P8.123 USEFULNESS OF STORM-SCALE MODEL GUIDANCE FOR FORECASTING DRY THUNDERSTORMS AT SPC

Benjamin S. Herzog[1, 2, 3*], Christopher J. Melick[2,4], Jeremy S. Grams[4], Israel L. Jirak[4], Andrew R. Dean[4], Steven J. Weiss[4]

[1]*NOAA/OAR/ National Severe Storms Laboratory*
[2]*CIMMS, University of Oklahoma*
[3]*School of Meteorology, University of Oklahoma*
[4]*NOAA/NWS/NCEP/Storm Prediction Center*

1. Introduction and Background

Wildfires are a serious threat to life and property in the United States and many of these wildfires in the U.S. are started by lightning. According to the National Interagency Fire Center, in 2011 alone, there were 10,249 wildfires started by lightning, resulting in over 3 million acres burned. Typically, lightning strikes which ignite wildfires are spawned from dry thunderstorms (hereafter, dry thunder). A storm is typically classified as dry thunder when lightning occurs with less than 0.1 inches of rain in a given location. Even with the advent of convection allowing models, though, errors still remain with respect to the guidance accurately predicting placement and timing of thunderstorms (e.g., Weisman et al. 2008). In addition, the forecasting challenge becomes more substantial when lightning is required to be co-located with minimal precipitation, especially since large amounts of precipitation are often a by-product of deep convection.

Fire weather forecasting is one of the duties of the National Weather Service (NWS) Storm Prediction Center (SPC). The SPC is charged with creating national fire weather guidance to be utilized by the NWS Weather Forecast Offices, as well as other entities. The SPC creates dry thunder outlooks out to three days that highlight critical areas at risk for fire weather. These critical areas are typically issued when widespread dry thunder is expected to occur where dried fuels exist.

**Corresponding author address*: Benjamin S. Herzog, NOAA/OAR/NSSL, 120 David L. Boren Blvd., Norman, OK 73072; e-mail: Ben.Herzog@ou.edu

To predict dry thunder, a forecaster must be able to determine if lightning will occur without significant amounts of precipitation. While precipitation has been a common and long-standing output of numerical models, explicit lightning output has only become available in the past few years to SPC forecasters through an algorithm (e.g., McCaul et al. 2009) developed that predicts lightning flash rate density (FRD; flashes (5 min)$^{-1}$ km^{-2}). In particular, McCaul et al. (2009) proposed a lightning threat parameter that is based on charge separation processes and the microphysical properties of ice-phase hydrometeor fields generated by numerical models such as the Weather Research and Forecasting (WRF) model.

The work presented by McCaul et al. (2009) proposed three methods for lightning prediction. The first method looked at upward fluxes of precipitating ice hydrometeors in the mixed phase region at the -15 ^{o}C level. The second method was based on the vertically integrated amounts of ice hydrometeors in each model grid column. It was found the first parameter did well in predicting the temporal variability of lightning activity, and the second method did well in predicting the spatial variability of lightning activity. Finally, a third parameter blending the positive aspects of the two methods was created. Given that this latter method has been shown in their study to perform the best, it will serve as the surrogate to represent explicit lightning in identifying dry thunder.

The purpose of this paper is to determine how well the National Severe Storms Laboratory (NSSL) version of the WRF-ARW (e.g., Kain et al. 2010) can predict dry thunder using various methods of defining a dry thunder event. For this particular model configuration, an hourly-max version of the lightning threat parameter was implemented in

2010 and a preliminary investigation (e.g., Miller et al. 2010) has identified some potential weaknesses. Still, taking this into consideration, an examination showing whether useful thresholds applied specifically for dry thunder could show some merit in delineating specific areas of interest. Experiments of classifying an event in this study were conducted by utilizing several values for the FRD minima, setting various criteria for the radius of influence (ROI; km) in a 'neighborhood' fashion (e.g., Harless 2010) to account for spatial uncertainty, and incorporating constraints for environmental moisture and precipitation. The ultimate goal would be to provide a calibrated, first-guess forecast on a routine basis that would ultimately improve products on the SPC fire weather desk.

In order to predict dry thunder, gridded forecasts of dry thunder events were created for each day during the summer (June, July, and August) of 2011 from the NSSL WRF-ARW. The gridded forecasts of dry thunder, or WRF event grids, were created by examining different thresholds of daily maximum lightning threat from McCaul et al. (2009), quantitative precipitation forecast (QPF), and 24-hour average precipitable water (PWAT). Observed cloud-to-ground lightning activity, precipitation, and analyzed PWAT fields were then used to also create observed event grids. Since PWAT is a continuous gridded field, it was included due to the fact that it can serve as a discriminator between dry, wet, and mixed thunderstorms where precipitation data are lacking. The observed event grids served to verify the WRF event grids through computations of traditional forecast verification scores (Wilks 2006), such as probability of detection (POD), bias, false alarm ratio (FAR), and critical success index (CSI).

