12B.3 UTILIZING THE HIGH-RESOLUTION ENSEMBLE FORECAST (HREF) TO PRODUCE CALIBRATED PROBABILISTIC THUNDERSTORM GUIDANCE AT THE STORM PREDICTION CENTER

David R. Harrison *
Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma, and
NOAA/NWS Storm Prediction Center, Norman, OK

Israel L. Jirak and Matthew Elliott
NOAA/NWS Storm Prediction Center, Norman, OK

1. INTRODUCTION

The Storm Prediction Center (SPC) issues Enhanced Thunderstorm outlooks that depict the probability of thunderstorms across the contiguous United States (CONUS) in 4- or 8-hour time periods. Specifically, these forecasts represent the probability of at least 1 cloud-to-ground (CG) lightning flash within 20 km (12 miles) of a point location during the valid forecast period. The increased temporal resolution of these Enhanced Thunderstorm outlooks aids NWS forecasters and partners in time-sensitive decisions related to thunderstorms.

Accurately predicting the timing and location of all thunderstorms across the CONUS can often be time consuming and mentally taxing on forecasters. To aid in the generation of these and other forecast products, the SPC post-processes the National Centers for Environmental Prediction (NCEP) Short-Range Ensemble Forecast (SREF; Du et al. 2014) system to provide operational probabilistic guidance for the prediction of lightning hazards. This guidance relies on physically-based parameters (e.g. Bright et al. 2005) to produce probabilistic forecasts of thunderstorms, which are then calibrated using CG lightning flash data from the National Lightning Data Network (NLDN) such that the predicted probabilities from an independent sample are statistically reliable against the verifying NLDN data. Although this method has generally shown both skill and reliability at predicting the occurrence of CG lightning flashes (Bright and Grams 2009), the temporal and spatial accuracy of the predictions is limited in part by the inability of the SREF to explicitly resolve convection.

Given these apparent limitations of the SREF, it was hypothesized that the addition of simulated radar reflectivity and other storm-attribute fields from a convection-allowing model (CAM) or ensemble may lead to improved probabilistic, calibrated thunderstorm predictions. To this end, a new suite of probabilistic thunderstorm guidance products have been derived from the NCEP High-Resolution Ensemble Forecast (HREF; Roberts et al. 2019) system and implemented operationally at SPC.

This paper will briefly describe how the HREF Calibrated Thunder suite was developed (section 2), then compare the performance and reliability metrics of the new guidance to that of the original SREF Calibrated Thunder products (section 3). All HREF Calibrated Thunder products described herein are available on the SPC website at https://www.spc.noaa.gov/exper/href/?model=href&product=guidance_thunder_hrefct_004h.

2. DATA AND METHODS

The HREF Calibrated Thunder forecast products were derived using prognostic data fields from the operational HREF version 2 (HREFv2) and an experimental version of the HREF, known unofficially as the HREFv2.1, currently being tested within the SPC. The operational version of the HREF is composed of eight ensemble members with four deterministic CAM configurations provided by the Advanced Research version of the Weather Research and Forecast Model (ARW; Skamarock et al. 2008), the National Severe Storms Laboratory

---

**Corresponding author address*: David R. Harrison, 120 David L. Boren Blvd, Norman, OK 73072; email: david.harrison@noaa.gov

version of the ARW (NSSL), the Nonhydrostatic Multiscale Model on the B Grid (NMMB; Janjic and Gall 2012), and the 3 km North American Mesoscale Forecast System (NAM). Each configuration is represented twice within the HREFv2 ensemble by including 12-hour time-lagged initializations of each member. A full list of the model cores, boundary conditions, microphysics schemes, and planetary boundary layer (PBL) schemes of each of the eight members is provided by Roberts et al. (2019), their Table 1.

The experimental HREFv2.1 adds the operational High Resolution Rapid Refresh (HRRR; Benjamin et al. 2016) and its 6-hour time-lagged run, giving the ensemble a total of 10 members for the first 30 forecast hours, 9 members through forecast hour 36, and 5 members through forecast hour 48. The inclusion of the HRRR has been shown to increase member spread and improve the overall skill of the ensemble (Gallo et al. 2018).

