Utilizing the High-Resolution Ensemble Forecast System to Produce Calibrated Probabilistic Thunderstorm Guidance

DAVID R. HARRISON,[a,b] MATTHEW S. ELLIOTT,[b] ISRAEL L. JIRAK,[b] AND PATRICK T. MARSH[b]

[a] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/NWS/Storm Prediction Center, Norman, Oklahoma*

(Manuscript received 3 January 2022, in final form 29 March 2022)

ABSTRACT: The High-Resolution Ensemble Forecast system (HREF) calibrated thunder guidance is a suite of probabilistic forecast products designed to predict the likelihood of at least one cloud-to-ground (CG) lightning flash within 20 km (12 miles) of a point during a given 1-, 4-, and 24-h time interval. This guidance takes advantage of a combination of storm attribute and environmental fields produced by the convection-allowing HREF to objectively improve upon lightning forecasts generated by the non-convection-allowing Short-Range Ensemble Forecast system (SREF). This study provides an overview of how the HREF calibrated thunder guidance was developed and calibrated to be statistically reliable against observed CG lightning flashes recorded by the National Lightning Detection Network (NLDN). Performance metrics for the 1-, 4-, and 24-h guidance are provided and compared to the respective SREF calibrated probabilistic lightning forecasts. The HREF calibrated thunder guidance has been implemented operationally within the National Weather Service and is now available to the public.

SIGNIFICANCE STATEMENT: The NOAA Storm Prediction Center has created a suite of new calibrated probabilistic thunderstorm guidance products from a convection-allowing model ensemble, the HREF. The new guidance is a notable improvement over the long-running SREF calibrated thunder guidance and is now operational across the National Weather Service.

KEYWORDS: Atmosphere; Lightning; Thunderstorms; Ensembles; Forecast verification/skill; Forecasting; Numerical weather prediction/forecasting; Operational forecasting; Probability forecasts/models/distribution

1. Introduction

The stated mission of the National Weather Service's (NWS) Storm Prediction Center (SPC) is to deliver timely and accurate forecast information about tornadoes, severe thunderstorms, lightning, wildfires, and winter weather across the contiguous United States (CONUS) to protect lives and property (SPC 2021a). As part of this mission, the SPC is responsible for issuing forecast products that indicate where and when cloud-to-ground (CG) lightning is anticipated. One such product is the Thunderstorm Outlook, which depicts the probability of thunderstorms across the CONUS in 4- or 8-h periods (Bright and Grams 2009; SPC 2021b) for the upcoming or current convective day. Specifically, these forecasts represent the probability of at least one CG lightning flash within 20 km (12 miles) of a point location during the valid forecast period. The increased temporal resolution of the Thunderstorm Outlook aids NWS forecasters and partners in time-sensitive decisions related to thunderstorms and lightning hazards (Stough et al. 2012; SPC 2021b).

Accurately predicting the timing and location of thunderstorms across the CONUS can often be a time consuming and mentally taxing challenge for forecasters. Many studies have been published over the past five decades showcasing a variety of automated, gridded thunderstorm probability guidance intended to aid in the prediction of lightning hazards. One of the earliest of these studies dates to the 1970s, when Reap and Foster (1979) created a multiple screening regression to generate medium-range thunderstorm probability forecasts from model output statistics (MOS; Glahn and Lowry 1972). More recent approaches to probabilistic lightning prediction have been incorporated within the NWS's National Blend of Models (NBM), a project intended to generate calibrated, high-resolution forecast guidance from statistically postprocessed multimodel ensembles (Tew et al. 2016; Hamill et al. 2017; Craven et al. 2018). Probabilistic lightning forecast products currently contained within the operational NBM include MOS guidance from the Global Forecast System (GFS; Kanamitsu et al. 1991; Hughes 2001), the North American Mesoscale Forecast System (NAM; Rogers et al. 2005; Maloney et al. 2009), and the European Centre for Medium-Range Weather Forecasts (ECMWF; Shafer and Rudack 2015) model. MOS thunderstorm forecasts in the NBM are primarily derived from a regression of deterministic large-scale numerical weather prediction (NWP) precipitation forecasts and lightning climatology in 3-h intervals. Although the GFS, NAM, and ECMWF MOS thunderstorm forecasts have demonstrated skill, their dependence on large-scale NWP forecasts ultimately limits the spatial and temporal detail of the guidance. Additionally, the reliance of MOS schemes on climatology tends to reduce the forecast skill of spatiotemporally less common lightning events (Shafer and Fuelberg 2008).

Corresponding author: David R. Harrison, david.harrison@noaa.gov

DOI: 10.1175/WAF-D-22-0001.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

The NBM also includes probabilistic lightning forecasts from the Localized Aviation MOS Program (LAMP; Charba et al. 2019), which combines the aforementioned deterministic large-scale MOS products with fine-scale model output from the deterministic High Resolution Rapid Refresh (HRRR; Benjamin et al. 2016) model, total lightning observations from the National Lightning Detection Network (NLDN; Cummins et al. 1998), and radar reflectivity from the Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016; Zhang et al. 2016) system. Objective verification of the LAMP by Charba et al. (2019) has shown the guidance performs very well in the first forecast hour via extrapolation of MRMS and NLDN observations. However, the influence of observations on the guidance was found to sharply decrease within the first four forecast hours. As such, much of the LAMP's forecast skill comes from shorter lead-time forecasts, with notably decreasing skill at longer lead times.