Section 2 of this paper describes the various thresholds used to forecast dry thunder and the different methods for creating the WRF event grids. Section 3 shows the verification statistics of the various forecasts. Finally, section 4 summarizes the results of this study.

2. Data and Methodology

2.1 *Data*

This study utilized several datasets to examine favorable conditions for the development of dry thunder encompassing a domain (Fig. 1) within the contiguous, United States and that matches the 24-hour period (12 UTC 12 UTC) valid for most fire weather outlooks issued at SPC. Model output from the 4-km NSSL WRF-ARW (Kain et. al, 2010) was the only data used to create the forecast WRF event grids. In order to verify the WRF event grids, observed event grids were created using three additional data sets. These were the: 1.) Cloud-to-ground (CG) lightning flashes obtained from the National Lightning Detection Network (NLDN), 2.) PWAT information retrieved from the SPC Mesoscale Analysis database (Bothwell et. al, 2002), 3.) and the observed, quantitative precipitation estimates obtained from gauge-corrected, radar amounts produced by the National Mosaic and Multi-Sensor QPE (NMQ) System (Zhang et. al 2011). All of the data were stored in gridded, GEMPAK (GEneral Meteorological PAcKage; desJardins et. al, 1991) format and manipulated using the software. A total of 88 days were examined in June, July, and August of 2011, with the periods of 12 UTC – 12 UTC on June 8 - June 9, August 2 – August 3, and August 23 – August 24 not included due to missing data.

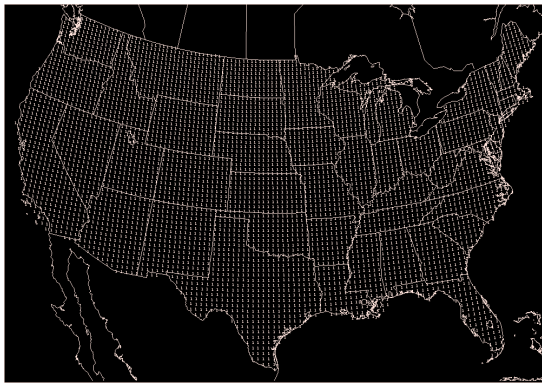

Figure 1: The domain of the study.

2.2 *Methodology*

Construction of binary (yes/no) event grids for the occurrence of dry thunder from the 4-km NSSL WRF-ARW guidance was performed for each day in the time period. These forecast event grids were created using both at-the-grid point and a binary 'neighborhood' approach. The binary WRF event grids were then compared to similar observation grids created from the SPC

Mesoscale Analysis, the NLDN, and the NMQ. Each observed event grid and WRF event grid had matching thresholds for precipitation amounts and PWAT.

As a first guess, grid points with daily maximum lightning threat ≥ 0.55 flashes (5 min)$^{-1}$ km^{-2}, average daily PWAT ≤ 0.5", and 24-hr QPF ≤ 0.1" were considered events. While the latter two conditions are roughly based on SPC guidelines, the lightning threat threshold selection is based on bias results by Miller et al. (2010). Using the NLDN as verification as well, their work indicated that an FRD value of 0.55 performed the best in discriminating where lightning would and would not occur. Since the conclusions from Miller et al. (2010) were limited to a few months' time period in 2010, though, other lightning threat thresholds were also examined. In addition, the PWAT threshold was eventually increased to ≤ 1.0" for most of the other tests as it became evident rather quickly that the default choice was too restrictive. Finally, various distances for the ROI were applied to hopefully improve placement errors in dry thunder and presumably overall verification statistics. Consequently, Table 1 shows each set of configuration tests for the WRF event grids, along with the associated thresholds and creation approach.

