# 6A.1 PREDICTING PROBABILISTIC LIGHTNING FLASH DENSITY FROM THE HREF CALIBRATED THUNDER GUIDANCE

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

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

## 1. INTRODUCTION

The HREF Calibrated Thunder (HREFCT) guidance is a suite of forecast products developed by the Storm Prediction Center (SPC) to predict the probability of at least one cloud-to-ground (CG) lightning flash within 20 km (12 miles) of a point location. As described by Harrison et al. (2022), this guidance is derived from prognostic storm-scale attributes and environmental parameters produced by the 10-member High-Resolution Ensemble Forecast (HREF; Roberts et al. 2019, 2020) system. Specifically, the HREFCT post-processes information from each HREF member's simulated radar reflectivity at the -10℃ temperature level, period-total accumulated precipitation, and most-unstable four-layer lifted index to provide a calibrated probabilistic thunderstorm forecast. The HREFCT has been shown to be both skillful and reliable when predicting CG flashes in rolling 1-, 4-, and 24-hour intervals up to 48 hours in advance of a storm (Harrison et al. 2022), and the guidance significantly outperforms similar thunderstorm forecast products derived from the 26-member National Center for Environmental Prediction (NCEP) Short-Range Ensemble Forecast (SREF; Du et al. 2014). The HREFCT became operational on NCEP's Weather and Climate Operational Supercomputing System (WCOSS) on 11 May 2021, and is now being distributed to the public by the National Weather Service (NWS). Additionally, the HREFCT is widely used by SPC forecasters as a "first guess" when issuing thunderstorm forecast products.

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

Despite the operational success of the HREFCT, there are additional aspects of lightning prediction that are not covered by the current guidance. Recall that the HREFCT output is defined as the probability of at least one CG flash within 20 km of a location over a given time period. This forecast product is a measure of *confidence* that there will be CG lightning within a specified spatiotemporal domain, but it does not provide any information about *how much* CG lightning is expected in that domain. A forecast of 70% could potentially verify as one CG flash during the forecast period or as 200 flashes. Such information about the spatiotemporal density of CG flashes is of importance to many NWS core partners and has been directly requested through formal communication. CG flash density is particularly relevant to partners in the aviation (Yoshikawa and Ushio 2019; He et al. 2020), fire weather (Rorig and Ferguson 1999; Dowdy and Mills 2012), and emergency management (Lopez et al. 1997; Vagasky 2022) communities where frequent or spatially dense CG lightning may pose a greater risk to decision makers' constituents. Additionally, researchers at the SPC have hypothesized that reliable CG flash density guidance may have applications for severe and fire weather forecasting.

To address these needs, this study expands upon the existing HREFCT product suite to provide explicit probabilistic forecasts of ≥25, 50, 100, or 200 CG flashes within 20 km of a point location during a rolling 4-hour period. This guidance was developed by applying traditional machine-learning techniques to train and calibrate four gradient-boosted classifier models. This paper will document the development of the new lightning density guidance (section 2) before demonstrating

preliminary results and verification of the products (section 3).

## 2. DATA AND METHODS

The CG flash density guidance was derived using a combination of HREF prognostic fields and 4-hour forecasts from the existing HREFCT guidance. For this study, full 48-hour 0000z and 1200z HREF and HREFCT forecasts were obtained for 13 June 2020 - 1 December 2022. These dates represent the full period of operational HREFCT forecasts and reforecasts available within SPC's internal archive. The HREF is natively produced on a 3-km grid; however, the HREFCT forecasts are produced and verified on the 40-km NCEP 212 grid. To ensure any derived CG flash density forecasts remain consistent with the existing HREFCT guidance, all fields contained within the HREF ensemble members were interpolated to the NCEP 212 grid using a maximum nearest neighbor approach. Similarly, hourly CG lightning flashes were obtained from the National Lightning Detection Network (NLDN) for the same 13 June 2020 – 1 December 2022 period and spatially mapped to the 40-km grid. This gridded hourly lightning data was used as the truth dataset for this study.

