

Paul X. Flanagan^{*1,2}, Christopher J. Melick^{3,4}, Israel L. Jirak⁴, Jaret W. Rogers⁴, Andrew R. Dean⁴, Steven J. Weiss⁴

¹*School of Meteorology, University of Oklahoma*

²*Oklahoma Climatological Survey, University of Oklahoma*

³*Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma*

⁴*NOAA/NWS/NCEP/Storm Prediction Center*

1. Introduction

Wildfires pose a serious threat to life, property and the nation's natural forests. In 2012, according to the National Interagency Fire Center, 9,443 wildfires were started by lightning, burning over 6 million acres of land. Forecasts for fire weather are performed on a routine basis by the National Weather Service (NWS) Storm Prediction Center (SPC). Of their product suite, the SPC produces outlooks of critical fire weather conditions for dry thunderstorms within the continental United States.

Dry thunderstorm events (hereafter, dry thunder) are a significant source of lightning caused wildfires given that little to no rainfall reaches the ground. The standard definition for dry thunder assumes one or more cloud-to-ground (CG) lightning flashes occurring in conjunction with no more than 0.1 inches of precipitation. However, it is unclear that the standard definition represents all or even most cases of lightning caused wildfires. Given the existing ambiguities in defining an important event, more precise specifications for dry thunder are required for forecasters and also for purposes of verification of forecasts.

The purpose of this study is to better identify the existence of dry thunder events using observational data by examining a variety of classification methods. The findings will hopefully increase understanding of what constitutes dry thunder events by creating a more consistent, realistic, and accurate means of documenting occurrence. If results are promising, any suggested refinements to the standard definition could be implemented at SPC in the future. The following section will describe the methodology used along with the various thresholds included in the study. Section 3 will discuss the results and section 4 will summarize the results for the study.

2.1 Data

This study used several observational data sets, similar to that in Herzog et al. (2012), in order to analyze occurrences of dry thunder within the contiguous United States for 148 days extending from May through September of 2013. The 24-hour period of each day was defined from 12 UTC - 12 UTC for the purpose to match the initial Day 1 SPC dry thunder outlooks. In order to develop the verifying event grids, quantitative precipitation estimates (QPE), obtained from gauge-corrected radar amounts produced by the National Mosaic and Multi-Sensor QPE (NMQ) System (Zhang et al. 2011), and cloud-to-ground (CG) lightning flash data obtained from the National Lightning Detection Network (NLDN) were placed onto a 40-km grid (NCEP 212; <http://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html>).

Precipitable water (PWAT) and surface relative humidity data retrieved from the SPC Mesoscale Analysis database (Bothwell et al. 2002) were also used for several classification methods to refine the number of events to the actual dry thunder occurrences. Since these data were available on an hourly basis, 24-hour average and daily minimum fields were determined for the PWAT and relative humidity fields, respectively. Finally, SPC dry thunder outlook products were converted to the same 40-km grid as the observations using a simple binary (yes/no) diagnosis, depending upon whether the grid point was located inside (value of 1) or outside (value of 0) of the threat area. These data were stored in gridded GEMPAK (GEneral Meteorological PAcKage; desJardins et al. 1991) format and manipulated using the software package. The periods of 12 UTC - 12 UTC on 17-18 May, 31 July - 2 August, and 18-20 August were not included due to missing data.

**Corresponding author address:* Paul X. Flanagan,
School of Meteorology Office 5234, 120 David L. Boren Blvd,
Norman, OK 73072; E-mail: pxf11@ou.edu

2. Data and Methodology

2.2 Methodology

2.2.1 Grid-Point Events

Binary (yes/no) event grids from the observations were created for each day by taking each classification method listed in Table 1 and testing each grid point with the prescribed threshold criteria. Any grid point, which met the constraints of the method, would be given a value of 1, while any grid point which did not would be given a value of 0. The observed event grids were then utilized to verify the forecast events created from the SPC dry thunder outlook data.

