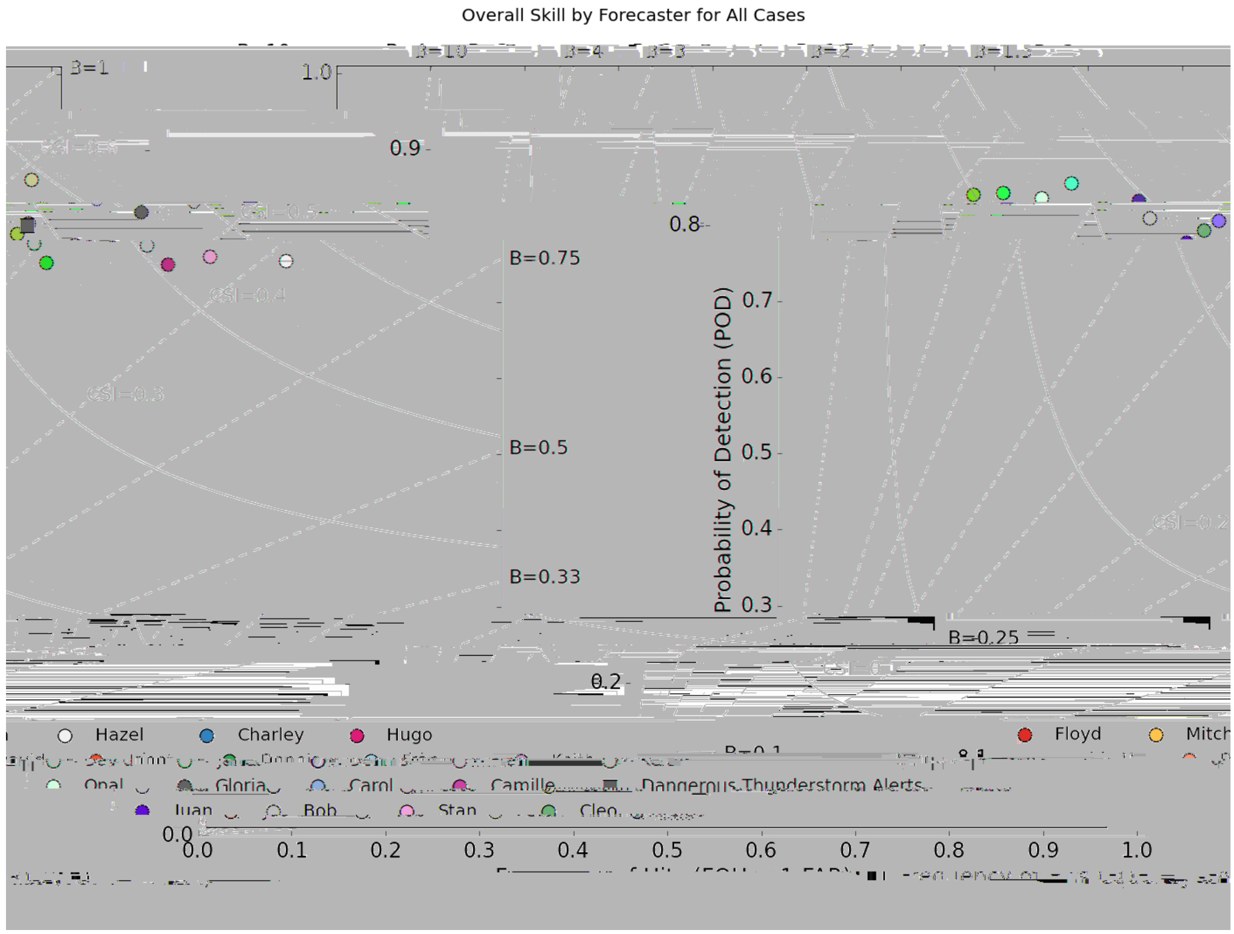


Figure 3. Performance diagram for each of the individual forecasters (circles) and the dangerous thunderstorm alerts (square) across the six scenarios. POD on y-axis and Frequency of hits (1-FAR) on x-axis.

study to judge how forecaster demand and stress varied depending on what tier of data was available. As part of the NASA-TLX analysis, forecasters rated their mental, physical, and temporal demands as well as their performance, effort, and frustration after every scenario using Likert-type scales (i.e. Strongly Agree, Agree, etc.). Forecasters were then required to quickly choose which source of workload was more important to that particular event (e.g., physical demand or performance). The patterns of forecaster choices were combined to create a summary workload score for each forecaster, for every event at each available data tier. This provided a scale such that the results vary less from person-to-person compared with unidimensional workload ratings (Hart 2006).

Tables 3 and 4 provide NASA-TLX scores by event and by data tier, respectfully. The most demanding case for all the forecasters to work was the Grand Junction, CO (GJT) event;

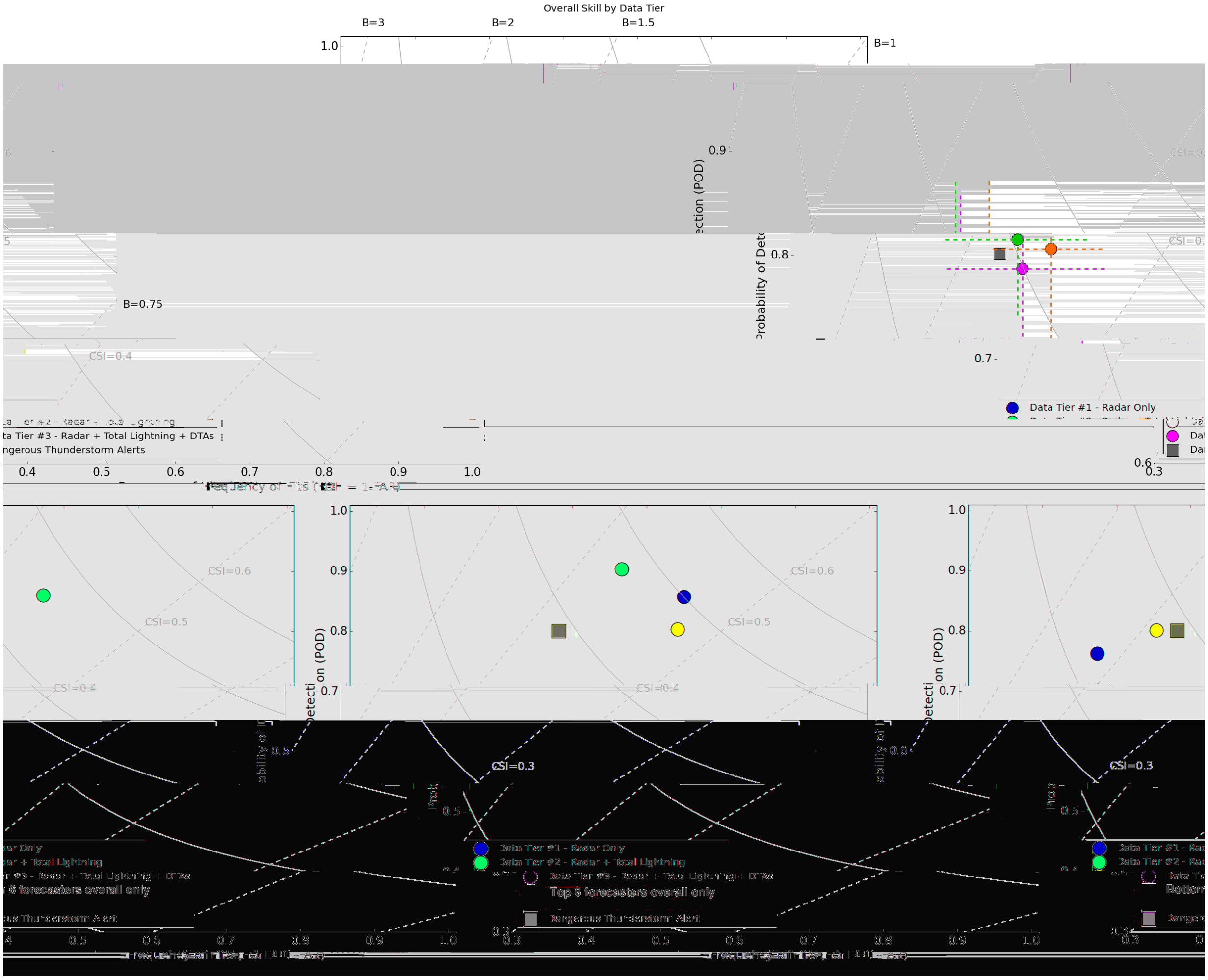


Figure 4. Performance diagrams across all six scenarios, separated by data tier (circles) and Dangerous Thunderstorm Alerts (squares). Top panel includes all forecasters, dotted lines indicate the range of the 95% confidence interval from bootstrap resampling. Bottom left, verification scores for bottom six performing forecasters. Bottom right, verification scores for top six performing forecasters.

this case had forecasters working a number of severe and sub-severe storms in a region of poor radar coverage. The easiest case determined by the forecaster workload was the Dallas-Ft. Worth (FWD) – a single non-tornadic supercell storm that produced multiple reports of large and damaging hail and severe wind; NASA-TLX scores were the lowest for this event for every forecaster. The other four cases clustered around similar values across participants, though the Sterling, VA (LWX) mesoscale convective system with an embedded EF1 tornado had the greatest variability from forecaster to forecaster. Meanwhile, little to no differences were seen on the forecaster workload when compared by data tier (i.e., radar only vs. the addition of total lightning data and/or DTA system tools; Table 4). Thus, the impact of the meteorological aspects of the cases had a higher impact on the overall forecaster workload than the addition of the 1-min total lightning data and/or total lightning-based decision-assistance tools.