### 2.1 *Feature Engineering*

Before performing any machine learning, it was first necessary to identify what variables would be used as input to the models. To this end, this study followed a similar approach to that implemented by Harrison et al. (2022) when deriving the original HREFCT algorithm. First, the number of NLDN CG lightning observations were tallied at each grid point in rolling 4-hour windows from 13 June 2020 - 13 June 2021. The 4-hour lightning counts were then converted into four binary fields, where a value of 1 was assigned to a grid point if the number of observed CG flashes was greater than or equal to 25, 50, 100, or 200 respectively. Next, the first 24 forecast hours of all 00z and 12z HREF cycles from 13 June 2020 - 13 June 2021 were compared to the corresponding binarized CG lightning observations, and a Pearson correlation coefficient was computed between the observed frequency of each

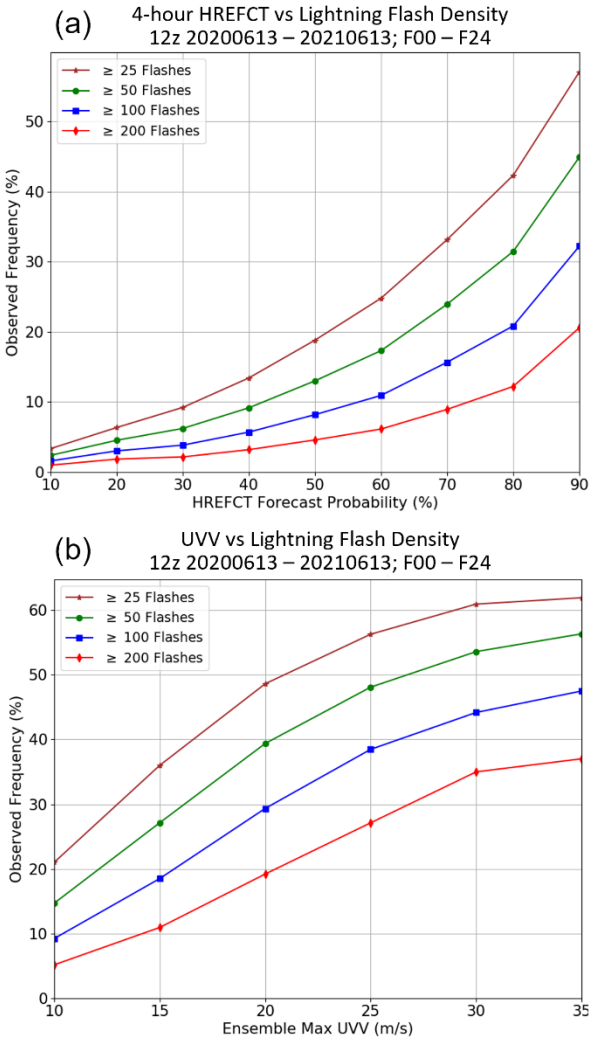


**Figure 1:** Relationship between the observed frequency of lightning density and (a) the 4-hour HREFCT forecast, and (b) the HREF ensemble-maximum updraft vertical velocity.

flash density threshold and all prognostic fields common across the HREF members. Finally, the correlations were averaged to produce an ensemble mean correlation for each field, and the relationships were plotted for additional evaluation (Fig. 1).