The first set of thresholds served as the control method, in which dry thunder was defined as less than or equal to 0.1” of QPE with 1 or more CG lightning flashes (LTNG) occurring over the prior 24 hours (Table 1). For the next two classification methods, the minimum LTNG threshold was increased from 1 to 3 and then to 10 or more flashes while keeping QPE constant. The same variation in LTNG occurred on the next three techniques (i.e. Methods 4-6 in Table 1) but the maximum amount of QPE was increased to 0.25”. The last four ways to define the occurrence of dry thunder (i.e. Methods 7-10 in Table 1) were similar to the control method except they also included environmental parameters to better isolate areas experiencing dry atmospheric conditions. More precisely, average precipitable water thresholds of less than or equal to 0.75” or 1” were applied along with a minimum relative humidity value of less than or equal to 30% or 15% (Table 1).

Method	LTNG (flashes per 24 hours)	QPE (inches)	MINRELH (%)	AVGINPW (inches)
	(>= count)	(<= count)	(<= count)	(<= count)
1 (control)		1	0.1	-
2		3	0.1	-
3		10	0.1	-
4		1	0.25	-
5		3	0.25	-
6		10	0.25	-
7		1	0.1	30
8		1	0.1	15
9		1	0.1	30
10		1	0.1	15

Table 1: Description of ten dry thunder methods used to define dry thunder occurrence where threshold criteria utilized in the study are displayed next to each method. The LTNG column refers to the minimum 24-hour CG lightning flash counts while the QPE column represents the maximum allowed precipitation amounts (in inches) over the 24-hour period. For the last two columns, MINRELH and AVGINPW represent upper limits for the daily minimum relative humidity and the daily average PWAT, respectively.

2.2.2 Neighborhood Events

Up to this point, the existence of dry thunder has been determined in an isolated sense strictly at the grid scale. The SPC Outlook areas, however, actually correspond to a 40% or greater coverage of events occurring in close proximity within a neighborhood surrounding the grid point. In order to address this concern, neighborhood fractional probabilities were calculated for each of the dry thunder event methods based on a procedure described in Schwartz et al. (2010). By applying their formula at each grid point, the number of surrounding grid boxes within a 120-km radius of influence (ROI) which met the criteria outlined in Table 1 was divided by the total number of boxes within that “neighborhood”. From these derived probabilities, binary (yes/no) “neighborhood” coverage event grids from the grid-point events were then created for each outlook day by testing each grid point for a greater than or equal to 40% coverage of events. Any grid point that met this coverage threshold was given a value of 1, while any which did not was given a value of 0. These “neighborhood” coverage event grids for all ten definitions were then utilized in a similar fashion to the grid-point event grids to verify the SPC dry thunder outlook forecasts.

2.3 Verification

The observed event grids formed from each technique were directly compared with the dry thunder outlook grids. As there were only two predicted outcomes, a 2x2 contingency table (e.g., Wilks 2006) was constructed to perform the verification for both the grid-point and neighborhood evaluations. Grid points with a 1 in both the forecast and observed event grids were counted as a “hit”. On the other hand, a 0 value for both datasets was counted as a “correct negative”. Finally, a 0 value calculated for the observed event grid but a 1 in the forecast event grid was counted as a “false alarm”, while a 1 in the observed event grid with a 0 in the forecast event grid was counted as a “miss”.

The total number of hits, false alarms, misses and correct negatives were summed over the entire domain for each day during the five-month time period of 2013. These counts were then used in a daily and accumulated fashion to compute standard forecast verification metrics (e.g., CSI [Critical Success Index], bias, POD [Probability of Detection], FAR [False Alarm Rate]). As revealed in Table 2, SPC outlooks for dry thunder were issued only on 15 of the 148 days examined. In order to narrow the focus to overlap

when both forecast and observations were available, a separate set of results just for the small subset was also produced. In addition, spatial plots of accumulated hits, false alarms, and misses were created for only this set of forecast dates for each of the verification methods to go along with the associated forecast verification statistics. By including a geographical analysis here, a more complete picture is obtained by revealing locations of dry thunder in the observations and how they match to the forecast areas. Consequently, the expectation is that a more definitive procedure for identifying these types of episodes will be revealed from the current investigation.