	BMX	**FWD**	**GJT**	**GRR**	**LWX**	**PAH**
Mean	59	35	67	53	56	53
Median	60	32	70	53	57	55
Standard Deviation	11.5	12	11.3	10.7	14.1	12.7

Table 3. The mean, median, and standard deviation of the NASA-TLX scores by event according to the local NWS county warning area: Birmingham, AL (BMX), Dallas-Ft.Worth (FWD), Grand Junction, CO (GJT), Grand Rapids, MI (GRR), Sterling, VA (LWX), and Paducah, KY (PAH). The scenario over FWD was found to be the easiest by all forecasters while the GJT warning scenario provided the highest workload.

	Tier 1: Radar Only	**Tier 2: Radar and Total Ltg**	**Tier 3: Radar, Total Ltg, and DTAs**
Average Overall Task Load	55	52	54

Table 4. The mean task load per data tier for all forecasters at that tier for all warning scenarios.

The exit survey completed by each of the forecasters had them rank the various total lightning and DTA system products that they had access to during the week along with several radar-based products by importance to warning operations. As expected, forecasters ranked 0.5 deg reflectivity as the most important product on the list (with an average ranking of 1.22 out of 10 with 1 being the most important and 10 being the least important). This was followed by 1-min total lightning points (3.61), total lightning-based cell tracking products (i.e., cell-based flash rates and tracks, 4.22), 5-min total lightning points (5.16), Dangerous Thunderstorm Alerts (Level 3, 5.5) and Vertically-integrated Liquid (VIL, 5.56). The lowest four products ranked in order of importance were: Composite reflectivity (6.72), Significant Thunderstorm Alerts (Level 2, 7.11), 10-min total lightning points (7.33) and General Thunderstorm Alerts (Level 1, 8.67). Forecasters were also asked how often (i.e., Never, Rarely, Occasionally, or Frequently) they would use (a) the total lightning data and (b) the DTA level 3 alerts in both warning decisions and for any purpose. Forecasters saw themselves using total lightning data more often than the DTA alerts for both warning decisions and everyday utility. Ten forecasters (55.5%) noted they would use the total lightning data frequently for warning decisions with the number increasing to 14 (77.7%) frequently using the data for any purpose. Whereas, only 16-20% of the forecasters saw themselves using the DTAs frequently for any purpose or in warning decisions.

Discussion periods were hosted weekly to better understand how forecasters incorporated total lightning data and DTA system products within the warning-decision process across the various cases and environments. The discussion periods also addressed how forecasters could utilize total lightning data and derived products at their home offices. Forecasters envisioned the biggest uses of total lightning information to be for decision-support services and as supplemental data in regions of poor radar coverage. Others noted use in determining storm strength and trends, particular for new and developing convection. Additionally, multiple forecasters mentioned the increased situational awareness during cases where they had access to the total lightning information and foresaw that as a primary use within their own office. Overall, forecasters saw very little value in alert polygons themselves, primarily due to lack of trust in the product given an uncertain relationship to local weather across multiple geographic domains and what appeared to be a relatively high false alarm ratio during the multiple cases. However, as captured in the surveys above, some forecasters did see utility in this product as a situational awareness tool, in addition to the total lightning cell tracking and storm cell histories, to help key in on storms that might need more attention. Forecasters generally felt the alerts would have more utility if the threshold could be configurable to local weather based on local research.

3 Year Two Experiment

3.1 AWIPS2 Implementation

The 2015 HWT experiment built upon the initial evaluation in 2014. Improvements were made to the algorithms and display based on forecaster feedback of the product use and incorporation into the warning-decision process. In addition to the display capabilities tested by forecasters in year one, two major changes were implemented based on the feedback: (1) development and display of storm-based time series information with tracked storms and (2) user-defined thresholds for the alerts.

3.2 Experiment Design

The goal for the second year was primarily to evaluate the feasibility of use and performance of the system under the stress of real-time warning operations within the full suite of operational products. The annual spring experimental warning program (EWP) was utilized for this part of the evaluation (e.g., Calhoun et al. 2014). The real-time EWP utilizes an operational environment similar to a local NWS Forecast Office with the same software (AWIPS2) and data flow to test the latest concepts, products, and algorithms aimed at improving short-term forecasts and warnings of severe weather. In addition to the ENI products available in the year one experiment, the forecasters had access to time series and customizable thunderstorm alerts.

The 2015 EWP Spring Experiment ran for 5 weeks (4 May - 12 June 2015) with six forecasters participating each week (5 NWS meteorologists and one broadcast meteorologist, weekly). The operational domain was chosen daily; areas expected to receive severe weather were given priority. Each day, the forecasters were separated into pairs, with each pair given a different NWS County Warning Area (CWA). Unlike the year one experiment, forecasters had access to all data currently available in NWS operations as well as other experimental products being developed for operations such as numerical model and satellite-focused algorithms. However, to help refine the forecaster focus and not conflate the analysis with feedback from similar experimental products, four of the six forecasters were provided access to the ENI products and tools for warning operations and restricted from using other total lightning data or algorithms. So that all forecasters participating in the experience had a chance to provide feedback on all products and algorithms, the pairs rotated daily.

The dependence upon the live weather environment allowed for testing and examination of the products across much of the CONUS for a variety of severe and near-severe weather environments. During the five weeks of the experiment, we operated in 42 different NWS county warning areas with forecasters working warning environments consisting of multicell storms, supercell and tornadic storms and multiple mesoscale convective systems. For continuity and to provide additional product expertise, forecasters that participated in the 2014 HWT ENI evaluation were encouraged to participate again within the 2015 EWP. Each week at least one of the forecasters had experience from the previous year.

3.3 Experiment Results and Feedback

All forecasters that had access to the data used it throughout the operational shifts to varying degrees. At the start of each week, forecasters were encouraged, though not forced, to use all products. The ENI lightning data and DTA system decision-assistance products were used primarily for situational awareness, to pick out the strongest or most intense storms, and to monitor storm trends. Initially, many forecasters found the large selection of products and options to be burdensome and that trying to use them all at once cluttered the screen. However, by the end of the week forecasters generally gravitated to a few select products that worked for them within their unique workflow and storm-interrogation process.

As part of a daily survey, forecasters were asked specifically about using the total lightning data and ENI DTA system storm-based tools for situational awareness and within their warnings decisions. 76% of respondents said "yes" that total lightning influenced their situational awareness, but the number decreased to 50% when asked if the total lightning data affected the warning confidence or timeliness. Similarly, 66% of the forecasters found that the DTAs and/or other decision assistance tools including the storm-based flash rate and time series influenced their situational awareness, but the number decreased again to 44% who thought these same tools affected their warning confidence or timeliness.

Additionally, the daily survey also asked the forecasters how useful they found each product in operations during their shift that day, ranking each either: Not at all useful (1), Marginally Useful (2), Somewhat Useful (3), Very Useful (4), Extremely Useful (5), or Not Applicable (no value). Ranked averages were calculated by assigning a value of 1 assigned to "Not at all useful" and increasing the value by one with each category to 5 for "Extremely Useful." The DTA system storm-based time series plots were considered the most useful product (a ranked average of 3.89) with 31% of responses finding them "extremely useful" and greater than 50% finding them at least very useful. The next most-used products included the raw 1 min total lightning points (ranked average of 3.63; 38% very or extremely useful) and current storm-based flash rate and past tracks (ranked average of 3.49; 49% very or extremely useful). The alerts were found less useful in daily operations (3.33, 2.66, and 2.06 ranked averages for the dangerous, significant, and general thunderstorm alerts, respectively), but some forecasters did use the top-level DTA for situational awareness to highlight stronger storms in a busy environment or to provide extra confidence in their warning decision. Many forecasters were quick to point out in the blogs, discussion and surveys that using all the alerts simultaneously or even simply the top-level DTAs alone cluttered the screen. Forecasters additionally noted the alerts were hard to follow, and were not a good indicator of severe thunderstorms. While some forecasters tried to incorporate the DTAs in the warning-decision process throughout the week, for the reasons listed above the many forecasters depended on the alerts less often as the week progressed.

While the configurable alerts were made an option to forecasters during the real time experiment, most forecasters (90%) did not try setting a threshold as either they found it inconvenient to do so, lacked interest, or lacked knowledge of what threshold to set instead of the defaults. However, through additional training and development of regional (i.e., office-level) knowledge and expertise, forecasters still felt strongly this would be an appropriate use of the alerting system. Multiple forecasters noted they wanted to keep this option to customize the tool to provide guidance in context to individual geographic areas if the DTA system were to transition to operations.

In regards to the total lightning data alone, most forecasters noted that it did provide "a quick reference for intensification" and was "consistent with radar data." The 1-min product was found slightly more useful than the 5-min product. However, personal forecaster preference and time-matching with radar were the primary reasons for choosing one time over the other, as the 1-min and 5-min data were interpreted similarly by forecasters. When the daily operations were located in the western US, forecasters noted in blog posts and discussions that they were less enthusiastic about the total lightning data as the storms had lower apparent flash rates than storms of similar radar appearance and strength in other regions of the country, potentially due to lower detection efficiency in that region.

An exit survey was given to the forecasters at the end of every week. Similar to year one, this survey asked the forecasters to rank in order of importance to warning operations the various experimental products they viewed during the week within many that are commonly used in warning operations (Table 5). Again, as expected and seen in the year one survey, 0.5 degree reflectivity from the nearest radar was ranked the highest. The top ranked lightning products were the cell identification and tracking with time series followed by the 1-min total lightning points and 5-min gridded total lightning data.