During feature selection, the authors hypothesized that additional predictive information may be derived by leveraging the smaller spatial resolution of the HREF's native 3-km grid. When converting the HREF prognostic fields from their 3-km grid to the 40-km NCEP 212 grid, each 3-km grid point is mapped to the nearest point in the 40-km grid. As such, each 40-km point is associated

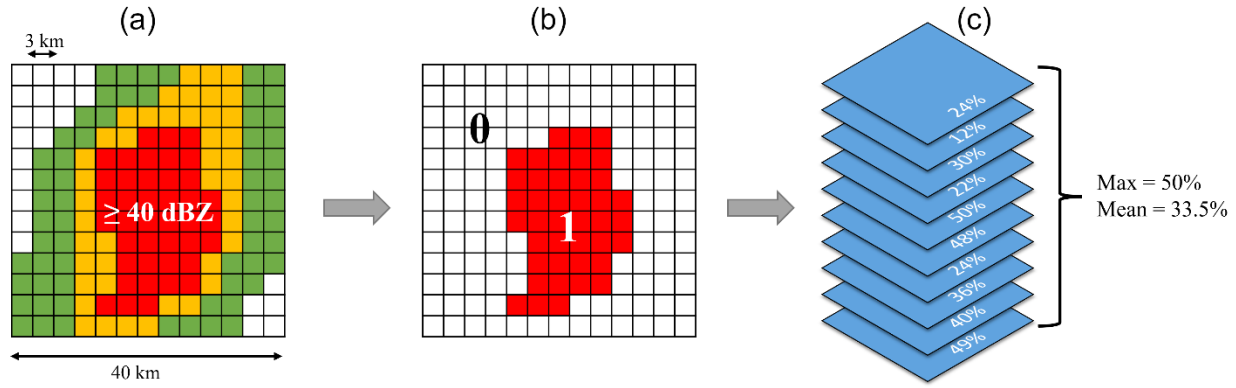


**Figure 2:** Example demonstrating how the ensemble mean and maximum fraction coverage is computed for $Z_{-10C} \geq 40\ dBZ$. The blue squares in panel (c) represent a single 40-km grid point compared across all 10 HREF members. See the text for more information.

with multiple values from the 3-km field. Previously, the maximum value of this data was used to fill in the 40-km grid; however, this discards information about the distribution of the higher-resolution data within each 40-km grid point. To determine if such information could be used for predicting CG flash density, we first identified all 3-km points mapped to a single 40-km grid point where the HREF simulated radar reflectivity at the -10℃ temperature level was greater than or equal to 40 dBZ ($Z_{-10C} \geq 40\ dBZ$; Fig. 2a). This combination of field and threshold was chosen based on its predictive skill demonstrated when deriving the original HREFCT guidance (Harrison et al. 2022). These values were then converted to binary, such that the 3-km data was 1 where the inequality was true and 0 everywhere else (Fig. 2b). Next, the mean of the binarized data was calculated for each 40-km point, yielding the fraction of 3-km points mapped to the 40-km point where the simulated reflectivity met the 40 dBZ threshold. For example, if there were 169 3-km grid points mapped to a single 40-km point and $Z_{-10C} \geq 40\ dBZ$ at 41 of those 3-km points, then the inequality was met in $41/169 = 0.24 = 24\%$ of that 40-km grid point. This process was repeated for every 40-km grid point and for all 10 HREF members individually. Finally, a point-by-point mean and maximum was taken across the entire HREF ensemble, resulting in the ensemble mean and maximum fractional coverage of $Z_{-10C} \geq 40\ dBZ$ respectively (Fig. 2c).

As before, the mean and maximum fractional coverage of $Z_{-10C} \geq 40\ dBZ$ were compared to the four CG flash density thresholds as shown in Fig. 3. Both features demonstrated strong positive correlation to the observed frequency of CG flashes, and this was particularly true for the mean fractional coverage data which exhibited a near one-to-one relationship with the 50-flash threshold. Other strongly related fields chosen as inputs to the machine learning models include the ensemble-maximum updraft vertical velocity (UVV) and the 4-hour forecasts from the original HREFCT guidance (Fig. 1).