Both 00z and 12z cycles of the HREFv2 were obtained for 1 July 2017 – 1 January 2019, and the HREFv2.1 cycles were collected for 1 January 2019 – 1 January 2020 (the full period available). The HREF is natively produced on a 3-km grid; however, SPC Enhanced Thunderstorm forecasts are verified on a 40-km grid. To ensure the HREF Calibrated Thunder forecast probabilities remain consistent with those being issued by the SPC forecasters, all data fields contained within the HREF ensemble members were interpolated to a 40-km grid using a maximum nearest neighbor approach. Similarly, the number of hourly CG lightning flashes were obtained from the NLDN for the same 1 July 2017 – 1 January 2020 period and spatially mapped to a 40-km grid.

2.1 DERIVING THE HREF CALIBRATED THUNDER ALGORITHM

The first step in deriving the HREF Calibrated Thunder guidance was to identify which HREF data fields best correlated to the occurrence of at least one CG lightning flash. To accomplish this, all 00z and 12z HREF forecasts from 1 July 2017 – 1 July 2018 were compared to the NLDN gridded CG lightning flash data for the same time period. A Pearson correlation coefficient was then computed between the CG flash data and all data fields common across the HREF ensemble members. The resulting correlations were averaged to provide an ensemble mean correlation for each field. The total 1-hour accumulated precipitation (Tot Precip) was found to have the highest mean correlation to at least one CG lightning flash, with a correlation of 0.27. The derived radar reflectivity at 4 km above ground level (4 km REFL) had the second highest correlation of 0.26, and the most unstable lifted index (MU LI) exhibited the third highest correlation of 0.25.

Once the best correlated data fields were known, the next step was to develop a statistical algorithm which could convert the input fields into a probabilistic thunderstorm forecast. This was largely a trial-and-error process, the details of which are too long for this extended abstract (see Harrison et al. 2019). Ultimately the most successful algorithm was determined to be of the form:

$$w_1 P(X \geq t_1) + w_2 P(Y \geq t_2) + w_3 P(Z \geq t_3) \qquad \textbf{(1)}$$

where w_1, w_2, and w_3 represent weights summing to 1; X, Y, and Z are input data fields; and t_1, t_2, and t_3 are threshold values corresponding to the respective input fields. The probability function $P(\,)$ is defined as the fraction of HREF ensemble members where the inequality is true. As an example, consider a single grid point where five of the ten HREF members predict the 4 km REFL will be greater than 40 dBZ over a given four-hour period. Then

$$P(4\,km\,REFL \geq 40\,dBZ) = 0.50 = 50\%. \qquad \textbf{(2)}$$

The final probability of lightning predicted by the algorithm for that grid point is then the weighted average of the probabilities that each input field is greater than or equal to that field's respective threshold value.

The final step was to determine which combination of weights and thresholds provide the optimal forecast. This was accomplished by performing a grid search, where thunder forecasts were computed for 1 July 2017 – 1 July 2018 using a subset of every possible combination of weights and thresholds. The combination of hyperparameters that resulted in the highest

	4 km REFL	**Tot Precip**	**MU LI**
1-hour forecast	$t_1 \geq 40\ dBZ;\ w_1 = 0.6$	$t_2 \geq 1\ mm;\ w_2 = 0.3$	$t_3 \leq -1;\ w_3 = 0.1$
4-hour forecast	$t_1 \geq 40\ dBZ;\ w_1 = 0.6$	$t_2 \geq 2\ mm;\ w_2 = 0.3$	$t_3 \leq -1;\ w_3 = 0.1$
24-hour forecast	$t_1 \geq 40\ dBZ;\ w_1 = 0.6$	$t_2 \geq 1\ mm;\ w_2 = 0.4$	

Table 1: The best thresholds (t) and weights (w) for each input data field and each forecast time interval. MU LI was excluded from the 24-hour forecast due to strong diurnal variations in the parameter.

forecast critical success index (CSI) was selected as the optimal configuration. This process was performed independently for thunder forecasts of 1-hour, 4-hour, and 24-hour intervals, and the best weights and thresholds for each are provided in Table 1.