A fifth lightning forecast product contained within the NBM is derived by SPC through postprocessing the 26-member National Centers for Environmental Prediction (NCEP) Short-Range Ensemble Forecast (SREF; Du et al. 2014). For nearly two decades, SPC forecasters have largely utilized the SREF calibrated thunder guidance (SREFCT) as their "first guess" when generating Thunderstorm Outlooks and other thunderstorm forecast products. The SREFCT aids in delineating areas favorable for CG lightning by identifying regions of appropriate instability and thermodynamic factors coinciding with precipitation within the SREF forecast grid (Bright et al. 2005). Specifically, the SREFCT highlights points within the SREF's 40-km NCEP 212 grid where the forecast LCL ≥ -10°C, CAPE > 100 J kg^{-1} in the 0° to -20°C layer, and the equilibrium level temperature is ≤ -20°C. As described by Bright et al. (2005), these parameters are believed to approximately delineate regions where mixed-phase hydrometeors are present above the charge-reversal temperature and coincide with updrafts sufficiently strong enough to replenish supercooled liquid above the charge-reversal zone (Saunders 1993). The SREFCT probability of at least one CG lightning flash within 20 km (12 miles) of a point location is derived by first combining the above environmental parameters into a single probabilistic ensemble composite known as the cloud physics thunder parameter (CPTP; Bright et al. 2005). The calibrated lightning probability is then obtained by calculating the relative frequency of CG flashes observed by the NLDN given the predicted CPTP probability and the SREF ensemble probability of accumulated precipitation ≥ 0.01 in. (0.254 mm) during the valid forecast period.

Verification studies of the SREFCT (e.g., Bright and Grams 2009) have shown that the algorithm provides reliable and skillful guidance, particularly as a "first-guess" product for forecasters when producing Thunderstorm Outlooks. However, the guidance does exhibit a notable bias toward underforecasting CG lightning probabilities in the Plains and Gulf Stream, especially during the warm season when nocturnal convection is more prevalent. Forecasters have also noted a tendency to overforecast lightning potential along the California coast and in the Pacific Northwest. Bright and Grams (2009) speculate that these biases may be in part due to the SREF's inability to explicitly resolve convection. For example, over the ocean it is common for the SREFCT's thermodynamic parameters to be met despite a dearth of convective precipitation. In these cases, the guidance may undesirably generate probabilities for gridscale precipitation originating from low clouds in the model's marine boundary layer (Bright and Grams 2009).

Given the apparent limitations of the SREFCT and other probabilistic lightning guidance, the authors hypothesized that the addition of simulated radar reflectivity and other storm-attribute fields from an ensemble of convection-allowing models (CAMs) may lead to improved probabilistic thunderstorm predictions. To this end, a new suite of probabilistic thunderstorm guidance products has been derived from the NCEP High-Resolution Ensemble Forecast (HREF; Roberts et al. 2019) system and implemented operationally across the NWS. This manuscript will document the design of the HREF Calibrated Thunder guidance (section 2) and then compare the performance and reliability metrics of the new guidance to that of the original SREFCT (section 3).

2. Data and methods

The HREF calibrated thunder (HREFCT) forecast products were derived using prognostic fields from the operational HREF version 2 (HREFv2) and an experimental version of the HREF (HREFv2.1) that was tested internally at SPC (Roberts et al. 2020). The HREFv2 is composed of eight ensemble members with four deterministic CAM configurations represented by the High-Resolution Window Advanced Research version of the Weather Research and Forecast Model (HRW ARW; Skamarock et al. 2008), the National Severe Storms Laboratory version of the ARW (HRW NSSL; Kain et al. 2010), the Nonhydrostatic Multiscale Model on the B Grid (HRW NMMB; Janjić and Gall 2012), and the 3-km NAM Nest (Rogers et al. 2017). Each configuration is represented twice within the HREFv2 ensemble by including 12-h time-lagged initializations of each member. A full list of the model cores, boundary conditions, microphysics schemes, and PBL schemes of each of the eight members is provided by Roberts et al. (2019), their Table 1.

The HREFv2.1 utilized in this study adds the operational HRRR and its 6-h 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). Note that as of 11 May 2021, the operational version of the HREF version 3 (HREFv3) replaces the NMMB member of the ensemble with a Finite Volume Cubed Sphere (FV3) model and extends the temporal range of the HRRR to 48 h (EMC 2021; NWS 2021). However, these changes were not available at the time of this study and thus were not included during the initial design and testing. This is discussed further in section 3.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

TABLE 1. HREF prognostic fields with the greatest ensemble mean Pearson correlation to 1-h NLDN CG lightning flashes computed between 1 Jul 2017 and 1 Jul 2019.

	Ensemble mean Pearson correlation
Total accumulated QPF	0.14
MUCAPE	0.13
Derived radar reflectivity at −10°C	0.12
Derived radar reflectivity at 4 km AGL	0.12
Maximum composite reflectivity	0.11
Precipitable water	0.10
Specific humidity	0.07
MU LI	0.06

Both 0000 and 1200 UTC cycles of the HREFv2 were obtained for 1 July 2017–1 January 2019, and the HREFv2.1 cycles were collected for 1 January 2019–11 May 2021 (the full period available). The HREF is natively produced on a 3-km grid; however, SPC Thunderstorm Outlooks are verified on the 40-km NCEP 212 grid. To ensure the HREFCT forecast probabilities remain consistent with those being issued by the SPC forecasters, all prognostic fields within the HREF ensemble members were remapped to the 212 grid using a nearest neighbor maximum, minimum, or average, depending on the variable (Mesinger et al. 1990; Mesinger 1996; Accadia et al. 2003). For example, a nearest neighbor minimum was used to remap lifted index forecasts because lower values indicate increased instability, while a nearest neighbor maximum was used to remap 1-h accumulated precipitation. Additionally, observed hourly CG lightning flashes were obtained from the NLDN for the 1 July 2017–11 May 2021 period and spatially mapped to the same 40-km grid.