3. Results

3.1 Accumulated Results

The ten previously described threshold combinations were investigated to determine which definition for dry thunder provided the best means to detect event occurrence. Table 3 shows the accumulated contingency table statistics for each of the methods (described in Table 1) for the entire 148-day period from 2013. The results show that no method provides an overall acceptable way to classify dry thunder events for an extended period of time. Further, the results also reveal the restrictive nature of including environmental parameters in the method, as there is a significant drop in the number of events. The removal of events in which the environmental conditions were not supportive of dry thunderstorms resulted in better agreement with SPC outlooks. The fact that method 9 (Table 3) has a relative increase in misses with respect to the other similar methods (7, 8, and 10) can be explained by the less restrictive PWAT value that is applied here. Finally, from this perspective, method 8 performed the best as it had the lowest number of missed events.

Method	Hits	False Alarms	Misses	CSI	Bias	FAR	POD
1	487	532	64182	0.00747	0.015757	0.52208	0.007531
2	358	661	44903	0.0078	0.022514	0.648675	0.00791
3	192	827	25108	0.00735	0.040277	0.81158	0.007589
4	664	355	107854	0.0061	0.00939	0.348381	0.006119
5	525	494	84410	0.00615	0.011997	0.484789	0.006181
6	328	691	56881	0.00566	0.017812	0.678116	0.005733
7	292	727	8641	0.03023	0.114071	0.713445	0.032688
8	161	858	2932	0.04075	0.329454	0.842002	0.052053
9	407	636	17110	0.02242	0.059542	0.609779	0.023235
10	219	824	4687	0.03822	0.212597	0.790029	0.044639

Table 2: Forecast days found for the study period, including the geographic location of the forecast area.

Forecast days	Area
6/7/2013	West New Mexico
6/28/2013	Southern Colorado and north New Mexico
6/29/2013	West Colorado and North/central New Mexico
6/30/2013	Four Corners region
7/1/2013	Western New Mexico
7/2/2013	Central Arizona
7/3/2013	Central Arizona and North west Nevada/southern Oregon
7/5/2013	Eastern Nevada/western Utah
7/17/2013	Northeast Oregon
8/8/2013	Southeast Oregon and southwest Idaho
8/9/2013	Eastern Oregon, southwest Idaho, northern Nevada and northeast California
8/10/2013	Central Idaho and east central Oregon
8/11/2013	Central Idaho and southwest Montana
8/12/2013	Central Idaho and northeast Oregon
8/13/2013	Central Idaho and northeast Oregon

Table 3: Total number of hits, false alarms and misses from the dry thunder study in 2013, as well as the relevant verification statistics for each set of methods outlined in Table 1.

The accumulated statistics are limited though, in giving insight into the validity of each method. In particular, all of the forecast verification metrics remained unfavorable (e.g., very low POD, high FAR) regardless of the technique explored (Table 3). More importantly, each method found observed dry thunder events on days with no SPC outlook, as revealed in Table 4. This caused a dramatic inflation in the number of misses in the accumulated tallies (Table 3), thereby making it difficult to distinguish an actual episode on days without an outlook area versus just noise. Consequently, incorporating the days without an outlook area tends to mask out any favorable results gained on forecast days. The few insights that can be seen from the accumulated results are that methods 7-10 drastically reduce the number of misses and thus improve the statistics that rely on that metric. Although method 4 has the largest number of hits, it also contains the largest number of misses. Methods 5 and 6 also show an increased number of misses, as compared to the first three methods. This finding is due to the less restrictive value for the QPE threshold allowing for more areas to be identified as dry thunder on days without an outlook area (Table 4). In general, methods 4-6 increase the number of hits and decrease the number of false alarms compared to methods 1-3, but the large increase of misses overshadows any gains made in the other two categories.