2.2 CALIBRATION

Calibration of the HREF probabilistic thunder guidance to be statistically reliable was performed by first generating thunder forecasts from 1 January 2018 through 1 January 2019. The raw forecast probabilities from the 1-year period were then stratified into 10% bins at each grid point, and the reliability of each category at each grid point was computed. For example, at a given grid point, the true probability of the 40% bin was defined as the fraction of 40% forecasts that verified with a lightning strike. If a given grid point received 40% probability forecasts 100 times throughout the year and lightning occurred at that grid point in 45 of those forecasts, then the true probability was 45% and the 40% bin had a reliability error of -5% (under-forecast). The reliability error was then saved for each grid point. This process was performed independently for the 1-hour, 4-hour, and 24-hour forecasts at every forecast hour. The end result is a 4-dimensional lookup table containing the mean reliability error at every grid point for every forecast hour, HREF cycle, and initial forecast probability.

Calibration is applied to new thunder forecasts by first matching the grid point, forecast hour, HREF cycle, and initial forecast probability to the corresponding reliability error in the lookup table. Once this is determined, the guidance is calibrated by simply adding that reliability error back to the original probability forecast at that grid point. For example, if at a given grid point the original guidance had a 47% probability and the mean reliability error for the 40% bin at that grid point at that forecast hour was -8% (under-forecast by 8%), then the final, calibrated probability for that grid point would be 47% + 8% = 55%.

2.3 INSTABILITY MASK

During initial testing of the HREF Calibrated Thunder guidance, forecasters and researchers identified a bias in the algorithm that would result in the prediction of thunder probabilities for locations that were subjectively analyzed to be unsupportive of deep convection or lightning. This most commonly occurred when several HREF members predicted moderate stratiform precipitation, which would activate the reflectivity and precipitation terms of the calibrated thunder equation. Although the lifted index is also considered when computing the probabilities, that term is weighted considerably less than the other two, and so a sufficiently large number of HREF members predicting high enough reflectivity and accumulated precipitation values could generate thunder probabilities even if the forecast environment was stable.

To correct this bias, an initial check was imposed on each member of the HREF to create an instability mask. The contribution from any HREF member that forecasts MU LI > 0 and 4 km REFL < 35 dBZ over the specified forecast period is set to zero when creating the probabilities for that grid point. As an example, consider a grid point where 8 of the 10 HREF members predict stratiform precipitation with a maximum 4 km REFL of 30 dBZ and a 4-hour accumulated precipitation total of 6.35 mm (0.25 in.) Only 1 of the 10 members predicts

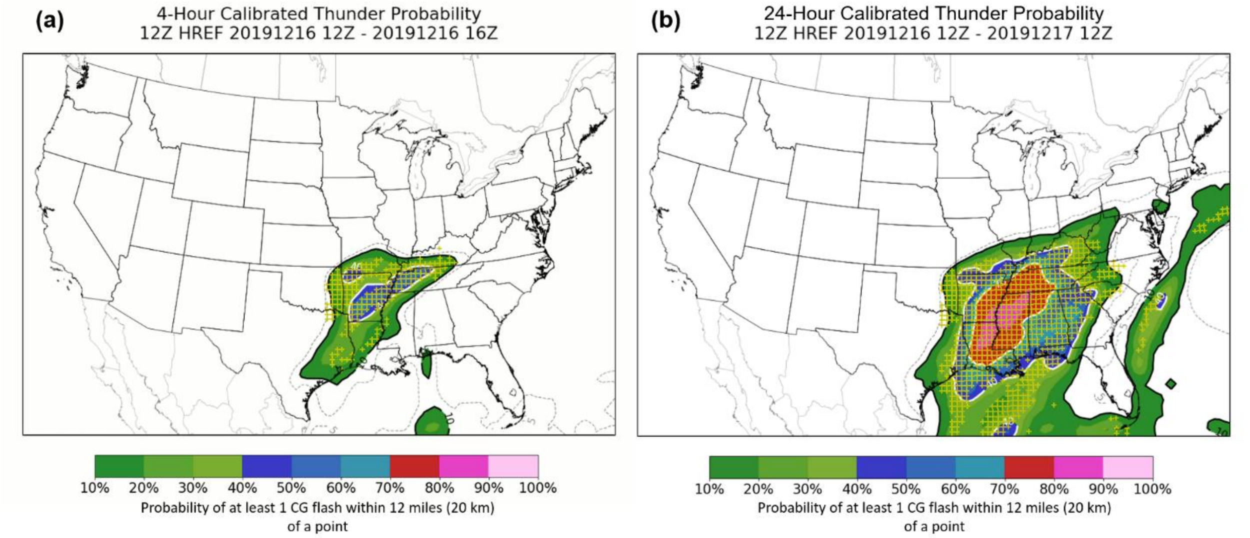

Figure 1: Calibrated Thunder (a) 4-hour and (b) 24-hour forecasts from the 12z HREF cycle on 16 December 2019. The yellow "+" symbols indicate grid points where there was at least one CG lightning flash detected during the valid forecast period.