a. The HREF calibrated thunder algorithm

Initial development of the HREFCT attempted to build upon the success of the SREFCT by first focusing on the same environmental parameters used in the existing guidance. However, recreating the SREFCT's algorithm within the HREF framework quickly proved unsuccessful as some HREF members lack the fields necessary to compute the equilibrium level temperature or CAPE in the 0° to −20°C layer. Instead, the HREFCT guidance was derived from scratch, with the first step to identify which HREF prognostic fields best correlate to the occurrence of at least one CG lightning flash. To accomplish this, the first 24 forecast hours of all 0000 and 1200 UTC HREF cycles from 1 July 2017 to 1 July 2019 were compared to the corresponding NLDN gridded CG lightning observations. A Pearson correlation coefficient was then computed between the CG flash observations and all prognostic fields common across the HREF members. The resulting correlations from each member were averaged to provide an ensemble mean correlation for each field. Total accumulated QPF was found to have the highest mean correlation to at least one CG lightning flash, with a value of 0.14. Most unstable CAPE (MUCAPE) had the second highest mean correlation, followed by the derived radar reflectivity at −10°C, derived radar reflectivity at 4 km above ground level, maximum 1-h composite reflectivity, precipitable water, specific humidity, and the most unstable four-layer lifted index (MU LI). Correlations for each field are provided in Table 1.

Once the best correlated prognostic fields were identified, the next step was to develop an algorithm to convert the data into a probabilistic thunderstorm forecast. Several regression analyses were tested during this stage and with various combinations of the aforementioned fields. Hourly probabilistic thunder forecasts were created from each algorithm/input combination for the first 24 forecast hours from the 0000 and 1200 UTC HREF cycles between 1 July 2017 and 1 July 2019. The mean critical success index (CSI) was computed for each 1-h forecast, and this process was iteratively repeated until the best combination (i.e., the combination with the greatest mean CSI) of algorithm and prognostic fields was subjectively determined. Ultimately, the derived radar reflectivity at −10°C ($Z_{-10°C}$), total accumulated QPF (QPF_{accum}), and MU LI were found to be the most successful combination of prognostic fields when paired with a linear regression model of the form:

$$w_1P(X \geq t_1) + w_2P(Y \geq t_2) + w_3P(Z \geq t_3). \quad (1)$$

Here, w_1, w_2, and w_3 represent weights summing to 1; X, Y, and Z are HREF prognostic fields; and t_1, t_2, and t_3 are threshold values corresponding to the respective HREF 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 $Z_{-10°C}$ will be greater than a threshold of 40 dBZ over a given 1-h period. Then $P(Z_{-10°C} \geq 40\ \text{dBZ}) = 5/10 = 50\%$. The probability of lightning predicted by the algorithm for a given grid point is then the weighted average of the probabilities that each prognostic field meets or exceeds its respective threshold value.

The final step in the initial derivation was to determine which combination of weights and thresholds provide the optimal forecast. This was accomplished by performing a randomized grid search (Bergstra and Bengio 2012), where thunder forecasts were again computed for 1 July 2017–1 July 2019 using a random subset of every possible combination of weights and thresholds. The combination of hyperparameters that resulted in the greatest forecast CSI was selected as the optimal configuration. This optimization was performed independently for rolling forecast windows of 1- and 4-h intervals, where the 4-h prediction at a given forecast hour represents the cumulative probability of at least one CG flash within 20 km (12 miles) of a point location over the previous 4 h. For example, the 4-h forecast at forecast hour f04 represents the f00–f04 period. Additionally, 24-h forecasts were generated and optimized for each convective day (1200–1200 UTC) contained within each HREF cycle (0000 UTC HREF f12–f36; 1200 UTC HREF f00–f24 and f24–f48.) HREFCT forecasts for intervals greater than one hour were computed using the maximum or minimum values of each input variable over the specified period. The best weights and thresholds for each forecast interval are provided in Table 2.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

TABLE 2. The best thresholds (t) and weights (w) for each HREF prognostic field and forecast time interval. MU LI was excluded from the 24-h forecast due to strong diurnal variations in the parameter.

	$Z_{-10°C}$; Z_{4kmAGL}	QPF_{accum}	MU LI
1-h forecast	$t_1 \geq 40$ dBZ; $w_1 = 0.6$	$t_2 \geq 1$ mm; $w_2 = 0.3$	$T_3 \leq -3$; $w_3 = 0.1$
4-h 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-h forecast	$t_1 \geq 40$ dBZ; $w_1 = 0.6$	$t_2 \geq 2$ mm; $w_2 = 0.4$	

Though the HREF parameters are not identical to those used in the SREFCT algorithm, the fields and thresholds chosen for the HREFCT formula capture many of the same environmental conditions and physical processes highlighted by the original guidance. For example, the thresholds chosen for MU LI broadly indicate where lapse rates may be steep enough to support sustained updrafts necessary to replenish supercooled liquid above the charge-reversal zone, and $Z_{-10°C}$ might be considered an approximation of mixed-phase hydrometeors present near or above the charge-reversal temperature. Note that some members of the HREFv2 did not initially provide $Z_{-10°C}$, so the derived radar reflectivity at 4 km above ground level (Z_{4kmAGL}) was used as a proxy with the same weights and thresholds prior to the implementation of the HREFv2.1. The improved performance of MU LI over MUCAPE was an unexpected result during the algorithm derivation, as MUCAPE exhibited much higher correlation to CG flashes (Table 1). Anecdotally, forecasts that utilized MUCAPE instead of MU LI tended to overforecast the spatial coverage of lightning, particularly in marginally unstable or capped environments. Perhaps the ability of MU LI to represent both stable and unstable environments as a single parameter gives it an advantage over MUCAPE in the HREFCT algorithm, as MUCAPE would need to be paired with another variable such as MUCIN to provide information about stable layers within the thermodynamic profile.