### 2.2 *Model Design*

Prior to model development, the dataset was separated into independent training, calibration, and testing sets. Examples from 13 June 2020 - 13 June 2021 were selected for the combined training and validation set, 13 June 2021 - 13 June 2022 was used for model calibration, and the test set contained examples from 13 June 2022 - 1 December 2022. Ten days were withheld between datasets to avoid cross-contamination from temporal autocorrelation within the features. To avoid oversaturating the dataset with trivial non-convective examples, the training set was populated by sampling points that fell within at least an SPC Day 1 Convective Outlook General Thunderstorm risk area. The SPC General Thunderstorm risk area denotes where SPC forecasters expect at least a 10% probability of lightning within 25 miles of a location during the convective day (12z - 12z). Masking the training

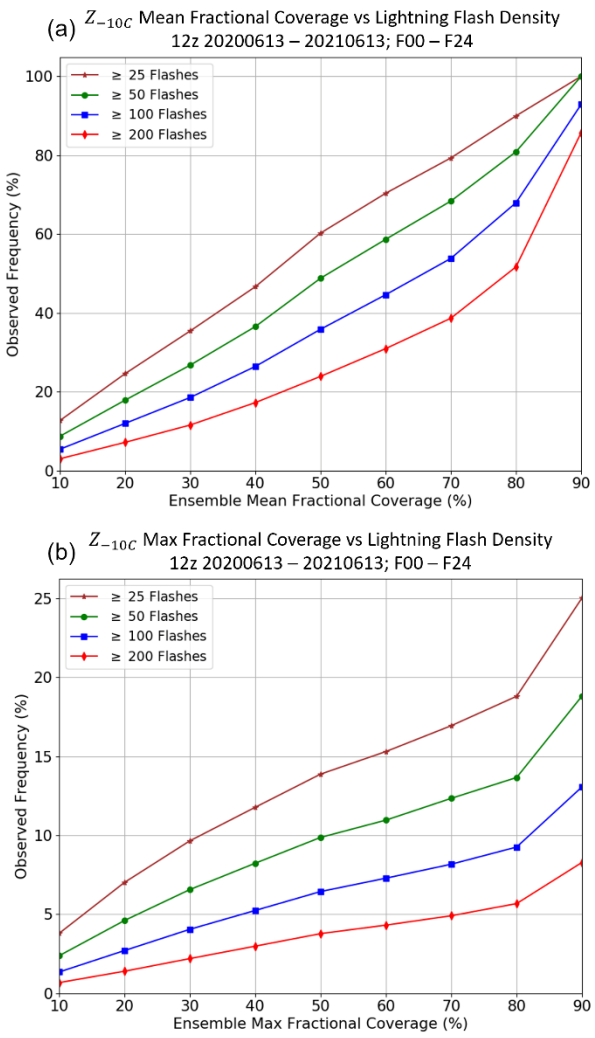


**Figure 3:** Relationship between the observed frequency of lightning density and the ensemble (a) mean and (b) maximum fractional coverage of $Z_{-10C} \geq 40\ dBZ$.

data by the General Thunderstorm outlook is intended to produce a set of negative examples representative of a diverse range of environmental conditions where either lightning did not occur, or the number of CG flashes did not meet the specified threshold. Because positive examples (where the CG flash density threshold was met) are relatively rare compared to the negative case, it was also necessary to undersample the negative examples such that there was an equal number of positive and negative classes in the training set. The calibration and test sets were not masked and contained data from all points within the CONUS.

Four gradient-boosted classifiers (GBC) were independently trained to predict the probability of ≥ 25, 50, 100, and 200 CG flashes within 20 km of a point location over a given 4-hour period respectively. A randomized grid search with 5-fold cross validation was used to train and tune each model's hyperparameters, and four isotonic regressors were independently trained and applied to calibrate the models' probabilistic outputs. This resulted in one GBC and isotonic regressor combination each for the 25-, 50-, 100-, and 200-flash thresholds.