Method	Misses (Outlook Days)	Misses (No Outlook Days)
1 (Control)	10471	53711
2	7603	37300
3	4300	20808
4	16898	90956
5	13467	70943
6	9004	47877
7	2081	6560
8	794	2138
9	3793	13317
10	1398	3289

Table 4: Number of accumulated misses for each dry thunder method obtained just for forecast (SPC Outlook) days as well as for days without an outlook area.

3.2 Grid-Point Events

3.2.1 Performance Diagram

The forecast verification metrics shown in Table 3 give no clear indication to a superior method for classifying dry thunder. As previously explained, though, the initial evaluation is heavily influenced by a large set of days with no SPC Outlook for comparison. Therefore, another set of accumulated statistics were created for display in Fig. 1 for just those 15 dates with a forecast area. The illustration in Fig. 1 is known as a performance diagram (Roebber 2009) and its advantageous feature is the ability to summarize multiple verification metrics derived from the contingency on one graph.

Using the bulk statistics from the filtered time period, none of the definitions for dry thunder examined excelled in an absolute sense (Fig. 1). Still, one notable characteristic in the performance diagram was that methods 7-10 relatively outperformed each of the other methods, as the CSI values were the highest around 0.1. On the other hand, methods 1-6 showed greater spread in FAR but, more importantly, exhibited very low PODs nearing the zero minimum (i.e. worst case scenario). Of these four methods that contain environmental information a slight edge was given to method 8 as the authors consider a reduction in missed dry thunder events to be more significant than a decrease in the false alarms. Further investigation of this method will be explored in the next section to provide more insight into how including environmental information increased performance over the control definition (method 1).

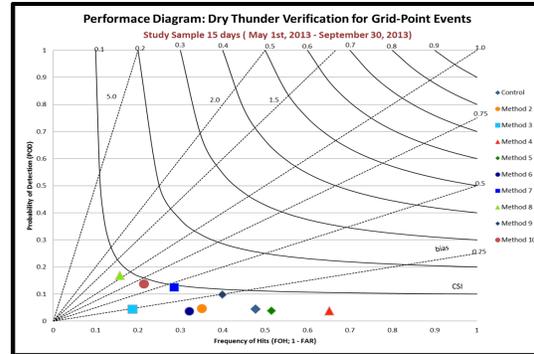


Figure 1: Performance diagram (Roebber 2009) showing accumulated results for contingency table forecast verification metrics from just the SPC Outlook days for dry thunder. These results were obtained for the grid-point events computed for all ten methods. The color legend reveals the matching type of verification method.

3.2.2 Spatial Aggregates

In order to get an understanding of the spatial distribution of the forecast verification, aggregates of misses, hits and false alarms of all the SPC Outlook days were created at each grid point for method 8. For means of comparison, a similar process was accomplished for the control method (method 1). From the forecast perspective, a count of the number of SPC Outlook days by grid point was also produced (Fig. 2). Figures 3 and 4 revealed that both methods 1 and 8, respectively, have misses occurring during the 15 days over the Intermountain West but a doubling in the maximum value for the control definition. In addition to this typical area for dry thunder occurrence, though, the unrestrictive nature for method 1 is further suggested as many more events are spread out over the southern and eastern United States (Fig. 3). In contrast, Fig. 4 indicates that this undesirable feature is lacking in method 8. Instead, an encouraging sign with this latter approach was the highest concentration of misses located just south of the area with the highest number of SPC Outlooks issued in 2013 (compare Fig. 4 with Fig. 2). The fact that dry thunder events happened nearby the composite forecast region appears reasonable, thus meaning that some actual observations were probably not captured.

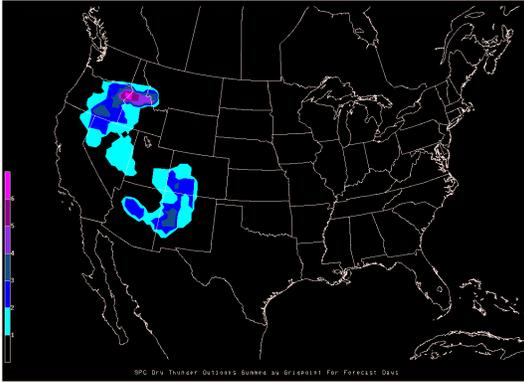


Figure 2: Number of SPC Outlook days for dry thunder in 2013 summed by grid point.