Product	**Overall Rank**
0.5 deg Reflectivity (nearest radar)	1.74
ProbSevere (Experimental Storm-based Probability of Severe Weather)	3.03
Earth Networks Cell Identification and Tracking with Storm Flash Rate	5.19
MRMS Maximum Expected Size of Hail (MESH)	5.19
1 minute total lightning (raw points, ENTLN)	5.48
Composite Reflectivity	6.39
5 minute total lightning (gridded, ENTLN)	6.58
Dangerous Thunderstorm Alerts (DTA, Purple Polygon, Level 3)	7.16
Vertically Integrated Liquid (VIL, D-VIL)	7.55
GOESR Overshooting tops algorithm	9.23
Significant Thunderstorm Alert (Orange Polygon, Level 2)	9.32
Thunderstorm Alerts (Green Polygon, Level 1, 3 flash min)	11.13

Table 5. The overall ranking of products in order of importance (with 1 being the highest or most important and 12 being the lowest) to warning operations by participants in the 2015 HWT.

The forecasters were also asked as part of the exit survey about expected frequency of use within their home offices of the total lightning data and the top-level DTAs for warning operations if they were to be made available nationwide. 71% (22 of 31) of forecasters said they would frequently use the total lightning data within warning decisions and the remaining 29% (9 of 31) believed they would use it at least occasionally. Fewer forecasters thought they would use the DTAs: 23% (7) would use them frequently, 55% (17) occasionally, 16% (5) rarely, and 6% (2) would never use them. As a follow-up to this question in the survey, forecasters were asked to explain why they chose their answer. Forecasters that would use DTAs frequently or occasionally noted that the presence of a DTA would be used primarily to build confidence on a warning decision based on radar data, for situational awareness to focus attention to the stronger or most significant storms, in regions of poor radar coverage, and for highlighting deviant storm motion. Forecasters that would rarely or never use the DTAs preferred to use base data and other tools such as the time series information for warning decisions, noted that the correlation of the product to storm severity was unproven, and/or that they did not trust the product.

Additionally, as part of the exit survey, forecasters were asked how comfortable they were including total lightning data in the storm interrogation process following a week of use in the HWT. The majority of forecasters felt much more comfortable after using the data than they were initially. Specifically forecasters stated they were now able to “see the utility in it”, “used it a lot more than first thought”, “have a better understanding on how to use these products”, and “now know how it can be used and it has the potential to increase lead time.” Additionally, multiple forecasters commented how total lightning data related to storm strength as it showed high value in helping to identify storms that are (re)intensifying. Forecasters also expressed an appreciation for how total lighting data and derivative tracking tools could be helpful in sparse radar coverage areas stating: “In areas that are poorly sampled by radar data…such as mountainous and farther offshore marine locales…the ENI data should prove to be very valuable. The ENI data could be the confidence factor in issuing or not issuing a special marine warning on a storm that looks to be marginally strong due to poor radar sampling.”

However, forecasters still want more training on what lightning flash rates and trends mean for storms in multiple storm environments (e.g., supercell storms compared to quasi-linear convective systems or tropical rainbands) and limitations of the data (e.g., due to detection efficiency or regional environmental controls). Multiple forecasters also mentioned problems that they thought would interfere with routine use within NWS operations, such as issues with storm tracking (storm splits and mergers), and added that the algorithms must be improved before they are implemented operationally. Finally, forecasters also again shared concerns regarding possible issues with the data in the western region and confusion along the ENI-defined east/west 104°W longitude border for cell tracking flash rate thresholds (see Table 1) following operations in this region.

3.4 Examples Within Operational Use

How forecasters integrated the lightning data and decision-assistance tools varied day-to-day across forecasters and weather scenarios, though similar product-use and storm-interrogation methodologies were seen. The following events are included to provide the forecaster perspective, provide details on how lightning data is used within operations, and highlight some difficulties that forecasters noted in real-time.

3.4.1 Midland and San Angelo, TX: 13-14 May 2015

For two days in a row we positioned forecasters over the Midland, TX (MAF) and San Angelo, TX (SJT) forecast offices to monitor possible storm development along the dryline and further east within the warm sector. Of the two days, the likelihood for severe storms was lower on 13 May with the region under a “marginal” risk for severe storms from the NOAA Storm Prediction Center (SPC). While expectations of severe weather were low, lift along the dryline combined with low-level moisture and surface heating were expected to support widely scattered thunderstorms with the possibility of severe hail and damaging winds. On 14 May, there was additional upper-level support for thunderstorm development as a weak mid-level short wave trough of low pressure moved through the area; accordingly, the area was increased to a “slight” risk from SPC. In addition to the similar low-level conditions as the 13th, there was an increase of CAPE due to more favorable mid-level lapse rates and shear was generally expected to increase throughout the day. Large hail and locally damaging winds were again the primary threat on 14 May.

Early in the shift on the 13th, with little activity over the Midland area, forecasters focused on the relationship between the $-20°$C reflectivity and total lightning rates over southern Texas (Fig. 5). As forecaster “PR” noted in an associated blog post: “There appears to be nice correlation between lightning and reflectivity aloft (which makes physical sense)...A ramp up in lightning could be indicative of larger reflectivity aloft, which would suggest an intensifying updraft.” This early evidence of the physical relationship gave the forecasters additional confidence in using the data throughout the rest of the

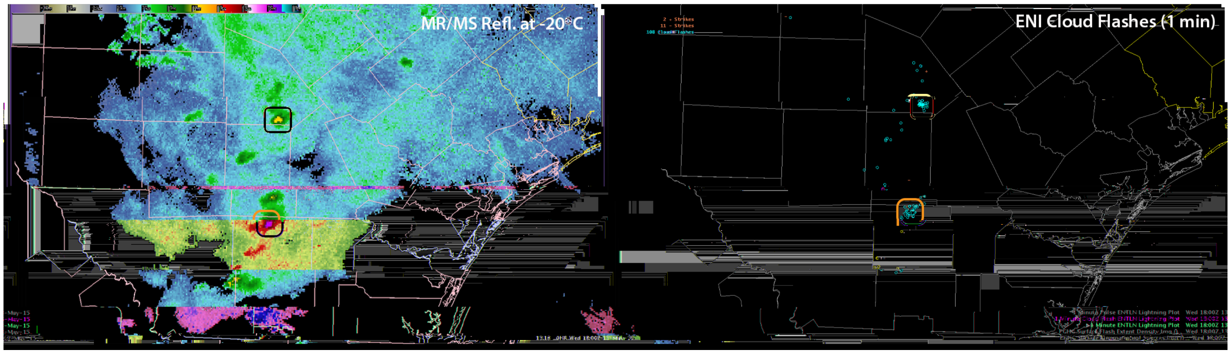


Figure 5. Screenshots from forecaster display showing the relationship between higher reflectivity at $-20°$C and presence of lightning at 1800 UTC, same locations are highlighted in the cream and orange circles on both images. Left: MRMS Reflectivity at $-20°$C. Right: 1 min ENI cloud flashes (blue circles).