## 3. VERIFICATION

Preliminary verification of the CG flash density guidance was performed on the 5-month test dataset of 13 June 2022 - 1 December 2022. Forecasts of ≥25, 50, 100, and 200 CG flashes were generated in rolling 4-hour windows for each 0000z and 1200z HREF cycle in the verification period out to 48 forecast hours. Each forecast was spatially smoothed using a gaussian kernel ($\sigma$=40 km), and the probabilities were then capped such that they couldn't exceed the original 4-hour HREFCT probabilistic forecast. As the HREFCT predicts the probability of at least one CG flash during the 4-hour forecast period, it would not mathematically make sense for the probability of at least 25, 50, 100, or 200 flashes to be greater than the probability of at least one flash during the same interval. Applying the HREFCT as an upper limit keeps the new guidance consistent with the original HREFCT forecast products and serves as a quality check against anomalous output from the GBC models. The resulting predictions were stratified into 10% probability bins and then compared to the observed 4-hour CG flash counts at each forecast hour (Fig. 4). The probability of detection (POD), false alarm (FAR), critical success index (CSI), and statistical reliability were then calculated for each probability bin and for each flash density threshold as shown in Fig. 5.

The GBC trained on the 25-flash density threshold exhibited the greatest preliminary performance metrics, with a maximum CSI of 0.19 (Fig. 5a). This was followed by the 50-flash model with a maximum CSI of 0.15 and the 100-flash model at 0.10. The GBC trained on the 200-flash