Experiment Number	Creation Approach	FRD [flashes (5 min)$^{-1}$ km^{-2}] ≥	QPF ≤	Average PWAT ≤
1	grid point	0.55	0.1"	1.0"
2	neighborhood on LTG, ROI = 40	> 0	0.1"	1.0"
3	neighborhood on LTG, ROI = 40	0.55	0.1"	1.0"
4	neighborhood on LTG, ROI = 40	1	0.1"	1.0"
5	neighborhood on LTG, ROI = 40	2	0.1"	1.0"
6	neighborhood on LTG, ROI = 40	3	0.1"	1.0"
7	neighborhood on LTG, ROI = 40	5	0.1"	1.0"
8	neighborhood on WRF event	0.55	0.1"	1.0"
9	neighborhood on LTG, ROI = 20	0.55	0.1"	1.0"
10	neighborhood on LTG, ROI = 10	0.55	0.1"	1.0"

Table 1: Experiment Configuration: The methods and thresholds used to create each set of WRF forecast grids for verification.

2.2.1 *The grid point approach*

The grid point approach focused on individual grid points for each day in the period. At each grid point, thresholds of lightning threat, QPF, and average PWAT were examined. Any grid point meeting those thresholds was given a value of 1; any grid point not meeting those thresholds was given a value of 0.

2.2.2 *The binary neighborhood approach*

An additional set of WRF event grids were created by using the binary neighborhood approach with an ROI at 10 km, 20 km and 40 km. This approach identified the neighborhood maximum value of lightning threat within a specified radius of influence (ROI). That value was then assigned to each grid point within the ROI. This resulted in more spatial coverage of the lightning threat field, and thus created more WRF events. From there, the WRF event grids were created in much the same way as in the grid point technique. The same QPF and average PWAT grids were used, and the new neighborhood lightning threat grid was used in place of the original lightning threat grid. Again, any grid point meeting those thresholds was given a value of 1; any grid point not meeting those thresholds was given a value of 0.

Another set of WRF event grids was also created using the neighborhood approach slightly differently. Initially, the WRF event grids were created using the grid point technique. Before the WRF event grids were transferred to the 40-km grid, the binary neighborhood technique was applied to the 4-km WRF event grid with an ROI of 40 km. By doing so, at each grid point with a value of 1, the value was spread to all grid points within a 40-km radius of the grid point.

2.2.3 *Creation of the observed event grid*

The observed datasets were placed on a 40-km grid (NCEP 212; http://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html) for evaluation purposes, similar to how the Short-Range Ensemble Forecast (SREF) calibrated thunder probabilities are verified at SPC against one or more NLDN CG lightning strikes (Bright et al. 2005). In order to perform a direct comparison against the

forecast events, bilinear interpolation was utilized to convert the 4-km NSSL WRF-ARW to the coarser 40-km grid. Points on the 40-km grid were turned on if the interpolated value was greater than zero (i.e. at least one surrounding grid box on the finer mesh from the NSSL WRF-ARW was true as well). In all verification approaches, there was a different observed event grid created to match each WRF event grid. The same PWAT thresholds as in the matching WRF event grid were examined, while the threshold for precipitation and lightning flash count remained at no more than 0.1 inches and at least 1 strike over the 24-hour period, respectively. Again, any grid point meeting those thresholds was given a value of 1; any grid point not meeting those thresholds was given a value of 0.

2.2.4 *Verification*

Finally, each WRF event grid was compared to the matching observed event grid. Since the verification involves discrete predictands with only two possible outcomes (i.e., yes/no), a 2x2 contingency table (e.g., Wilks 2006) was constructed to tally all possible outcomes. Grid points that had a 1 (a forecasted and observed event) in both the WRF event and observed event grids were considered a "hit", or good forecast. Grid points where the WRF event grid had a 1 and the observed event grid had a 0 were considered a "false alarm". Grid points where the WRF event grid had a 0 and the observed event grid had a 1 were considered a "miss". Grid points where the WRF event grid had a 0 and the observed event grid had a 0 were considered a "correct negative".

For each set of event grids, the number of hits, false alarms, misses and correct negatives were totaled over the entire domain for every day during the summer of 2011. From there, standard forecast verification statistics were computed. Finally, summary plots for the period were created containing the total number of hits, misses, and false alarms at each grid point for the best set of thresholds identified from the skill scores. This analysis perspective will hopefully shed some light on the spatial distribution of the hits, misses, and false alarms.

3. Results

As mentioned, several sets of thresholds were examined in order to determine which configuration would have the best forecasting capabilities. Table 2 shows each set of WRF grids and the verification statistics associated with those grids. Interestingly, none of the methods have a CSI greater than 0.25, while all results have a FAR of at least 0.70. Also, most of the approaches display a bias greater than 1. In the few cases where the bias is less than 1, the POD is unfavorably low. As for the other tests examined, the forecasted events tend to be more numerous than the observed events given the high FAR values along with a large bias signal. In fact, the FAR does not change much, even with meeting the more restrictive criteria.