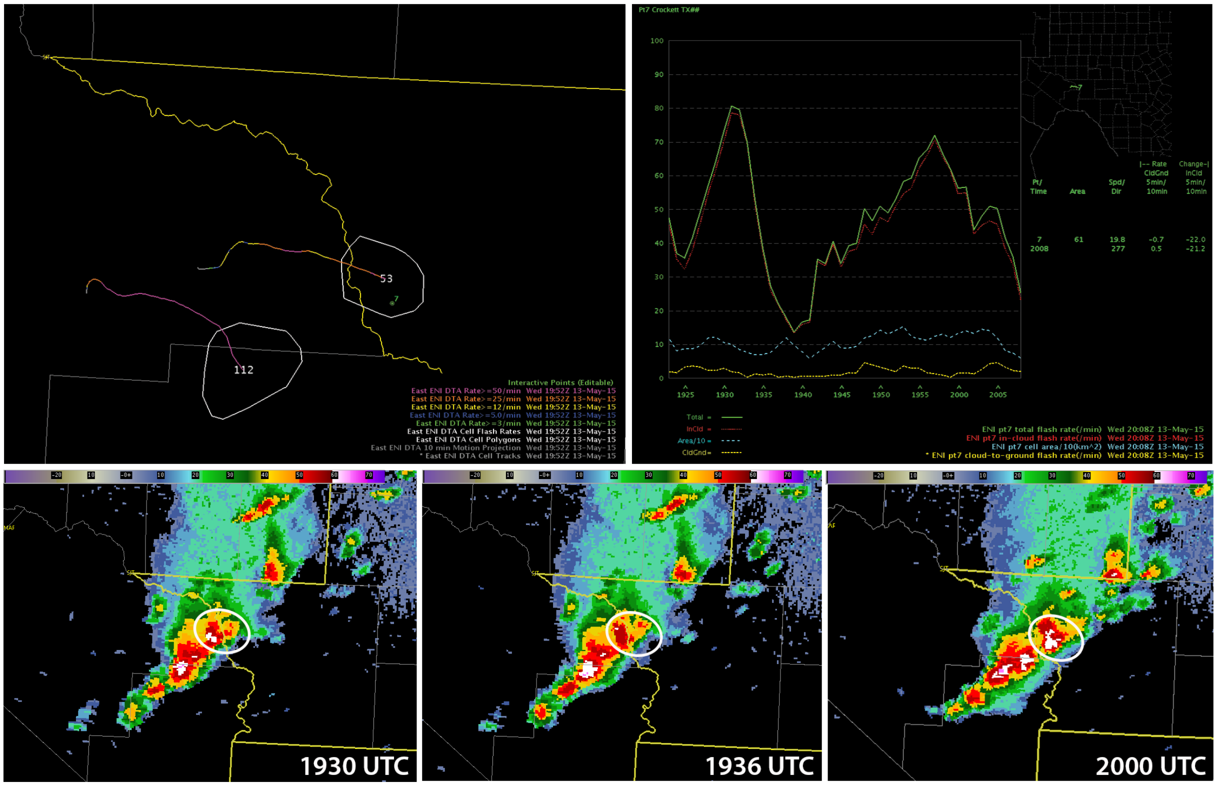
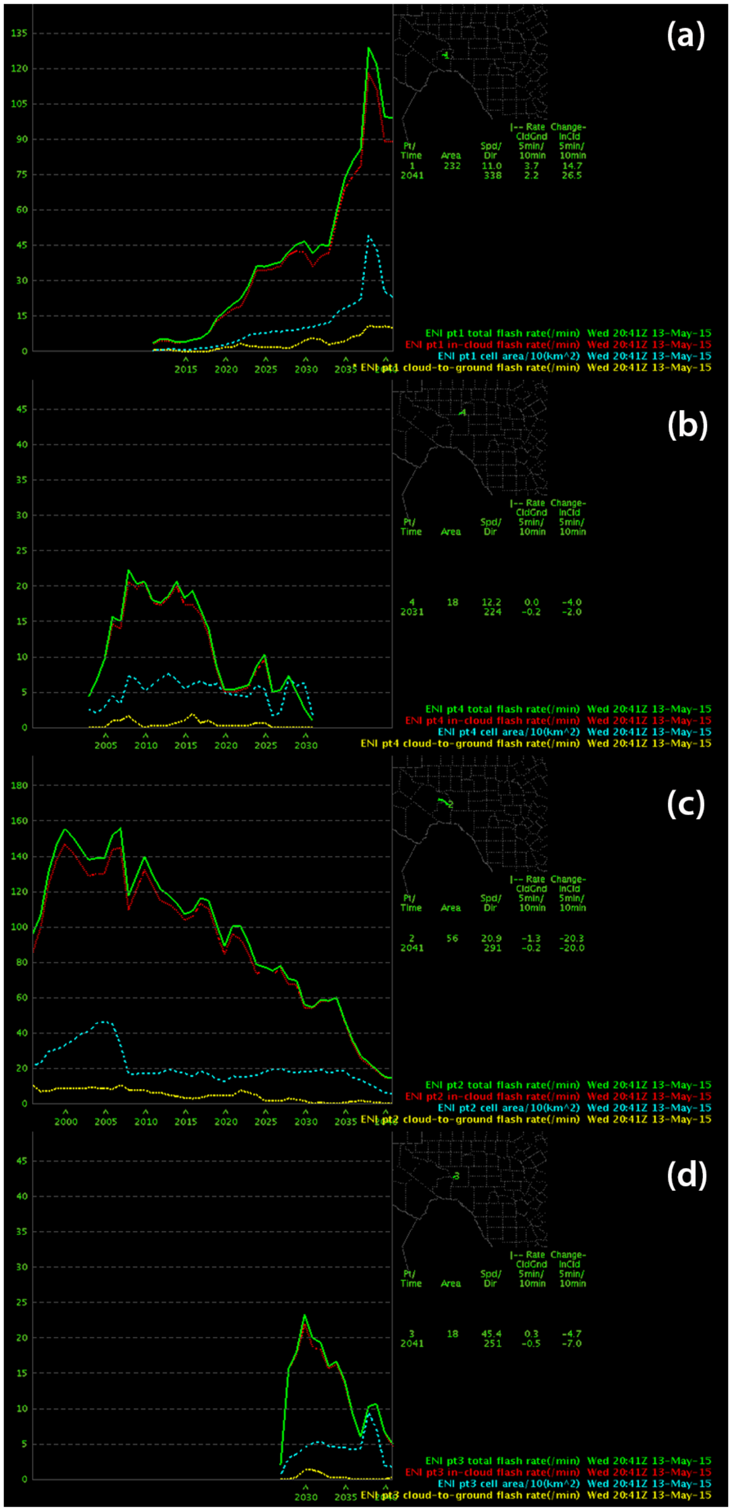


Figure 6. Series of screenshots from forecaster "Regina" between 1930 and 2000 UTC depicting the correlation between peak storm reflectivity and lightning flash rates as seen in the time series for the tracked lightning cell. Top left: ENI tracked lightning storm polygons, current flash rate and path history at 1952 UTC. Top right: Time series data from 1920-2007 for northern tracked lightning cell. Bottom row: series of composite reflectivity at 1930, 1936, and 2000 UTC highlighting changes in peak reflectivity for northern tracked cell (denoted by white circle).

shift.

From 1930 - 2015 UTC on 13 May, forecaster "Regina" was monitoring storms on the border of the MAF and SJT offices. She became focused on how trends in the total lightning data seemed to always precede radar trends in part "since the lightning data comes in more often than/ahead of the radar imagery." She noted that "the time series plots could be used to anticipate increases or decreases in storm cell strength, potentially aiding in the warning process (i.e., whether to issue (lead time), continue, or let a warning expire)." A series of images from her screen show this relationship well for both storm intensification and decrease (Fig. 6). Later in the shift, she decided to continue on a warning on one storm and expire another warning for a nearby storm partially due to lightning trends on each.

As the forecasters monitored the storms in the region, they continued to note that rapid



```
                      |-- Rate  Change-
                       CldGnd   InCld
 Pt/            Spd/   5min/    5min/
 Time    Area   Dir    10min    10min
   1      232   11.0    3.7      14.7
 2041            338    2.2      26.5
```

```
                      |-- Rate  Change-
                       CldGnd   InCld
 Pt/            Spd/   5min/    5min/
 Time    Area   Dir    10min    10min


   4       18   12.2    0.0      -4.0
 2031            224   -0.2      -2.0
```

```
                      |-- Rate  Change-
                       CldGnd   InCld
 Pt/            Spd/   5min/    5min/
 Time    Area   Dir    10min    10min

   2       56   20.9   -1.3     -20.3
 2041            291   -0.2     -20.0
```

```
                      |-- Rate  Change-
                       CldGnd   InCld
 Pt/            Spd/   5min/    5min/
 Time    Area   Dir    10min    10min


   3       18   45.4    0.3      -4.7
 2041            251   -0.5      -7.0
```

Figure 7. Forecaster screenshot of four-panel time series display (shown as a column here for magnification) of four different storms in southwest Texas covering the period 2000 - 2040 UTC. Includes: Total flash rate (IC+CG, green), IC flash rate (red), cell area/10 (km2, blue), and CG flash rate (yellow). Panel (a) highlights the importance of noting whether a flash rate increase is tied to cell merger and/or increased storm size.

increases in the flash rate appeared to be well correlated to storm intensification. Additionally, the forecasters found it easy to diagnose this rapid flash rate increase using the time series panels (Fig. 7). However, they also saw rapid increases in the flash rate due to cell mergers. While increases in the cell area trend on the times series plots could be used to diagnose that the increase was not related to storm intensification (Fig. 7a), the forecaster later stressed that deeper investigation of flash rate increases should be an important aspect of training and best practices on using the data.