b. Calibration

Calibration of the HREF probabilistic thunder guidance to be statistically reliable was performed by first generating thunder forecasts from 13 June 2019 to 13 June 2020. These dates were chosen for calibration to avoid inconsistencies in the HREF members that were present during the initial transition from HREFv2 to HREFv2.1. Noise in the raw probability fields was removed by applying a 2D Gaussian filter (σ = 80 km) to spatially smooth the forecasts. The smoothed probabilities from the 1-yr period were then stratified into 10% bins centered on every 10% (5%–15%, 15%–25%, etc.), and the reliability of each bin at each grid point was computed (Fig. 1a). For example, at a given grid point, the true probability of the 40% bin was defined as the fraction of 35%–45% forecasts that verified with at least one observed CG lightning flash. If a given grid point received 40% probability forecasts 100 times throughout the year and lightning occurred at that grid point in 30 of those forecasts, then the true probability was 30% and the 40% bin had a reliability error of +10% (an overforecast). The reliability error was then recorded for each grid point. This process was performed independently for the 0000 and 1200 UTC HREF cycles, the 1-, 4-, and 24-h forecast products, and for each of the HREF's 48 forecast hours. Thus, the calibration step resulted in a five-dimensional lookup table containing the mean reliability error at every grid point for every forecast hour, HREF cycle, and binned forecast probability.

This calibration process revealed a systematic bias in the uncalibrated HREFCT probabilities that generally led to an overforecast of CG lightning flashes in the 1- and 4-h forecast products and an underforecast in the 24-h product (Fig. 1b). The 1-h probabilities exhibited a consistent mean reliability error of about +5% through at least the first 24 forecast hours, while the 4-h uncalibrated guidance had an error from +5% to +10%. The 24-h uncalibrated guidance averaged an underforecast from −5% to −10%. The reliability errors of the 1- and 4-h forecasts varied considerably in the last 18 forecast hours, likely due in part to predictability error in the spatial placement of convection at longer lead times. Forecast probabilities between 25% and 55% exhibited the greatest mean reliability error for all three products (not shown). Both the 1- and 4-h uncalibrated guidance averaged an overforecast from +10% to +15% at these probabilities, while the 24-h guidance underforecast by up to −5% on average. Notably, all three products were found to slightly underforecast at probabilities $< 5\%$ and overforecast at probabilities $> 95\%$ on average. As such, the resulting calibration tends to move the final forecast probabilities away from these extremes. Because of this, the calibration rarely changes the areal coverage of HREFCT probabilities or zeros out probabilities in the uncalibrated guidance.

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 subtracting that reliability error from the original forecast probability. For example, if at a given grid point the pre-calibrated guidance had a probability of 44% and the mean reliability error for the 40% bin at that grid point at that forecast hour was +10% (an overforecast by 10%), then the final, calibrated probability for that grid point would be 44% − 10% = 34%. Note that the calibration is only applied to points where probabilities already exist in the grid. As such, the calibration cannot introduce probabilities where there were none before the calibration step. An example 4-h and 24-h calibrated thunderstorm forecast from 17 March 2021 is shown in Fig. 2.

c. Instability and reflectivity mask

During initial testing of the HREFCT guidance, forecasters and researchers identified a bias in the algorithm that would