Exp No	Hits	False Alarms	Misses	POD	FAR	CSI	BIAS
1	5755	11744	15155	0.28	0.67	0.18	0.84
2	18657	81389	2253	0.89	0.81	0.18	4.78
3	15406	47629	5504	0.74	0.76	0.22	3.01
4	13581	39878	7329	0.65	0.74	0.22	2.56
5	8694	27396	12216	0.42	0.76	0.18	1.73
6	4677	17883	16233	0.22	0.79	0.12	1.07
7	992	6542	19918	0.05	0.87	0.04	0.36
8	522	7694	654	0.44	0.93	0.06	6.99
9	12958	34927	7952	0.62	0.73	0.23	2.29
10	10807	26520	10103	0.52	0.71	0.23	1.79

Table 2: The total number of hits, false alarms, and misses, as well as verification statistics for each set of verification methods outlined in Table 1.

Along these lines, the desirability of lower thresholds can further be illustrated in Figure 2 by examining just experiments 2-7 (as defined in Table 1). The thresholds for PWAT and QPF on experiments 2-7 are the same, whereas the lighting threat is allowed to vary from greater than 0 (experiment 2) to greater than or equal to 5 (experiment 7). Once the lightning threat gets above 1 (experiment 4), Figure 2 indicated that the POD drops dramatically. In addition, the decrease in total number of hits, misses, and

false alarms with increasing FRD from Table 2 reveals that most of the lightning threat values reside below 1. Consequently, it seems that the best statistical outcome from the current work is to attempt to have a high POD with less restrictive criteria for defining dry thunder.

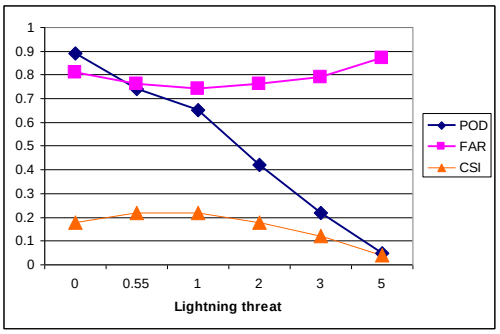

Figure 2: Plot of POD, FAR and CSI with respect to increasing lightning threat values (flashes (5 min)$^{-1}$ km^{-2}). All other sensitivity tests are kept constant at an ROI=40-km and PWAT ≤ 1.0".

Experiments 2 and 3 are not particularly restrictive, so they are able to capture most of the events (Table 2). Those results display the highest POD, and are therefore the most useful forecasts. While those forecasts still have high FAR and low CSI, they can capture approximately 75% of the observed dry thunder events. Experiment 2 has a higher POD than experiment 3, however it also has a higher FAR, a lower CSI, and a larger bias. With that information, we conclude that the thresholds associated with experiment 3 (lightning threat ≥ 0.55, PWAT ≤ 1.0" and QPF ≤ 0.1") produce the best forecast for the three month time period in the summer of 2011.

Figures 3, 4, and 5 highlight spatial plots of the total number of hits, misses, and false alarms at each grid point, respectively, for the experiment 3 grids. Each indicate that most of the forecasted events and observed events for the experiment 3 thresholds occurred in the western United States. This result is expected, as this region generally has the highest frequency occurrence of fire weather events. Interestingly, the maximum number of hits and false alarms is co-located with the minima in misses in south central/southeast Colorado. In that area, nearly half of the days in the period were considered events, with around 40 out of 88 days either hits or false alarms.

There is also a secondary maximum in false alarms over the Appalachian Mountains. These maxima in the mountainous regions are not unexpected. The mountains provide orographic forcing for the storms. This additional forcing can cause more storms to initiate than normal. We believe that the reason most of the dry thunder forecasts in the Appalachian region are not realized is the fact that most of the storms that occur in that region do not meet the ≤ 0.1" of precipitation requirement for storms to be considered dry.

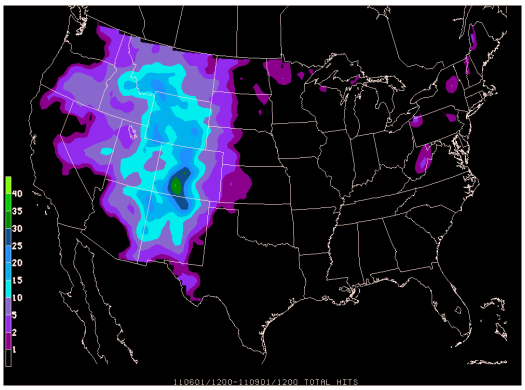

Figure 3: The total number of hits found at each grid point throughout the domain of the study.