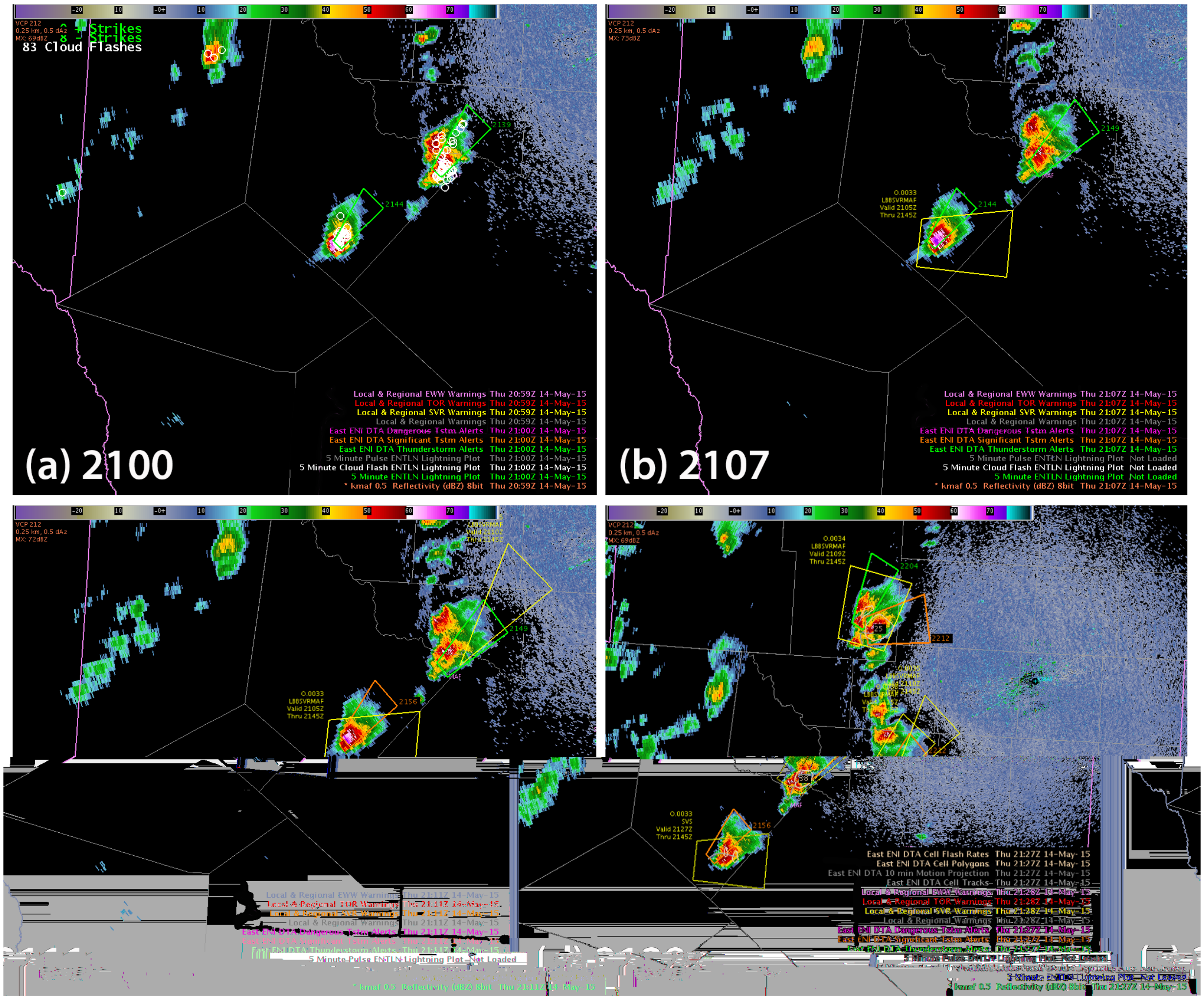


Figure 8. Series of forecaster screenshots from 2100 - 2128 UTC on 14 May 2015 highlighting the lack of a dangerous thunderstorm alert on a storm with a radar-evident three-body scatter spike. (a) 2100 UTC first appearance of the ENI general thunderstorm alert (green polygon). (b) 2107 UTC forecaster issues severe thunderstorm warning (yellow polygon), still no Significant or Dangerous Thunderstorm Alert on storm. (c) 2111 UTC, ENI Significant Thunderstorm Alert appears (orange polygon). (d) Overview of storms in the region, including three storms with severe thunderstorm warnings and Significant Thunderstorm Alerts, but no Dangerous Thunderstorm Alerts.

On 14 May, convection initiation occurred earlier, between 1800 - 1900 UTC, and farther southwest than on the 13th. The first few storm cells quickly reached 40-50 dBZ in the western part of the Midland CWA prior to 1930. By 2130, the forecasters had issued three different severe thunderstorm warnings, but had difficulty relying on the flash rates and ENI top-level thunderstorm DTAs. Multiple storms in the area had three-body scatter spikes apparent in the KMAF WSR-88D radar reflectivity, but did not have either significant thunderstorm or dangerous thunderstorm alerts present (Fig. 8). Forecasters were unsure if this was due to the storms simply having lower flash rates than expected or related to detection efficiency in the area. They stressed in both discussion and a blog post that this is "why DTAs are not the "best" thing to wait for or look at for a severe thunderstorm."

Later, as storms became more clustered as opposed to isolated, the forecasters found the tracking algorithm frustrating to use as it moved from multiple cells to cells within cells to one large cell and back in the period of 10 min (Fig. 9). Additionally, during these periods where one cell was tracked within another cell the time series information could not decipher the different cells and would simply provide details regarding one cell. The forecasters were left guessing at this point at what region to attribute the time series properties.

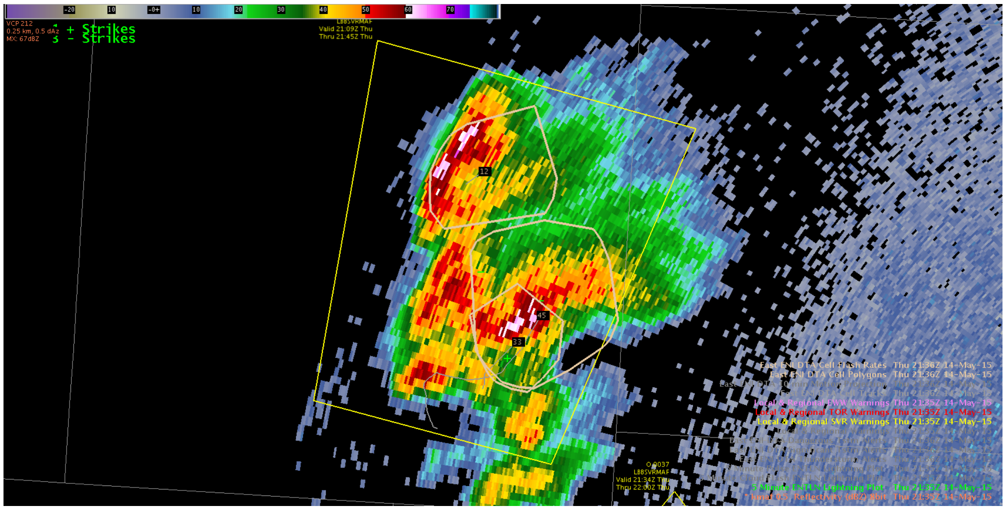


Figure 9. Forecaster screenshot at 2139 UTC depicting multiple tracked storms and polygons within one severe thunderstorm warning. Forecaster noted how cell polygons changed across multiple time steps during this period making it difficult to follow storm-based trends. Additionally, the forecaster found it unclear as to how and where to attribute flash rates when tracking included concentric circles as shown above.

At the end of the day, forecasters voiced disappointment and frustration with the ENI system over the region compared to previous days. Specifically they noted: "When using ENI data, forecasters must have a good handle on the "typical" amount of lightning displayed in "average" storms. There have been numerous severe storms in MAF CWA, but they have been low on lightning strikes." Additionally, forecasters found the delineation of flash rate thresholds at 104°W longitude between the east and west regions confusing (Fig. 10), stating: "Working in west TX, it was a nightmare trying to use the lightning data because the line cut through 1/3 of the CWA...A better boundary has to be chosen thats more friendly for these CWAs."

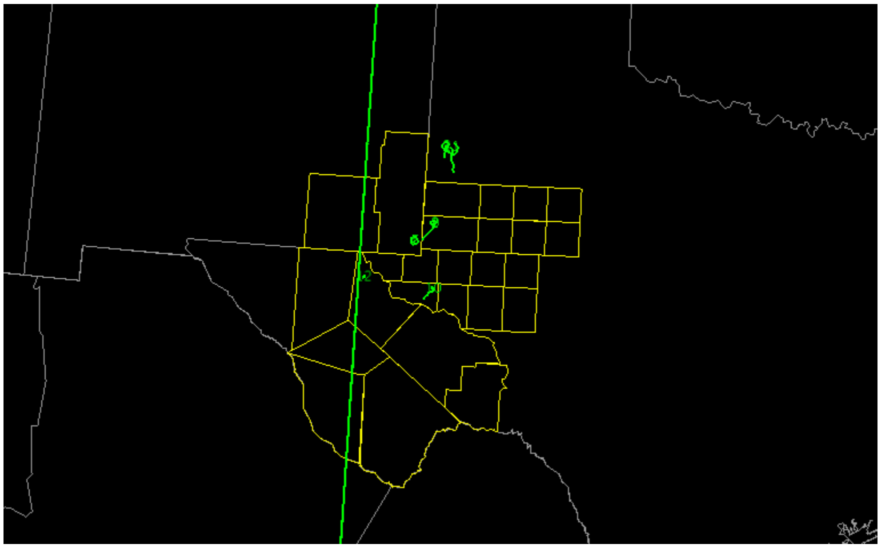


Figure 10. ENI east-west boundary at 104°W longitude (green line) and MAF county warning area (yellow). Multiple forecasters were unsure how to handle the different files and alert levels.

3.4.2 Wilmington, NC: 21 May 2015

The shift for Wilmington, NC (ILM) forecasters "Rocky" and "Holaday" began around 18 UTC with disorganized clusters of convection already ongoing in central NC around Raleigh. SPC had this region under a "slight risk" of severe storms with strong winds expected to be the primary hazard as an upper-level short wave trough of low pressure and associated surface low moved across the region. We anticipated additional development within this warm sector over the ILM region as the day progressed.

Throughout the event the forecasters monitored a number of borderline severe storm clusters with the strongest storms primarily along the cold front. However, many small multicell storms also developed across the warm sector as there was little to no convective inhibition. Initially, lightning rates remained low for these storms – from 10-13