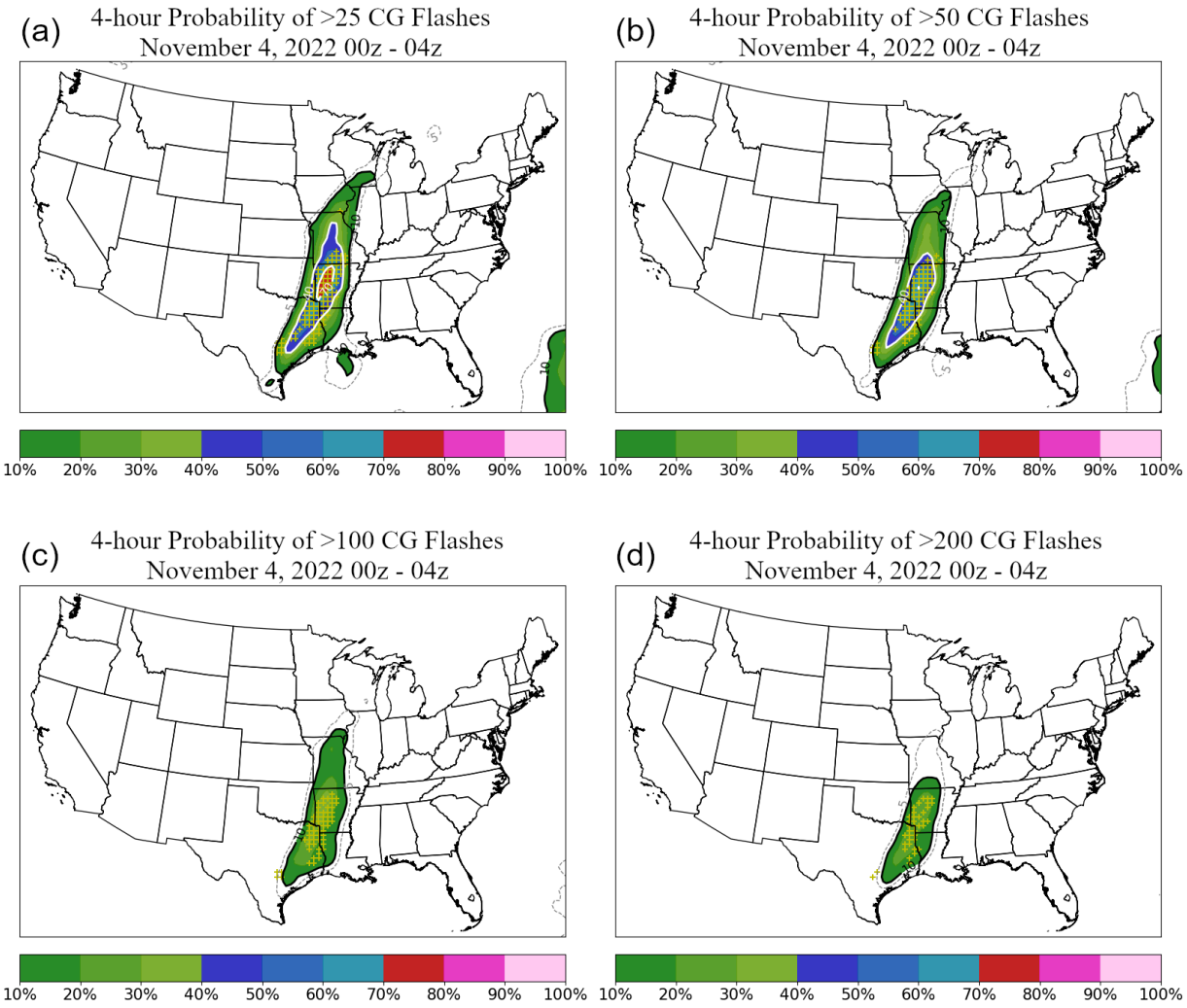


**Figure 4: P**robability of at least (a) 25, (b) 50, (c) 100, and (d) 200 CG flashes within 20 km of a point for 4 November 2022 0000z – 0400z. Yellow "+" markers indicate where the number of CG flashes exceeded the specified threshold.

density threshold demonstrated the weakest performance on average with a maximum CSI of about 0.07. This notable decrease in performance at the 100- and 200-flash thresholds may be due in part to the limited sample sizes at those higher densities. Forecast probabilities were found to generally decrease as the CG flash density threshold increased, and the probability of ≥ 200 flashes over a 4-hour period rarely exceeded 30%. Indeed, there were anecdotally many days in the test dataset where the 100- and 200-flash models did not produce probabilities >10% while the lower threshold models did. This is expected and desirable model behavior, as the higher flash densities are considerably rarer than the lower thresholds. Within the training and calibration datasets, there were a combined 2,880,963 cases where the lightning density exceeded 25 CG flashes within 20 km of a point during a 4-hour period. This tally decreased to 1,979,809 for the 50-flash threshold and 1,249,658 at the 100-flash threshold. There were only 710,765 grid points in the 2-year training and calibration datasets where the number of CG flashes exceeded 200 during a 4-hour period.

All GBC models were found to be relatively well calibrated during the 5-month verification period, with forecast probabilities generally falling within