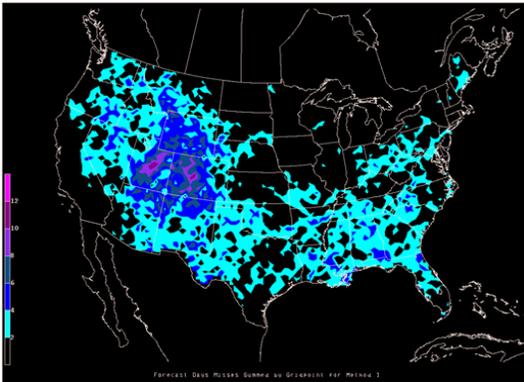


Figure 3: Aggregate plot of the total number of misses computed from the grid-point events for SPC Outlook days in 2013 from method 1.

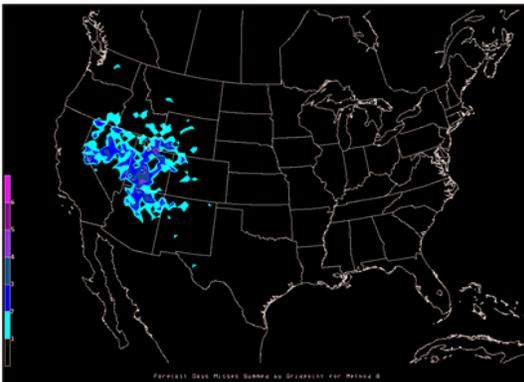


Figure 4: Same as in Fig. 3 except for method 8.

The aggregate plot of hits for method 1 and 8 are displayed in Figs. 5 and 6, respectively. Both techniques show nearly the same analysis of two separate regions, which verify, with method 1 having a larger number of hits (compare Figs. 5 and 6) presumably the results of being less restrictive than method 8. However, the unrealistic increase of misses from method 1 has to be taken into account as well. Upon examination of Figs. 7 and 8, both sets of verification approaches indicate an overforecast for dry thunder events for a couple

of areas in the Western part of the United States for the 2013 season. More precisely, the number of false alarms increased for method 8 within across central Idaho and eastern Oregon when limiting the number of observed events by incorporating environmental information (Fig. 8). Since this region has the highest number of SPC Outlook days, this impacted the verification metrics of method 8. Still, the authors consider the increase in false alarms to be a minor concern overall relative to the decrease in misses.



Figure 5: Aggregate plot of the total number of hits computed from the grid-point events for SPC Outlook days from method 1.

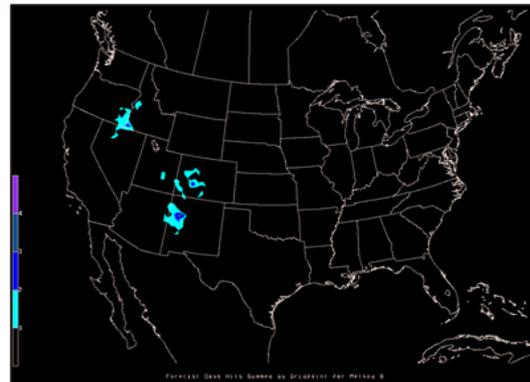


Figure 6: Same as Fig. 5 except for method 8.

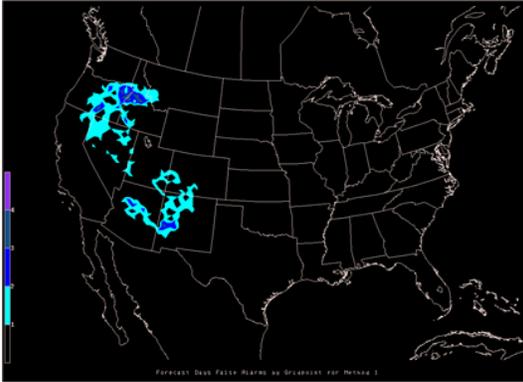


Figure 7: Aggregate plot of the total number of false alarms computed from the grid-point events for SPC Outlook days from method 1.