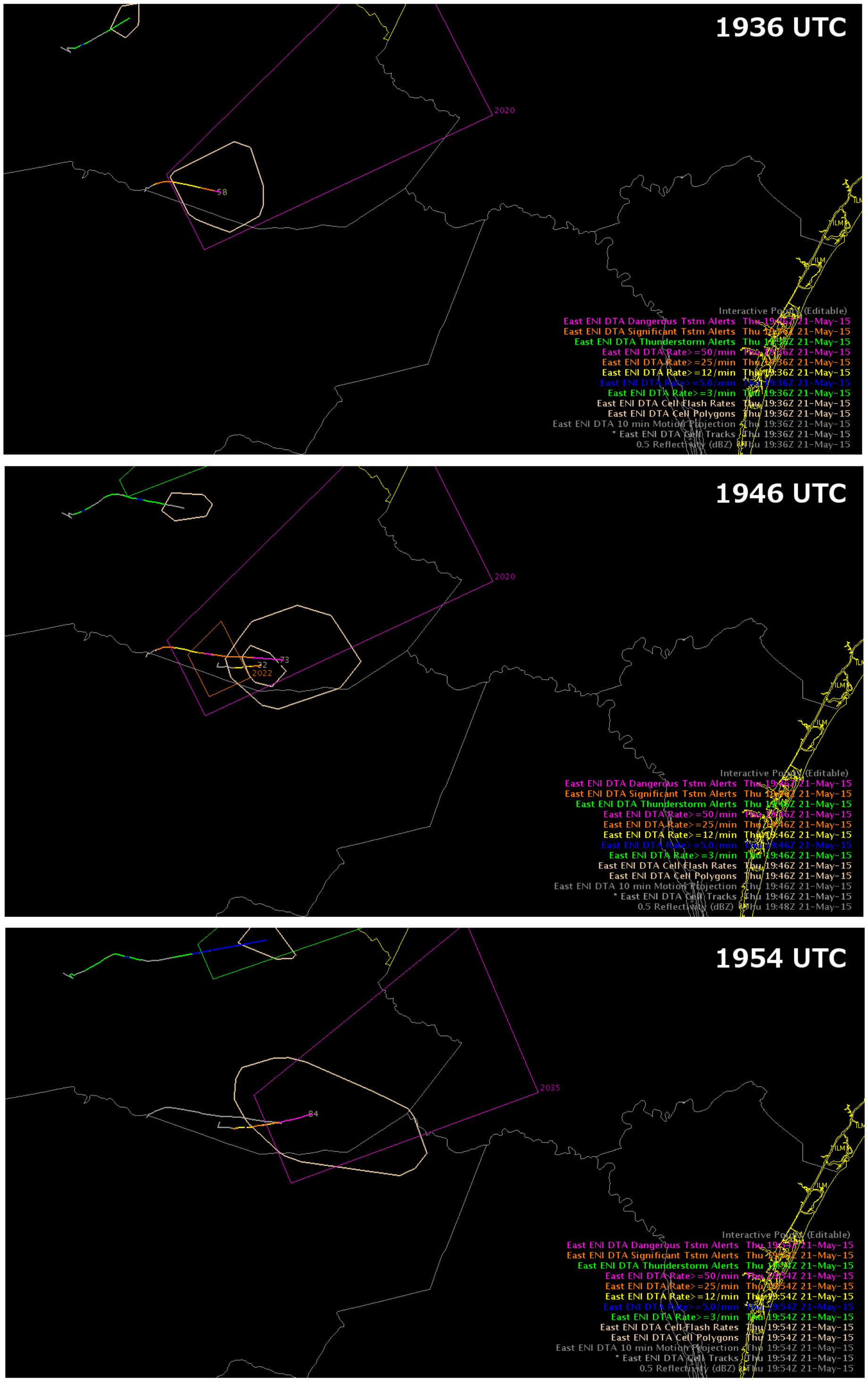


Figure 11. Screenshots from forecaster “Holaday” at 1936, 1946, and 1954 UTC on 21 May 2015 in Wilmington, NC CWA. ENI-based cell polygons, flash rates, past track, and all three layers of alerts polygons. Forecaster highlighted confusion with the merger process at 1946 UTC and visual of a cell within another storm cell.

flashes per min to a maximum of 30 flashes per min. Though there were small increases in the flash rate for a few storms during this time, combined with "unremarkable" reflectivity data, the forecasters opted to issue no warnings during the first hour of the shift.

At 1930, Holaday issued a severe warning for hail, but the ENI tracking algorithm was having difficulty continuously tracking the storm clusters. An example of these tracking errors is depicted in image from Holaday's screen when an area of smaller multicell storm merged with a larger severe storm and produced one cell within a larger one (Fig. 11). Similar to the event on 14 May in south Texas, this resulted in errors with the flash rate trends as it was unclear how the flashes were associated to the clusters and unknown if the same flashes were being counted by both clusters within the algorithm. This type of error occurred at least six different times across an hour of tracking the storm.

However, the forecasters still used the total lightning flash rates and flash rate trends in combination with radar data for severe and eventually tornado warnings at other points during the shift. A severe thunderstorm warning was issued at 1950 UTC for Bladen

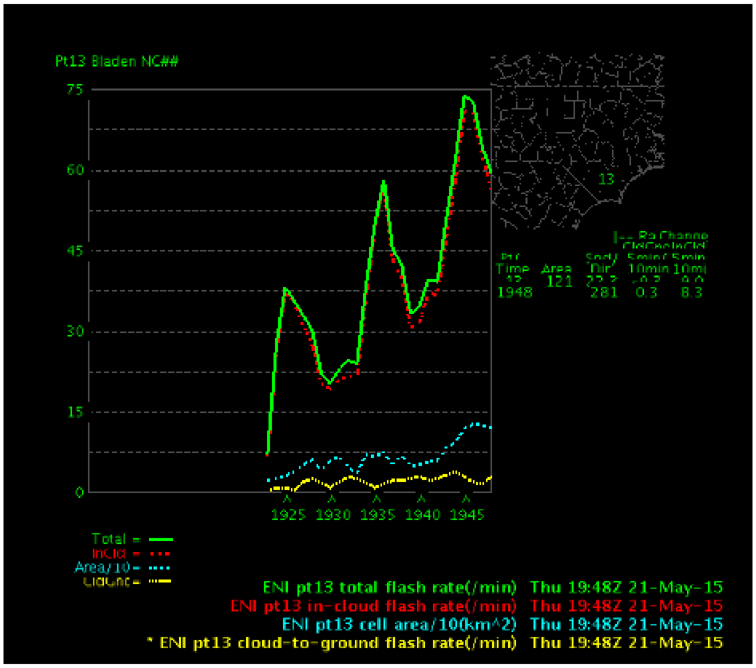


Figure 12. Time series for storm in Bladen County, NC between 1920 - 1948 UTC on 21 May 2015. Forecaster observed multiple jumps in the flash rate which, combined with radar data, helped influence the severe thunderstorm warning on the storm.

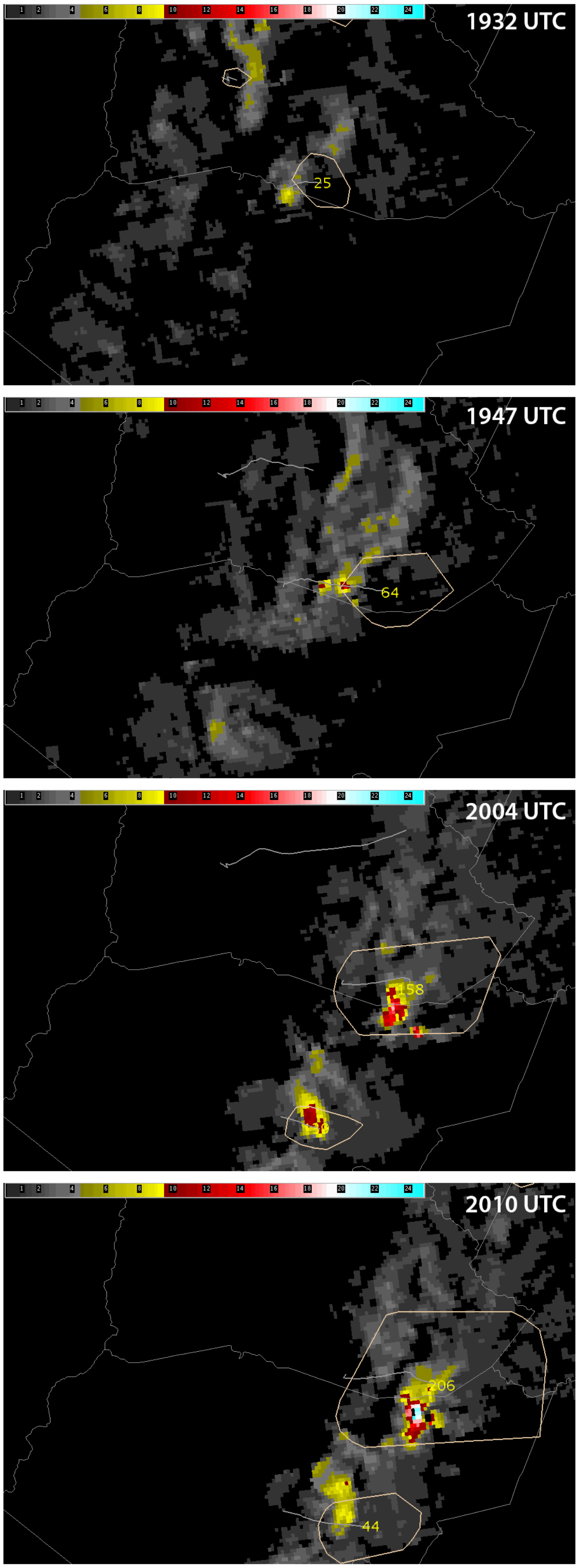


Figure 13. Gridded MRMS low-level (0-2 km AGL) azimuthal shear, ENI cell polygons (tan), cell history track (gray) and 1-min storm-based flash rates (yellow) at 1932, 1947, 2004, and 2010 UTC on 21 May 2015.

County, NC after a series of jumps in the lightning flash rates from the 10-30 per min to near 60 per min (first bringing the forecasters attention to this particular cell) and again to greater than 70 per min (Fig. 12) followed by local WSR 88D reflectivity data showing 70 dBZ at -20 C (25kft). A tornado warning was later issued by "Holaday" on a storm near Columbus County, NC just after 2000 UTC. Radar-indicated rotation was the primary reason for the warning issuance, but deeper inspection by forecaster "Rocky" of the low-level (0-2 km) Multi-Radar Multi-Sensor (MRMS) azimuthal shear depicted the total lightning flash rate increase as a precursor to the low-level rotation increase (Fig. 13) and EF1 tornado which began at 2013 UTC (*StormData*).