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

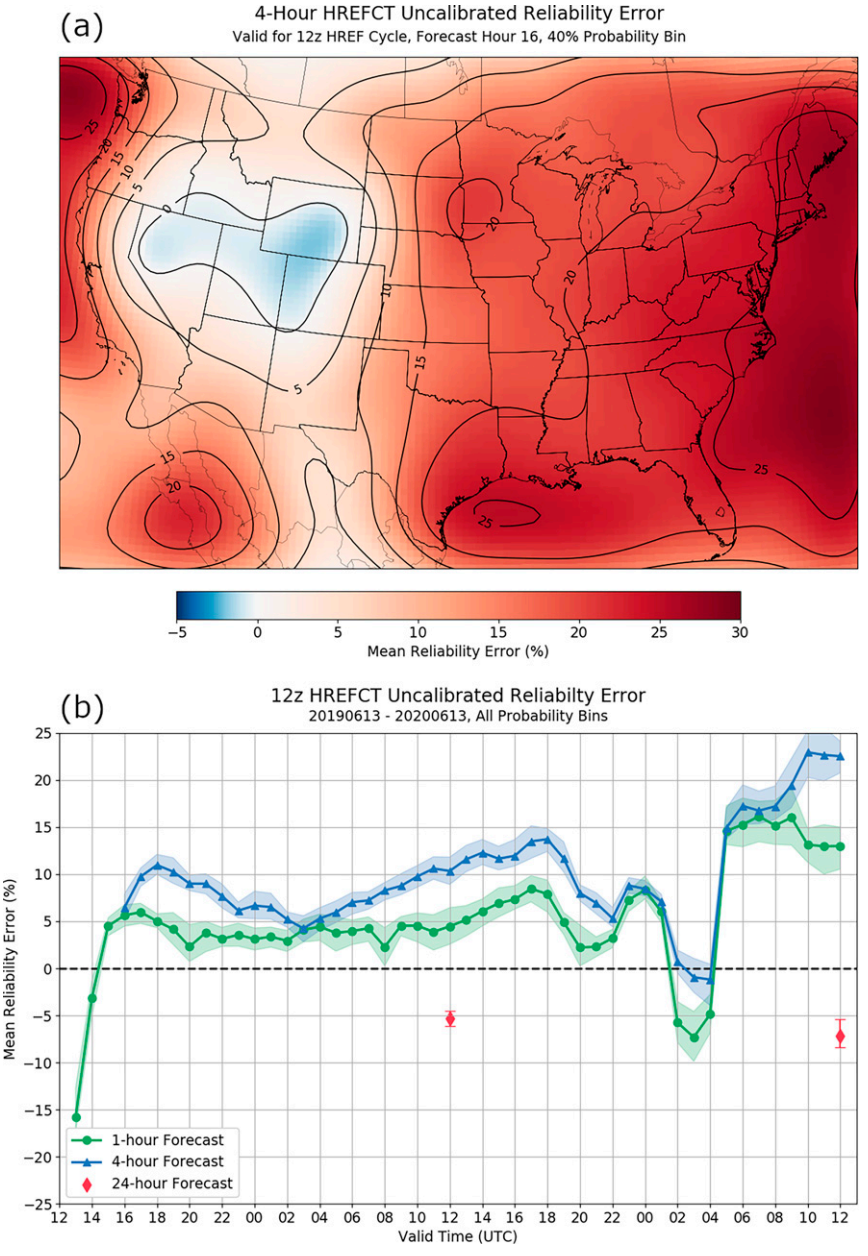

FIG. 1. (a) Mean uncalibrated reliability error of the 1200 UTC HREFCT 4-h lightning probabilities at forecast hour 16 for the 40% probability bin. Positive values represent an overforecast compared to NLDN observations from 13 Jun 2019 to 13 Jun 2020. (b) Mean uncalibrated reliability error of the 1200 UTC HREFCT as a function of lead time. The shaded regions and error bars represent 95% confidence intervals from 1000 bootstrapped samples. Forecast lead time increases to the right.

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 $\mathrm{QPF}_{\mathrm{accum}}$ term of the HREFCT equation. Although such relatively stable environments might fail to meet the thresholds for $Z_{-10^{\circ}\mathrm{C}}$ and MU LI, a sufficiently large fraction of HREF members predicting moderate accumulated precipitation values could still generate lightning probabilities.

To correct this bias, a filter was imposed on each member of the HREF to create a simple instability and reflectivity mask. The contribution from any HREF member that forecasts MU LI > 0 and $Z_{-10^{\circ}\mathrm{C}} < 35$ dBZ over the valid forecast

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

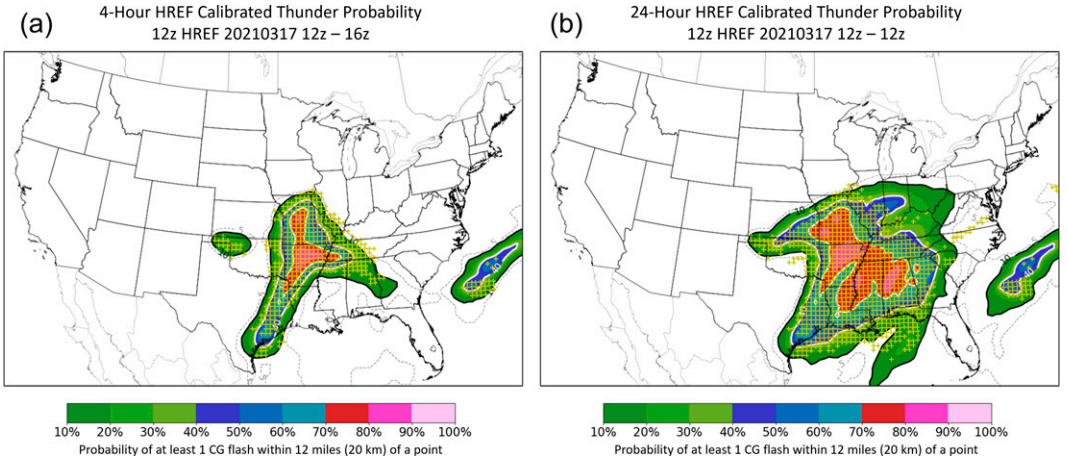

FIG. 2. HREFCT (a) 4- and (b) 24-h forecasts from the 1200 UTC HREF cycle on 17 Mar 2021. Yellow "+" symbols indicate grid points where there was at least one CG lightning flash detected during the valid forecast period.

period is set to zero when creating the probabilities for a given grid point. As an example, consider a grid point where eight of the ten HREF members predict stratiform precipitation with a maximum $Z_{-10°C}$ of 30 dBZ and a 4-h QPF_{accum} of 0.25 in. (6.35 mm). Only 1 of the 10 members predicts MU LI < 0, while the others are all >0. Without the instability mask, this grid point would be given a 25% uncalibrated 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-, 4-, and 24-h thunder forecasts were regenerated for 13 June 2019–13 June 2020 with the new mask applied, and preliminary verification revealed a slight improvement in the bulk performance of the guidance (not shown). Furthermore, anecdotal case studies and real-time application by SPC forecasters found that the mask was successful at removing most nonmeteorological regions of low thunder probabilities, particularly in the Pacific Northwest and in stratiform precipitation regions of extratropical cyclones. All discussion of the HREFCT hereafter refers to the HREFCT with the instability and reflectivity mask applied.

3. Results and discussion

Verification of the 0000 and 1200 UTC HREFCT forecast products was performed on the 11-month independent dataset of 13 June 2020–11 May 2021. Calibrated 1-, 4-, and 24-h thunder forecasts were generated for the full verification period, and the probabilities from each forecast were stratified into 10% bins as during calibration. The forecasts were then compared to the observed NLDN CG lightning flashes for each forecast hour, and the POD, FAR, CSI, and statistical reliability were computed for each probability bin. Additionally, 95% confidence intervals for the metrics were computed from 10 000 bootstrapped samples. The following discussion will focus on the 1200 UTC HREFCT guidance, but similar results were noted for the 0000 UTC guidance as well.

Verification of the HREFCT generally improved as the valid forecast window increased (Fig. 3a). The 24-h forecast product was found to have the greatest performance over the verification period with a maximum CSI of 0.43 (0.41–0.45) at the 40% probability bin. The 4-h forecast product exhibited the next best performance with a maximum CSI of 0.28 (0.26–0.30) at 30%, and the 1-h HREFCT guidance had the lowest average performance with a maximum CSI of 0.19 (0.17–0.20) at 20%. All three forecast products were found to be statistically reliable over the verification period, but the 24-h forecast tended to underforecast the observed CG lightning flashes by about 5%–10% at forecast probabilities $\geq$ 40% (Fig. 3b). In contrast, the 1- and 4-h forecasts were, on average, reliable within 5% of observations at all probability levels. Note that the 1-h HREFCT guidance rarely predicted probabilities $\geq$ 75% during the verification period (Fig. 3c), and so the higher bins were excluded when calculating the performance and reliability of the product.