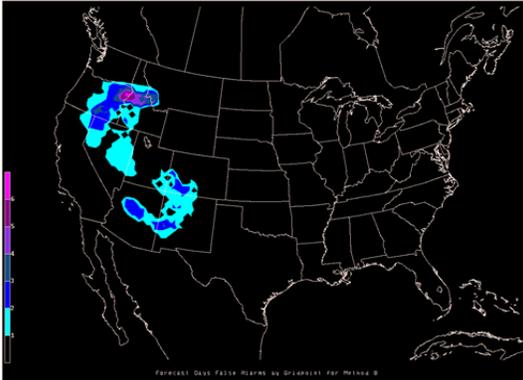


Figure 8: Same as Fig. 7 except for method 8.

3.3 Neighborhood Events

3.3.1 Performance Diagram

Forecast verification based on a neighborhood concept, which takes into account the spread of close observations, will now be addressed. Similar to that of the grid-point events verification, a performance diagram was constructed for Fig. 9 using the accumulated results for each method from just the 15 forecast dates. In most cases, the requirement to have some coverage in the grid-point events was beneficial to the verification metrics. While there was more of a range in the skill scores when comparing Fig. 9 to Fig. 1, knowledge of atmospheric moisture in defining dry thunder continued to be important as the first six methods underperformed once again. In contrast, the highest POD from method 8 had exceeded 0.2 for this evaluation (Fig. 9). Actually, in terms of CSI, statistical results from method 7 showed a slight advantage compared to the other 3 approaches of incorporating environmental data (methods 6, 8-10). Still, Method 8 performed comparatively well and for consistency will still be used for further investigations.

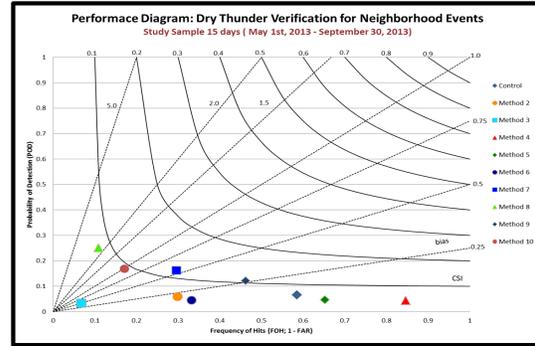


Figure 9: Same as Fig. 2, except for neighborhood events.

3.3.2 Spatial Aggregates

A series of spatial aggregates of hits, misses, and false alarms were also compiled from the neighborhood events for methods 1 and 8. These were created to discern the coverage pattern in observational datasets over the 15 SPC dry thunder Outlook days forecasted. The hope would be that those verifying regions with scattered to widespread grid-point events (40% or higher) overlapped with the forecast area (shown in Fig. 2). For the control method, the obvious feature when comparing Fig. 10 to Fig. 3 was the elimination of missed events over the Central and Southern United States as most of these occurrences tended to be spotty in nature. Coincidentally, this reduction also extended farther west into locations where dry thunder is usually more typical. On the latter point, a similar contrast for method 8 is evident between the neighborhood and grid-point event analyses (compare Fig. 11 to Fig. 4).

The evaluation thus far has been favorable with ensuring the existence of numerous nearby events. However, the following analyses illustrate there are negative trade-offs with the neighborhood approach. Figures 12 and 13 show the total number of hits for methods 1 and 8, respectively, based on the spatial coverage criteria for dry thunder grid-point events. First, the location and number of hits appear very similar between the two definitions. Still, there is a decrease in the overall counts when comparing the neighborhood to the grid-point verification approaches, which is especially the case for the control method (compare Fig. 12 to Fig 5). Finally, the undesirable consequences of being more restrictive also impact the false alarm perspective. Both Figs. 14 and 15 reveal an increase in tallies over the Southwestern and Northwestern United States in relation to the analyses provided in Figs. 6 and 7, respectively. As a result, the extent of false alarms in method 8 was much larger than what was observed with the control method.