4 Verification of the Dangerous Thunderstorm Alerts and NWS Warnings

With interest in using the DTAs within severe storm interrogation, it is important we understand how well the top-level DTAs themselves perform as compared with current NWS severe thunderstorm and tornado warnings. The thunderstorm alerts have been publicized to provide improvements in lead times for various thunderstorm hazards; an ENI press release stated "Dangerous Thunderstorm Alerts (DTAs) improved median lead times by 50%, or an additional 9 minutes, over the 18 minute lead time afforded by NWS Warnings." In determining these metrics, ENI used a grid-based approach due to the perceived nuances of the automated DTA system and availability of storm report information.[2] This study employed the traditional methodologies, following NWS Verification Directives detailed below, to compile verification statistics of ENI top-level DTAs, NWS severe and tornado warnings. *NOAA StormData* reports from January 2013 through September 2015 were used for verification to determine probability of detection (POD), false alarm ratio (FAR), and lead times.

To complete the verification statistics, we followed section 2.1 from NWS Verification Procedure Manual for storm-based warning verification, specifically, event specific ver-

---

[2] ENIs methodology for calculating lead times includes taking into account the more frequent issuances of DTAs while following a specific storm cell. In such instances, subsequently issued DTAs automatically replace prior existing DTAs and often cover much of the same area as the previous. As such, multiple successive DTAs often intersect with an eventual storm report location. The DTA lead time is calculated using the issuance time of the first DTA to cover the storm report location and the event time. Following this methodology, an event lead time can be greater than the 45 minute DTA valid time. In this manner, extended lead times are achieved.

ification (Table 6). The event specific verification is described as each warning type, severe thunderstorm and tornado, can only be verified by the corresponding event type, hail or thunderstorm wind and tornado respectively. Therefore, a severe thunderstorm warning can only be verified by a hail or thunderstorm wind report meeting NWS warning criteria (i.e., 1 inch and 58 mph) and a tornado warning can only be verified by a tornado.

Warning Type	**Event Specific Verification** (Each warning type is only verified by the corresponding event type.)
Severe Thunderstorm	Non-tornadic severe thunderstorm hail (i.e., 1 inch or greater) or wind (i.e., 58 mph or greater).
Tornado	Tornado only.

Table 6. Warning type and event specific verification details; adapted from the NWS "Verification Procedure Manual for Storm-based Warning Verification."

Event	**Total Number**
NWS Severe Thunderstorm Warnings	33,049
NWS Tornado Warnings	3,681
ENI Dangerous Thunderstorm Alerts	109,116
StormData Severe Reports	41,912
StormData Tornado Reports	2,059

Table 7. Total number of NWS severe and tornado warnings, ENI DTAs, and *StormData* events from January 2013 through September 2015 included in the verification study.

There are a several caveats relative to using NWS verification methods with the top-level DTAs that may affect the verification statistics. DTAs are automatically updated every 15 minutes and each is valid for 45 minutes, while the NWS warnings can vary in length. This can result in significantly more individual DTAs to verify for the same storm and thus may contribute to higher FARs. Second, the DTAs are issued for multiple storm threats (tornado, wind, and hail), whereas the NWS has specific warnings for the storm threats. In order to apply the NWS verification methodology to ENIs DTAs, they were verified three different ways: only severe events, only tornado events, and severe and tornado events combined. A third caveat is the lack of geographic boundaries. The only boundary of consideration for the ENI data is the east/west boundary at $104°$W, which has different thresholds for when the DTA gets issued, however the DTAs can still overlap this boundary. The NWS has to consider county warning area boundaries; this can affect the size of a polygon as well as the duration of a warning. Fourth, because there are no geographic boundary limitations, the ENI DTAs can also be issued and cover ocean and lake regions. Since there are no reports in these locations, these DTAs were not included in this verification study. A shapefile of the continental United States was used to keep only DTAs that occurred within the CONUS and any that overlapped the coast. After this methodology was applied, the study included 109,116 DTAs and 36,730 NWS warnings for verification from 1 January 2013 through 30 September 2015 (Table 7). A fifth and final caveat is that since *StormData* is primarily dependent upon human reporting, there tends to be a relative dearth of reports in regions with sparse populations. Since DTAs are automated and continuously issued regardless of population, this may contribute to lowering their performance metrics.

From here POD and FAR were calculated using the following equations:

$$POD = \frac{A}{A+C} \tag{1}$$

$$FAR = \frac{B}{A+B}. \tag{2}$$

The above variables A, B, and C were found using a contingency table (Fig. 14). Point A is considered a hit because an observed event fell temporally and spatially inside a forecasted event. Object B is considered a false alarm because an event was forecasted but no event was observed. Point C is a miss because an event was observed but there was no event forecasted.

2x2 Contingency Table		Event Observed	
		Yes	No
Event Forecasted	Yes	A	B
	No	C	D

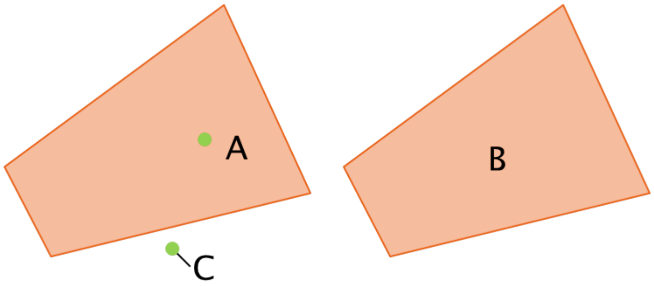


Figure 14. Contingency table used to calculate POD and FAR (left). Example storm-based warning or DTA polygons (orange) and reports (green dots) (right). Points, A, B, and C represent a Hit, False Alarm, and Miss, repectfully.

Lead times for each of the observed events were calculated, again based off NWS Verification Directives using one of two methods. The first method is for an instantaneous event, a single location. An example is an isolated thunderstorm wind report at a single time. If the event fell inside of a warning, both spatially and temporally, the lead time is the event start time warning start time (point A in Fig. 15). The second method is when an event starts at one location and ends at another location over a period of time (track event). An example of a track event is a tornado that started in one location and moved to another. Two assumptions need to be made for a track event: the event travels in a straight path between the event start/end locations and the event travels at a constant speed between the event start/end locations. Once these assumptions are made, the location of the event is estimated every minute for the duration of the event and a lead time is evaluated at each of those time steps. For example if an event lasted from 12:30 to 12:34, the lead time would be evaluated at 5 locations and times (Points $B_1 - B_5$ in Fig. 15). The lead time of the first point is the initial lead time of the event, while the average of the lead times at each time step is the mean lead time. This study was only concerned with the lead time at the beginning of the event so the initial lead time was used. If there is no warning at the report time, the lead time is 0 (point C in Fig. 15).

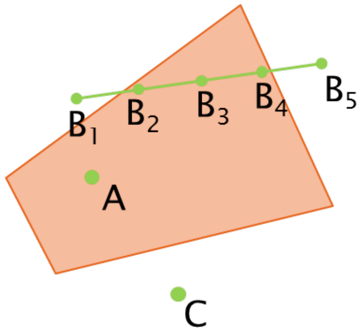
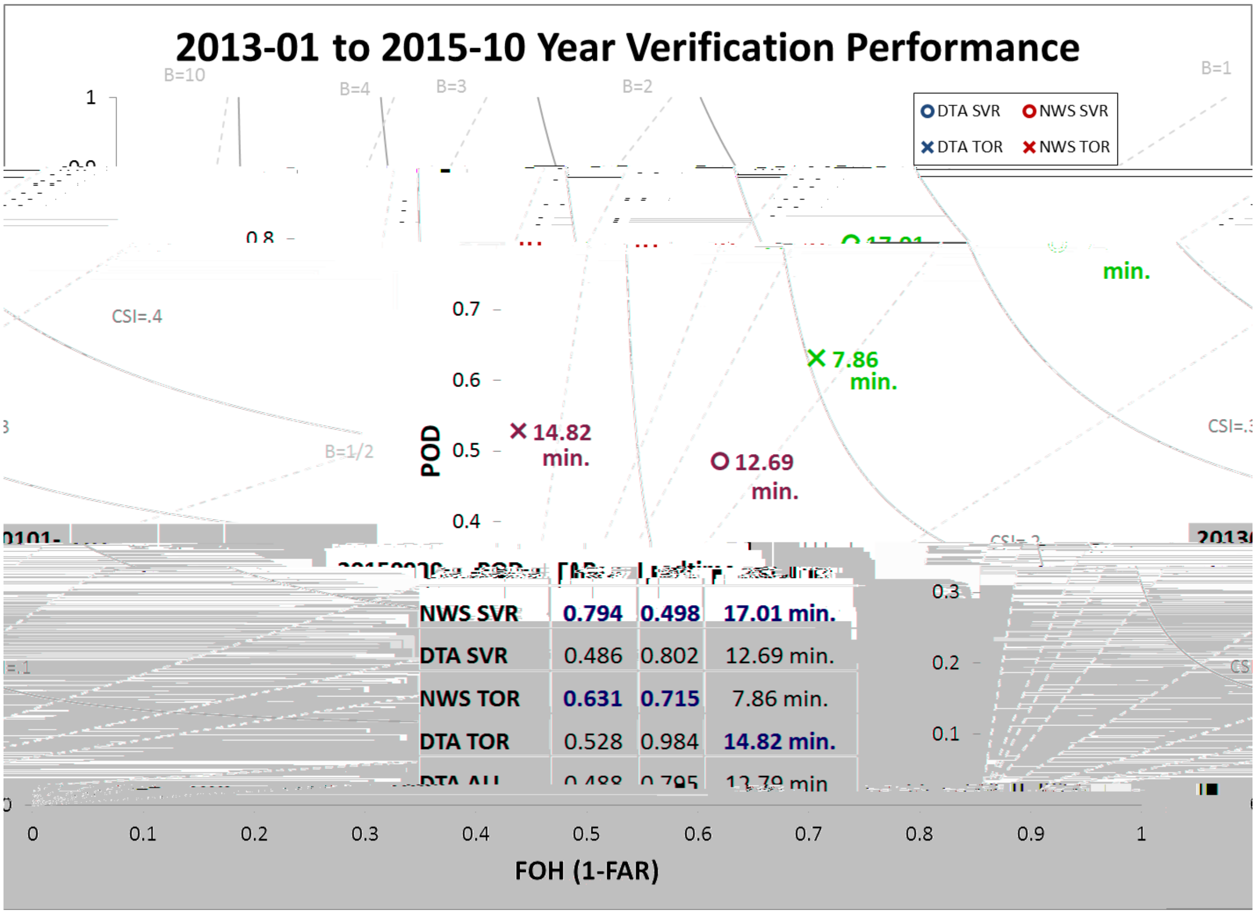


Figure 15. Example warning (orange polygon), instantaneous events (Points A and C, green dot) and event track (Points ${}_1 - {}_5$). Point C is an example of an instantaneous event with 0 min lead time because spatially and temporally it falls outside the warning.