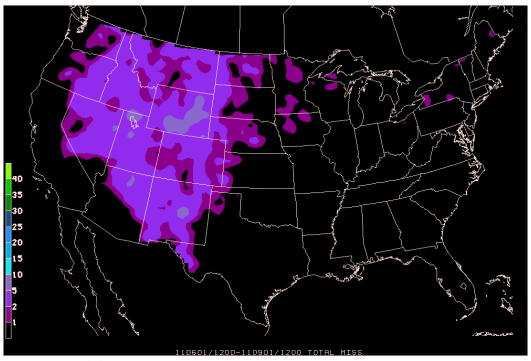

Figure 4: Total number of misses found at each grid point throughout the domain of the study.

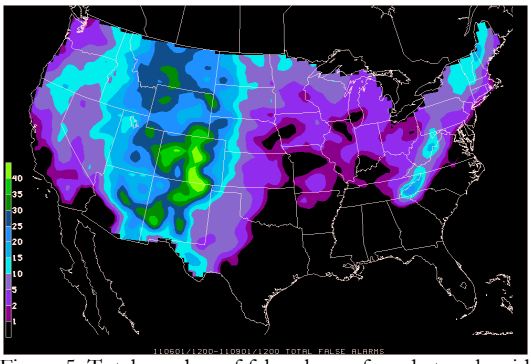

Figure 5: Total number of false alarms found at each grid point throughout the domain of the study.

In summary, findings from the 2x2 contingency table in this study suggest a complex picture in trying to classify what constitutes dry thunder. Still, one conclusion is evident from the plots (Fig. 3, 4, and 5) and the skill score statistics (Table 2 and Fig. 2) given earlier; the thresholds associated with experiment 3 tend to be more accurate but also have a problem in substantially over-forecasting events. However, it is encouraging that those forecasts are able to capture a large majority of the dry thunder episodes in the domain.

4. Summary and Conclusions

In this study, the NSSL version of the WRF-ARW (Kain et al. 2010) was used to create gridded, deterministic forecasts of dry thunder events. Grid points where certain thresholds of the FRD (McCaul et al. 2009), PWAT, and QPF were met were considered dry thunder events. These forecasts were verified using gridded observed data. The observed event grids were created with observed lightning, PWAT, and precipitation data with thresholds matching those used in creating the gridded forecasts.

Several thresholds of lightning threat and PWAT, along with two different approaches of creating these grids were examined to determine which created the most useful forecast. Each experiment was found to have a very high FAR and a low CSI. However it was found that the forecast created by utilizing a binary neighborhood approach on the lightning threat grid, thresholds of lightning threat ≥ 0.55, a PWAT ≤ 1.0" and a QPF ≤ 0.5" created a forecast with a high POD. This forecast was able to detect nearly 75% of the dry thunder events that occurred in the time period of this study.

It should be noted that that the sample size for this study is relatively small. Only about 3 months (i.e. June, July, and August) of data from 2011 were examined, so additional study is likely needed to fully understand the capabilities of this forecasting method. Given some of the encouraging results discussed already, we also feel that a more thorough investigation is warranted. As long as the limitations of the forecasting technique are acknowledged and accounted for, this may indeed provide valuable first-guess guidance to assist fire weather forecasters. That first guess may streamline the fire weather forecasting process, and simplify what can prove to be a very complicated forecast.

Acknowledgements. We would like to thank Jack Kain of NSSL for providing the NSSL WRF-ARW model and Carrie Langston of NSSL for providing the NMQ data available to SPC.

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.

Bright, D.R.*, M.S. Wandishin, R.E. Jewell, and S.J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. *Preprints*, Conf. on Meteor. Applications of Lightning Data, San Diego CA

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.

Harless, A.R., 2010: A report-based verification study of the CAPS 2008 storm-scale ensemble forecasts for severe convective weather. M.S. Thesis, School of Meteorology, University of Oklahoma, 143 pp.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high resolution forecast models: Monitoring selected fields and phenomena every time step. Wea. Forecasting, 25, 1536-1542.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, D. J. Cecil, 2009: Forecasting Lightning Threat Using Cloud-Resolving Model Simulations. *Wea. Forecasting*, **24**, 709–729.

Miller, S. D., G. W. Carbin, J. S. Kain, E. W. McCaul, A. R. Dean, C. J. Melick, S. J. Weiss, 2010: Preliminary investigation into lightning hazard prediction from high resolution model output.

Preprints, 25th Conf. Severe Local Storms, Denver CO.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23,** 407–437.

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.