MU LI < -1, while the others are all > 0. Without the instability mask, this grid point would be given a 25% probability of thunder, largely driven by the accumulated precipitation term. With the mask applied, however, all but one member would be set to zero in the calculation because the predicted reflectivity is < 35 dBZ and the MU LI is > 0. This would then produce a thunder probability of 4% for the grid point prior to calibration.

Calibrated 1-hour, 4-hour, and 24-hour thunder forecasts were regenerated for 1 July 2017 – 1 July 2018 with the new mask applied, and the resulting verification revealed a slight improvement in the overall performance of the guidance (not shown). Furthermore, anecdotal case studies found that the mask was successful at removing un-meteorological regions of low thunder probabilities, particularly in the Pacific Northwest.

3. VERIFICATION

Verification of the 00z and 12z HREF Calibrated Thunder products was performed on the 1-year independent dataset of 1 January 2019 – 1 January 2020. Calibrated 1-hour, 4-hour, and 24-hour thunder forecasts were generated for the full verification period, and the probabilities from each forecast were stratified into 10% bins. The forecasts were then compared to the NLDN CG lightning flash data for each forecast hour (Fig. 1), and the Probability of Detection (POD), False Alarm Ratio (FAR), CSI, and statistical reliability were computed for each bin. This process was then repeated for the equivalent SREF Calibrated Thunder products, and the results were compared as shown in Fig. 2.

All HREF Calibrated Thunder products outscored their SREF counterparts in terms of CSI during the 1-year period (Fig. 2a). The 24-hour HREF Calibrated Thunder forecast exhibited the greatest performance, with a maximum CSI of 0.46. In contrast, the SREF 24-hour thunder forecast had a maximum CSI of 0.31, or 0.15 less than the HREF guidance. Similarly, the 4-hour HREF Calibrated Thunder forecast had a maximum CSI of 0.29 (compared to 0.21 for the SREF), and the 1-hour forecast had a maximum CSI of 0.22 (compared to 0.13 for the SREF). These results indicate that the new HREF Calibrated Thunder guidance is a notable improvement over the original SREF Calibrated Thunder guidance, and supports the initial hypothesis that the addition of simulated radar reflectivity and other storm-attribute fields from a CAM may lead to improved probabilistic, calibrated thunderstorm predictions.