The observed tendency of the 1-h guidance to produce lower probabilities than the 4- or 24-h guidance is largely a result of forecast uncertainty manifested by the spread of the HREF members. As described in section 2, the HREFCT algorithm is a linear combination of probability functions. These probability functions describe not only how favorable the predicted environment is for lightning, but also the ensemble uncertainty that those conditions will be met. Generating high probabilities in a 1-h forecast requires an equally large number of HREF members to meet the forecast thresholds at the same grid point and forecast hour (i.e., large ensemble agreement.) Conversely, the 4-h guidance only requires the HREF membership to meet those forecast thresholds at any time during the valid 4-h interval. As such, it is generally easier to achieve higher probabilities in the 4-h (and 24-h) guidance than in the 1-h guidance. This holds true at shorter lead times too, as many HREF members utilize different dynamical cores, initial and boundary conditions, microphysics schemes,

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

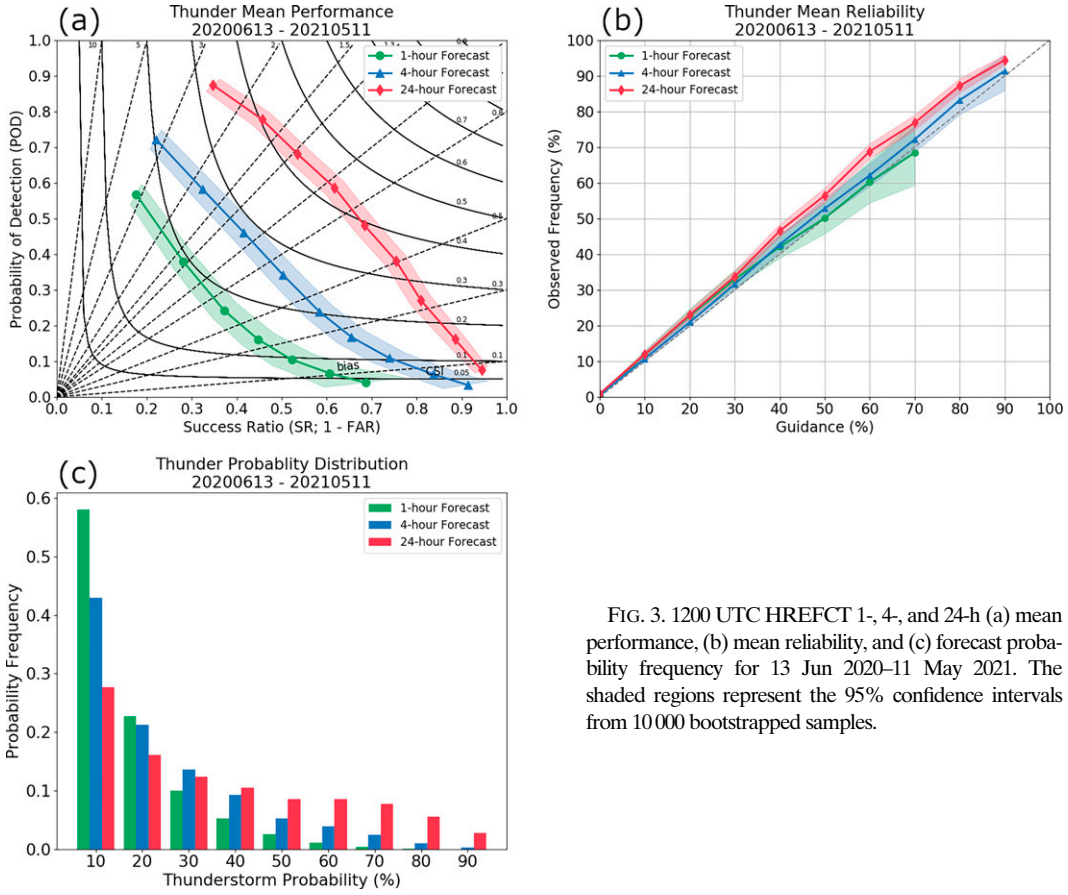

FIG. 3. 1200 UTC HREFCT 1-, 4-, and 24-h (a) mean performance, (b) mean reliability, and (c) forecast probability frequency for 13 Jun 2020–11 May 2021. The shaded regions represent the 95% confidence intervals from 10 000 bootstrapped samples.

and PBL schemes that lead to differing solutions early in the forecast cycle.

Additional nuance in the HREFCT's performance was revealed by analyzing a spatial and temporal breakdown of the 4-h guidance. The average reliability error of the product over the verification dataset was produced at each point within the NCEP 212 grid and across six 4-h intervals as shown in Fig. 4. Despite calibration, the HREFCT continued to exhibit a tendency to overforecast CG lightning probabilities on average across most of the CONUS and at most hours of the day. Even so, this error was typically within 5%–10% of observations and much improved from the 20% to 25% error noted in the uncalibrated guidance (Fig. 1). These results serve as an example of how an underdispersive or overconfident ensemble may be corrected by applying calibration to reduce probabilities at most locations and times (Raftery et al. 2005; Berrocal et al. 2007; Kann et al. 2009). More notable overforecasting was observed along and east of the Appalachians between 0400 and 1600 UTC, with reliability errors of 10%–15% common across that region. Other forecast biases include a broad area of +10% reliability error along the Gulf Coast from 0000 to 0400 UTC, a small area of +10% error in the central Plains from approximately 0800 to 2000 UTC, and a slight underforecast of up to 5% across the Southwest from 1600 to 0400 UTC.