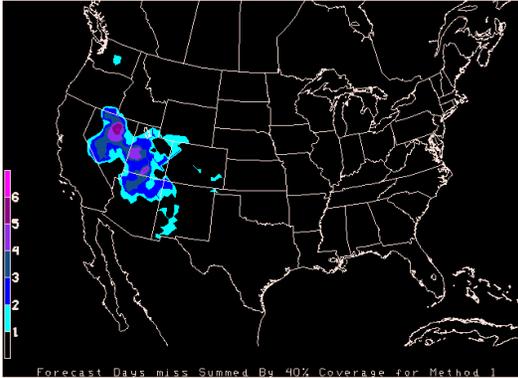


Figure 10: Aggregate plot of the number of misses computed from the neighborhood approach for SPC Outlook days from method 1.

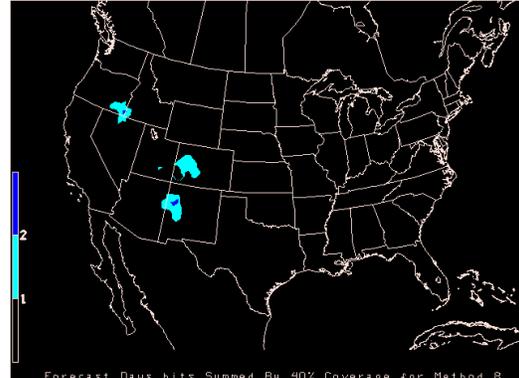


Figure 13: Same as Fig. 12 except for method 8.



Figure 11: Same as Fig. 10 except for method 8.

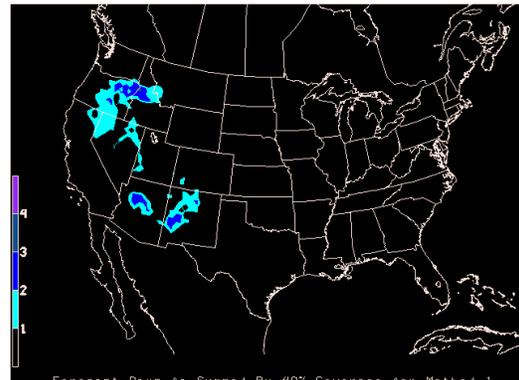


Figure 14: Aggregate plots of the number of false alarms computed from the neighborhood approach for SPC Outlook days from method 1.

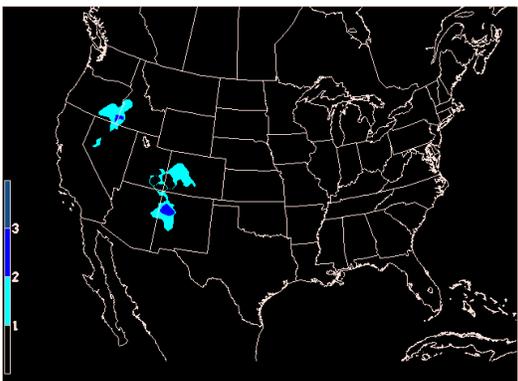


Figure 12: Aggregate plot of the number of hits computed from the neighborhood approach for SPC Outlook days from method 1.

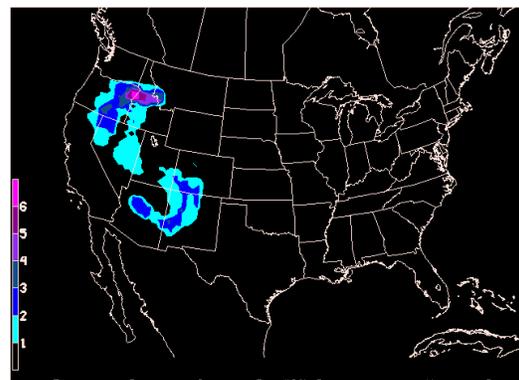


Figure 15: Same as Fig. 14 except for method 8.