POD, FAR, and lead times were calculated for each NWS severe warning, NWS tornado warning, and ENI top-level DTA using NWS *StormData* reports during the period 1 January 2013 - 30 September 2015(Fig. 16). The Frequency of Hits (FOH), which is 1-FAR, is plotted on the x-axis and the POD is plotted on the y-axis. Biases and CSI lines are also plotted for reference. Ideally the top right is the best area to be on the graph, with high POD and low FAR. First looking at NWS SVR and DTA SVR, NWS SVR has a much higher POD (79%) and lower FAR (50%) than DTA SVR (49% and 80%, respectively). NWS SVR also has a better lead time by 4.32 minutes. Looking next at NWS TOR and DTA TOR, again NWS TOR has higher POD and much lower FAR. DTA TOR has a better lead time of almost 7 minutes, however FAR is 98% and the POD is only 53%. The table in Fig. 16 also shows the DTA results for verifying against any storm event and is very similar to the DTA SVR results. Compared to DTA SVR, the POD is slightly higher and the FAR is slightly lower. The lead times are also very similar.

Since the mid-80s the probability of detection has doubled, however the National Weather Service is still trying to work on improving the false alarm, specifically for tornado warnings. Although ENI DTA has a higher lead time for tornadoes, having a FAR is close to 100% is too high to be useful to NWS forecasters in this regard.



[illegible]	POD	FAR	Lead time
NWS SVR	0.794	0.498	17.01 min.
DTA SVR	0.486	0.802	12.69 min.
NWS TOR	0.631	0.715	7.86 min.
DTA TOR	0.528	0.984	14.82 min.
DTA ALL	0.488	0.795	13.79 min

Figure 16. Performance diagram for NWS warnings and ENI DTAs during verification period 1 January 2013 through 30 September 2015. POD (y-axis), FOH (1-FAR, x-axis), and lead time for NWS severe thunderstorm warnings (red circle), NWS tornado warnings (red X), DTAs as verified by severe reports (blue circle), and by tornado reports (blue X). CSI (gray lines) and Bias (dashed lines) also shown. Table summaries POD, FAR, and lead time for each, top performance for severe and tornadoes highlighted in green text.

5 Recommendations for Future Implementation

Based on the variety of feedback and operational use, it appears that the multiple levels of thunderstorm alerts (including the top-level DTAs) are not useful in most operational warning settings for most forecasters. However, the additional DTA system total lightning-based storm tracking products, such as the current cell polygons, the current storm flash rates, time series, and trends, can be useful for both forecaster situational awareness and within storm interrogation for warning decisions. As one forecaster stated: "Since the lightning data comes in more often than/ahead of the radar imagery, the time series plots could be used to anticipate increases or decreases in storm cell strength, potentially aiding in the warning process (i.e., whether to issue (lead time), continue, or let a warning expire)." Based primarily on forecaster feedback during the year two real-time experiment, it appears that without having access to and the ability to use these total lightning derivative tools to diagnose storm trends, many forecasters would have limited use for the lightning data within the context of severe thunderstorm warning-decision process. This is primarily due to a combination of lack of experience with the data and time needed to interpret and understand trends from the raw data alone.

If the DTA system were to become operational, we strongly suggest that additional development be completed on the storm-tracking algorithm before it is implemented across the NWS. Each week of the real-time evaluation, forecasters complained about issues with lines of storms, storm mergers and splits, and odd artifacts such as smaller cells contained within larger ones. While any tracking algorithm faces these issues, the artifacts seen here hindered the use operationally and additional quality control measures need to be completed to reduce the number of times this type of error occurs. Some forecasters noted in the exit survey that many people would be quick to dismiss the system entirely based on these issues alone.

The time series data was considered the most useful ancillary product in the HWT real time evaluation, however, a number changes within the AWIPS2 implementation and display need to be occur before this product is made operational. First, the current tool methodology was burdensome on the forecaster as it required manual intervention. The forecaster in the HWT had to drag a point to the storm in question and at times move it along with the storm. Some forecasters had difficulty determining which point was associated with which storm in busy work environments. A better implementation would be to allow the forecaster to right-click on any storm in their domain and a temporary pop-up is displayed which he/she could then move to an additional panel in the display if wanted. It was also suggested to denote lightning jumps (as defined by Schultz et al. 2009) on the time series.

For all levels of the thunderstorm alerts, forecasters were absolutely certain that the 45 min expiration window for the alerts was much too long. The forecasters either need to be able to dismiss the alerts after they appeared, configure the expiration time, and/or set to 20 min or less expiration time by default. Additionally, though not widely used within the HWT, forecasters strongly felt that the configurable thresholds would be an important part of the system at the local office level.

Finally, whether or not the ancillary ENI DTA tools and products are implemented within NWS operations, a thorough training program should be created for all NWS forecasters on total lightning use in general operations as well as specifically within severe storm interrogation and diagnosis. This training should build on the foundation of basic research beginning with relationships between storm electrification and lightning with meteorology. This should be followed by a background and comparison of the various lightning detection networks available operationally. Additionally, training needs to review the use of lightning data within decision support, public safety, winter weather, and fire weather forecasting. This training should also incorporate a best practices for operational displays and product integration as well as an overview of future algorithm and tools. While many forecasters felt more comfortable following a week of total lightning use in the HWT, there were still questions as to how to interpret the data. Specifically, forecasters were left wondering what "baseline" lightning flash rates and trends could or should be used for diagnosing severe storms in different environments, what are the limitations of the data, and what regional dependencies exist.

Acknowledgments

This research was supported by Bill Callahan and Amanda Long of Earth Networks who were instrumental in facilitating the delivery of Earth Networks Total Lightning (ETLTN) data, DTA System data, and training materials. Although they may not agree with all of the interpretations/conclusions of this publication their comments were welcomed and greatly improved the manuscript. Additional support was provided by Daphne LaDue, Matthew Elliot, and Benjamin Herzog. The PIs would like to specifically acknowledge Daphne LaDue for adapting the NASA-TLX to this study and acting as an observer during the enhanced case reviews during year one of the experiment. Finally, we thank all of the NWS forecasters who participated in this study within the HWT and provided valuable feedback. Funding for this study was provided by NOAA Oceanic and Atmospheric Research (OAR) Sandy Supplemental Grant NA14OAR4830164.

Notice from NOAA

Mention of a commercial company or product does not constitute an endorsement by NOAA/OAR. Use of information from this publication concerning proprietary products or the tests of such products for publicity or advertising purposes is not authorized. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration.

References

Calhoun, K. M., D. R. MacGorman, C. L. Ziegler, and M. I. . Biggerstaff, 2013: Evolution of lightning activity and storm charge relative to dual-doppler analysis of a high-precipitation supercell storm. *Mon. Wea. Rev.*, **141**, 2199–2223, doi:10.1175/MWR-D-12-00258.1.

Calhoun, K. M., T. M. Smith, D. M. Kingfield, J. Gao, and D. J. Stensrud, 2014: Forecaster use and evaluation of real-time 3DVAR analyses during severe thunderstorm and tornado warning operations in the Hazardous Weather Testbed. *Wea. Forecasting*, **29**, 601–613, doi:10.1175/WAF-D-13-00107.1.

Goodman, S. J., D. E. Buechler, P. D. Wright, and W. D. Rust, 1988: Lightning and precipitation history of a microburst-producing storm. *Geophys. Res. Lett.*, **15**, 1185–1188.

Hart, S. G., 2006: NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA, Hum. Fac. Erg. Soc., 904–908, 32, doi:10.1177/154193120605000909.

Hart, S. G. and L. E. Staveland, 1988: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., North Holland Press.

Saunders, C. P. R., H. Bax-Norman, C. Emerisic, E. E. Avila, and N. E. Castellano, 2006: Laboratory studies of the effect of cloud conditions on graupel/crystal charge transfer in thunderstorm electrification. *Quart. J. Roy. Meteor. Soc.*, **132**, 2653–2673.

Schultz, C. J., W. A. Petersen, and L. D. Carey, 2009: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. *J. Appl. Meteor.*, **48**, 2543–2563, doi:10.1175/2009JAMC2237.1.

Schultz, C. J., W. A. Petersen, and L. D. Carey, 2011: Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. *Wea. Forecasting*, **26**, 744–755, doi:10.1175/WAF-D-10-05026.1.