These regional biases may be at least partially attributable to systematic error in the underlying HREF forecast. The overforecast region in the central Plains, for example, anecdotally correlates to the approximate time and location of a number of MCS events that occurred during the 2020 warm season. Although the ability of CAMs to predict MCS events has improved over recent years, some HREF members such as the HRRR have been shown to commonly overforecast MCS convection in the Plains during the overnight hours (Clark et al. 2007; Pinto et al. 2015). As such, the HREFCT may have overforecast the probability of CG lightning because some members of the HREF consistently predicted too many MCS events in that region. One important caveat to this analysis is that the relatively short 11-month verification period potentially makes the reliability error at any given location sensitive to a small number of events. For instance, there is a consistent area along the Pacific coast where the guidance overforecast the lightning potential by up to 25% on average. However, the sample size of forecasts in that region is extremely limited, and so these results may not be fully representative of the longer-term performance of the HREFCT in that area. These sample-size limitations are also applicable to the underlying HREFCT calibration, which was necessarily performed on just one year of data. Recalibration and verification on multiple years of forecasts is planned for future

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

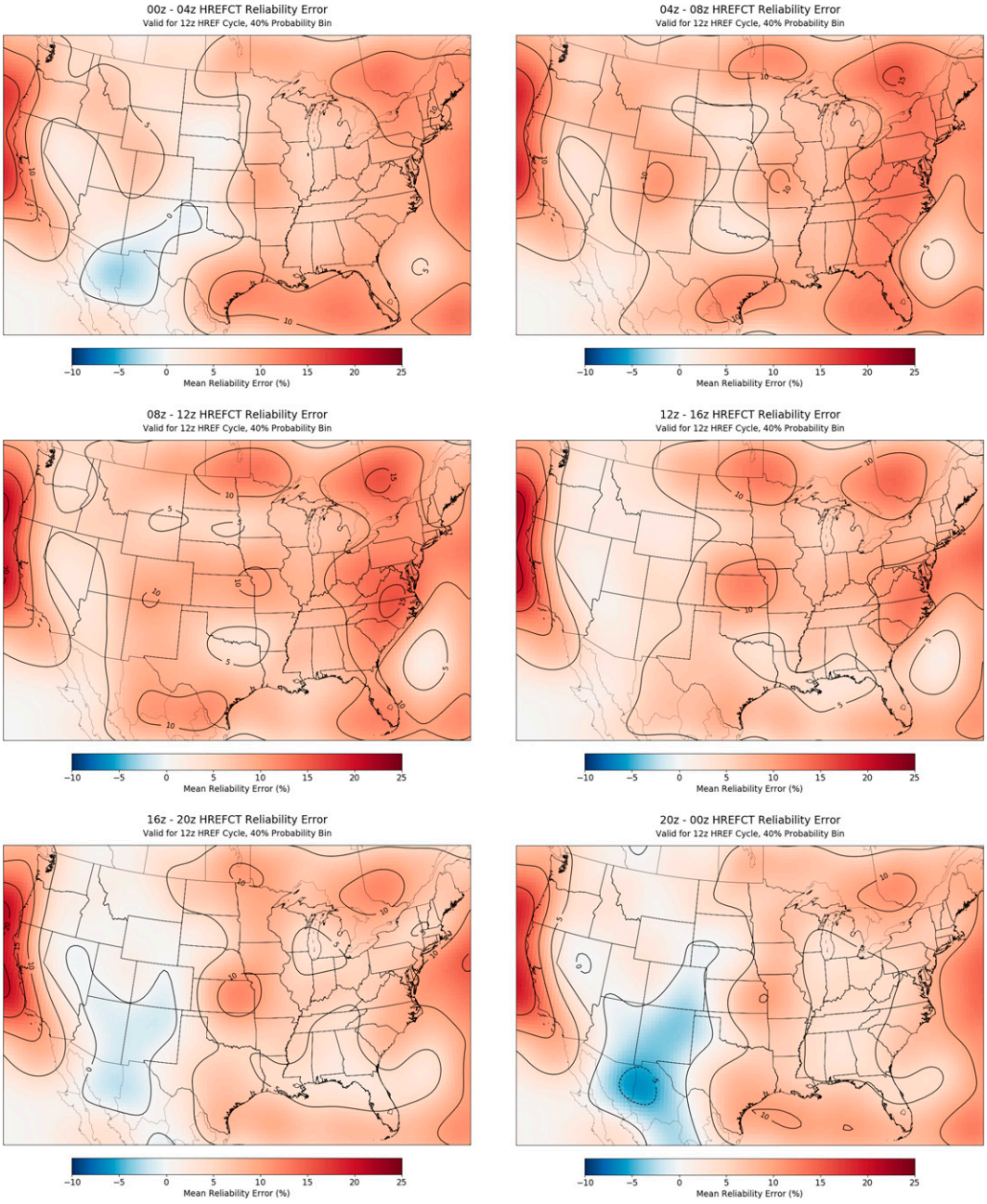

FIG. 4. 1200 UTC HREFCT mean spatial reliability error across six 4-h periods. Positive values (warmer colors) represent an overforecast and negative values (cooler colors) represent an underforecast. The reliability error was calculated for the 40% probability bin from 13 Jun 2020 to 11 May 2021.

updates to the operational guidance as a longer period of record becomes available.