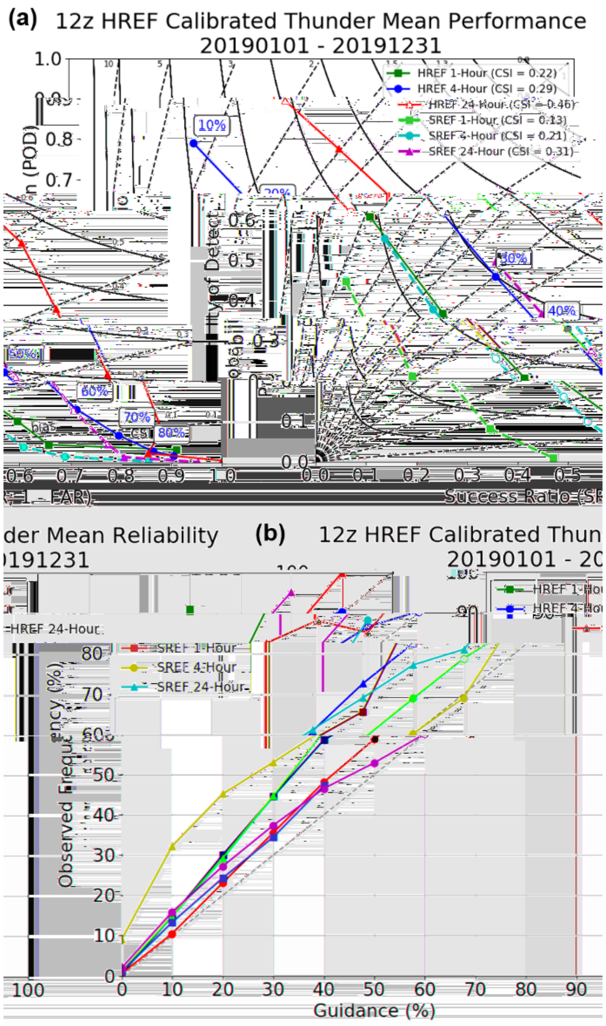

Figure 2: Comparison of the (a) performance and (b) reliability of the HREF and SREF Calibrated Thunder 1-hour, 4-hour, and 24-hour forecasts from 1 January 2019 – 1 January 2020.

Both the SREF and HREF Calibrated Thunder products exhibited similar behavior in terms of statistical reliability (Fig. 2b). For both versions of the guidance, the 4-hour calibrated thunder forecasts were generally the most reliable, with reduced reliability noted for the 1-hour and 24-hour forecasts. All forecast products tended to under-predict the true probability of lightning by approximately 5 – 15%, with the HREF 24-hour forecasts under-predicting by about 20% at probabilities > 40%. In general, the 1-hour and 24-hour forecast reliability decreased as the predicted probability increased, but this may be due in part to a smaller sample size at the higher probability range. Additional calibration of the HREF thunder products may help to further reduce these reliability errors.

4. CONCLUSION

A new suite of calibrated thunderstorm forecast products has been developed using a combination of simulated radar and environmental fields from the HREF. These products have been shown to skillfully predict the probability of at least one CG lightning flash over a given 1-hour, 4-hour, or 24-hour forecast period. In addition, the HREF Calibrated Thunder guidance has been shown to outperform the original calibrated thunder products from the non-convection-allowing SREF, which has been in use by the SPC for over a decade.

The HREF Calibrated Thunder suite has now been implemented operationally at the SPC. Initial feedback from forecasters has been positive, and the guidance is actively being used to help produce the daily Enhanced Thunderstorm forecasts. All HREF Calibrated Thunder products are now available to the public on the SPC website at https://www.spc.noaa.gov/exper/href/?model=href&product=guidance_thunder_hrefct_004h.

5. ACKNOWLEDGMENTS

This extended abstract was prepared by David Harrison with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

6. REFERENCES

Benjamin, S., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

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 Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.3.

Bright, D.R. and J.S. Grams, 2009: Short Range Ensemble Forecast (SREF) calibrated thunder probability forecasts: 2007-2008 verification and recent enhancements. *4th Conference on the Meteorological Applications of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc, 6.3.

Du, J. and Co-authors, 2014: NCEP regional ensemble update: current systems and planned storm-scale ensembles. Preprints, *26th Conf. on Wea. Forecasting*, Atlanta, GA, Amer. Meteor. Soc., J1.4.

Gallo, B. T., B. Roberts, I.L. Jirak, A. J. Clark, C. P. Kalb, and T. Jensen, 2018: Evaluating Potential Future Configurations of the High Resolution Ensemble Forecast System. *29th Conf. on Severe Local Storms*, Stowe, VT, Amer. Meteor. Soc., 76, https://ams.confex.com/ams/29SLS/webprogram/Paper348791.html.

Harrison, D., I. L. Jirak, and N. J. Nauslar, 2019: A Preliminary Investigation of the High-Resolution Ensemble Forecast (HREF) for Generating Calibrated Probabilistic Thunderstorm Forecasts. *9th Conference on the Meteorological Application of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc., 4.3. [Available online at https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/351851.]

Janjić, Z. I., and R. L. Gall, 2012: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB). Part 1: Dynamics. NCAR Tech. Note NCAR/TN-489+STR, 75 pp., https://doi.org/10.5065/D6WH2MZX.

Roberts, B., I.L. Jirak, A.J. Clark, S.J. Weiss, and J.S. Kain, 2019: PostProcessing and Visualization Techniques for Convection-Allowing Ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, https://doi.org/10.1175/BAMS-D-18-0041.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.