4. Summary and Conclusions

In this study, observational data were used to create gridded classifications of dry thunder events by using ten various combinations of thresholds from QPE, lightning flash data, and environmental parameters of precipitable water and surface relative humidity. Different approaches were examined in order to reveal which provided the best means for verifying forecasts. To accomplish this, each set of these

methods was used to compare against the initial Day 1 SPC dry thunder Outlooks issued from May-September 2013. For the evaluation process, a 2x2 contingency table was created for each method and summary forecast verification statistics calculated for individual days as well as in an accumulated sense for the entire period. Two different approaches were used in the evaluation, one that strictly focused on grid-point-to-grid-point comparisons and a neighborhood technique which required greater coverage of dry thunder occurrence to better match the SPC Outlook definition. For this purpose, the spatial extent of the events was also examined by determining the fractional coverage of dry thunder in the 40 km ROI around each grid-point.

The cumulative statistics for the five-month period were found to obscure the details with only 15 days of forecasts available for comparison. Thus, results were filtered to focus on how the methods performed on dates with an issued Day 1 SPC Outlook. First, the current work revealed that the forecast verification scores were mediocre overall regardless of how the event was defined. In a relative sense, though, the standard definition [one or more CG flashes with no more than 0.1" of precipitation] for dry thunder resulted in many more observed than forecast events with a high FOH, but very low POD. Unfortunately, the spatial plots also showed the misses from the control method to be spread over a large portion of the country in locations, which were atypical for dry thunder occurrence. On the other hand, a more realistic assessment of conditions was produced by incorporating low RELH and PWAT limits from environmental datasets using four different approaches. Specifically, improved skill scores compared to the control method were common and method 8 was chosen as best, owing to the relatively higher reduction in misses.

The examination into the spatial coverage of dry thunder grid-point events by classifying events in a neighborhood approach provided a more in depth analysis into the performance of each method. The low frequency and isolated nature of the observed events was eliminated from the southern and eastern United States by requiring 40% coverage to dry thunder. In other words, the benefit of this restrictive nature was to emphasize the majority of the activity in climatologically favored areas. The examinations also revealed the added usefulness of specifying a neighborhood approach in combination with utilizing environmental observations in the definition of dry thunder. In particular, most of the forecast verification metrics improved with

method 9 exhibiting the best CSI values and the highest POD observed for method 8. As was discussed earlier, though, the complete assessment is often complex as there are caveats that come along with increased restrictions. It was found that more realistic neighborhood approaches resulted in fewer observed events but also a decrease in hits and a corresponding increase in false alarms.

It should be emphasized that the conclusions established here are limited given the short time period of the study. Nonetheless, the addition of environmental parameters and a spatial-coverage criterion into the dry thunder definition reduced the number of observed events to better agree with climatologically favored areas. Future work should expand to include Day 1 SPC Outlooks from more years and possibly more thresholds. Other environmental parameters could also be taken into account, including fuel dryness.

References

- Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. Preprints, 21st Conf. on Severe Local Storms, Amer. Meteor. Soc., San Antonio, TX, J117-120.
- desJardins, M. L., K. F. Brill, and S. S. Schotz, 1991: Use of GEMPAK on Unix workstations. Preprints, *Seventh International Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, New Orleans, LA, Amer. Meteor. Soc., 449-453.
- Herzog, B.S., C.J. Melick, J.S. Grams, I.L. Jirak, A.R. Dean, and S.J. Weiss, 2012: Usefulness of storm-scale model guidance for forecasting dry thunderstorms at SPC. Preprints, 26th Conf. Severe Local Storms, Nashville, TN, P8.123.
- Roeber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601-608.
- Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263-280.
- Wilks, D.S., 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd Ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.
- Zhang, Jian, and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321-1338.