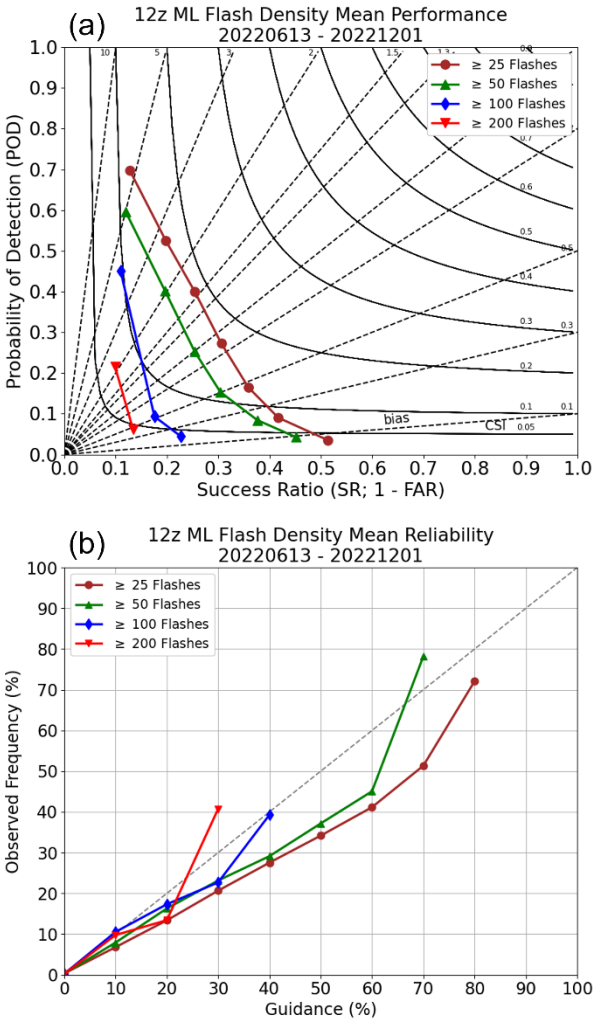


**Figure 5:** Comparison of the (a) performance and (b) reliability of the gradient-boosted classifiers trained on 25, 50, 100, and 200 flash density thresholds.

about 10% of the observed frequency on average (Fig. 5b). The 25- and 50-flash models did tend to generally overforecast compared to observations, particularly at probabilities >50%. The reliability error at these higher probabilities was found to exceed 15% for the 50-flash model, and approached about 20% for the 25-flash model. This trend was not as notable for the 100- and 200-flash models, but this again may be in part due to the smaller sample sizes and lower forecast probabilities produced by those models. Future evaluations will include a longer period of study and should increase the available sample size to improve the robustness of these results. Further study is also required to determine how the model performs in different regions of the CONUS, and a region-specific calibration may be warranted to improve forecast reliability.

## 4. CONCLUSION

Four gradient-boosted classifiers have been trained and calibrated to predict the probability of ≥ 25, 50, 100, or 200 CG flashes within 20 km of a location over a given 4-hour window. The resulting forecasts demonstrated moderate skill at identifying the time and location where each flash density threshold will be met, though the guidance tended to somewhat overforecast the lower flash-density thresholds. Future work will continue to evaluate the model guidance, assess regional performance and reliability, and explore additional predictive variables that may lead to more robust forecasts. The CG flash density guidance is expected to be evaluated by SPC forecasters during the upcoming warm season and will eventually be added to the publicly available HREFCT suite of forecast products.

## 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 #NA21OAR4320204, 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

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.

Harrison, D. R., M. S. Elliott, I. L. Jirak, and P. T. Marsh, 2022: Utilizing the High-Resolution Ensemble Forecast System to produce calibrated probabilistic thunderstorm guidance. *Wea.*

*Forecasting*, **37**, 1103–1115, https://doi.org/10.1175/WAF-D-22-0001.1.

He, Y., S. Lindbergh, C. Graves, and J. Rakas, 2020: Airport Exposure to Lightning Strike Hazard in the Contiguous United States. *Risk Analysis*, **41**, 1323 – 1344, https://doi.org/10.1111/risa.13630.

López, R. E., R. L. Holle, A. I. Watson, and J. Skindlov, 1997: Spatial and temporal distributions of lightning over Arizona from a power utility perspective. *J. Appl. Meteor.*, **36**, 825–831, https://doi.org/10.1175/1520-0450-36.6.825.

Roberts, B., B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What does a convection-allowing ensemble of opportunity buy us in forecasting thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, https://doi.org/10.1175/WAF-D-20-0069.1.

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.

Rorig, M. L., and S. A. Ferguson, 2002: The 2000 fire season: Lightning-caused fires. *J. App. Meteor.*, **41**, 786–791, https://doi.org/10.1175/1520-0450(1999)038<1565:COLAWF>2.0.CO;2.

Vagasky, C., 2022: One Strike, You're Out: Lightning during Professional Baseball Games. *Wea. Clim. Soc.*, **14**, 365 – 371, https://doi.org/10.1175/WCAS-D-21-0099.1.

Yoshikawa, E. and T. Ushio, 2019: Tactical Decision-Making Support Information for Aircraft Lightning Avoidance: Feasibility Study in Area of Winter Lightning. *Bull. Amer. Meteor. Soc.,* **100**, 1443 – 1452, https://doi.org/10.1175/BAMS-D-18-0078.1.