A primary goal when developing the HREFCT was to improve upon the existing SREFCT guidance. As such, it was necessary to evaluate how the new HREFCT performance compared to that of the original SREFCT. The 1-, 4-, and 24-h forecasts from the SREFCT were regenerated for the 13 June 2020–11 May 2021 verification period, and the POD, FAR, CSI, and reliability were computed as before (Fig. 5). There is no 1200 UTC SREF cycle to directly compare against the 1200 UTC HREF, so the 0900 and 1500 UTC SREF cycles were used instead. The 0900 UTC 4-h SREFCT exhibited a maximum CSI of 0.20 (0.19–0.21) at the 20% probability bin, while the 1500 UTC SREFCT had a maximum CSI of about 0.21 (0.19–0.21) also at 20%. This performance was notably less than the 4-h HREFCT's maximum CSI of 0.28 (0.26–0.30) at the 30% bin. Unfortunately, several months of missing data in the local SREFCT archive prevented a direct comparison of

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

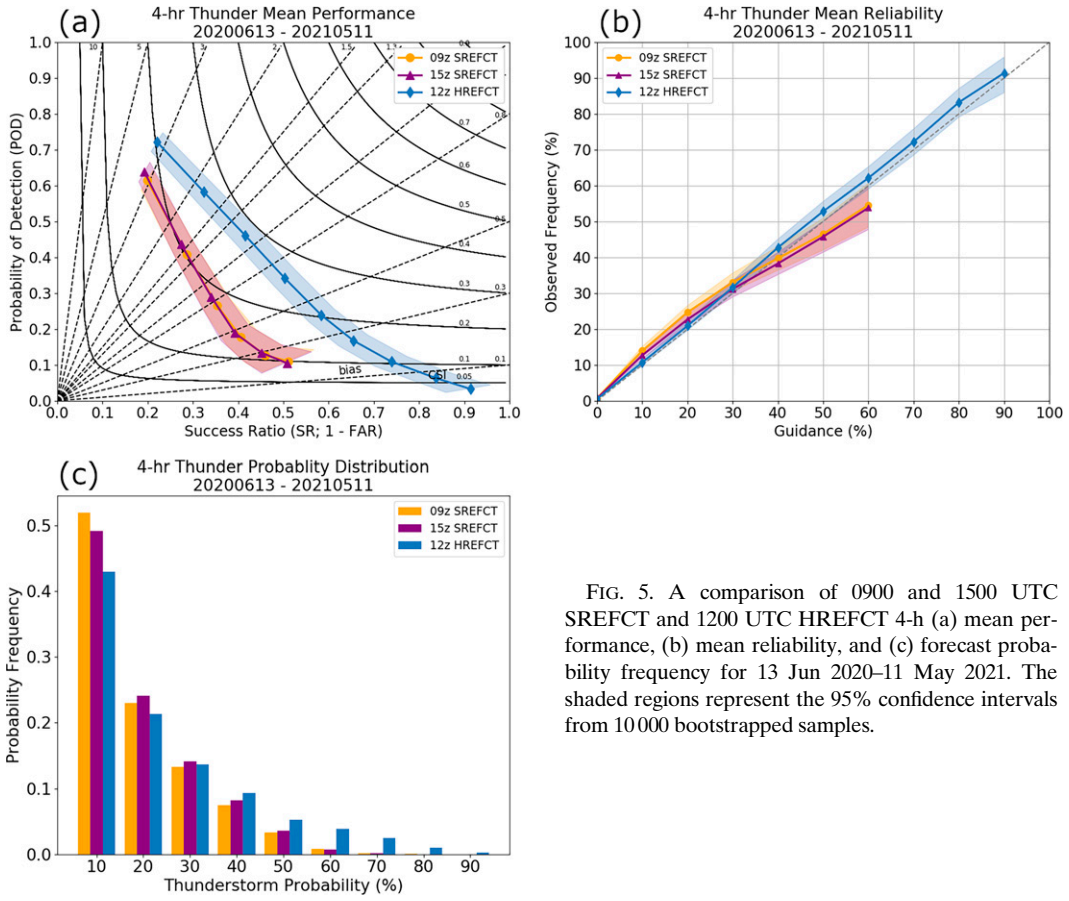

FIG. 5. A comparison of 0900 and 1500 UTC SREFCT and 1200 UTC HREFCT 4-h (a) mean performance, (b) mean reliability, and (c) forecast probability frequency for 13 Jun 2020–11 May 2021. The shaded regions represent the 95% confidence intervals from 10 000 bootstrapped samples.

the 1- and 24-h products. However, indirect comparisons using the incomplete dataset reveal a similar improvement in the HREFCT guidance over that of the SREFCT (not shown). Despite the notable improvements in CSI, the SREFCT and HREFCT products exhibited similar statistical reliability (Fig. 5b). The 4-h forecasts for the 0900 and 1500 UTC SREFCT tended to slightly underforecast probabilities $< 25\%$ while slightly overforecasting probabilities $\geq 35\%$. In contrast, the HREFCT tended to slightly underforecast CG lightning potential at all probabilities. Additionally, the forecast probability distribution of the HREFCT was notably shifted toward higher probabilities compared to that of the SREFCT (Fig. 5c). Approximately 10% of HREFCT 4-h forecast values exceeded 55% during the verification period compared to only 2% of the 0900 and 1500 UTC SREFCT forecasts. In fact, the SREFCT 4-h forecast rarely exceeded 65% during the verification period. This demonstrated ability of the HREFCT to produce higher, statistically reliable forecast probabilities is a noteworthy improvement over the SREFCT.

Next, the mean CSI and reliability error of the SREFCT and HREFCT 4-h forecasts were computed at each forecast hour to reveal diurnal and lead time variations in the guidance (Fig. 6). Both the SREFCT and HREFCT 4-h forecasts achieved their greatest mean CSI at approximately 2300 UTC on day 1 (1200 UTC HREF f11; 0900 UTC SREF f14; 1500 UTC SREF f08). As before, the HREFCT showed marked improvement over the SREFCT with a mean CSI of about 0.31 (0.30–0.32) compared to 0.24 (0.22–0.25) for the 0900 and 1500 UTC SREFCT forecasts. After a steady decline in mean performance between approximately 0100 and 1800 UTC, all versions of the guidance then experienced a secondary peak at about 2300–0000 UTC on day 2 (1200 UTC HREF f36; 0900 UTC SREF f38; 1500 UTC SREF f32). This time, the HREFCT achieved a mean CSI of about 0.26 (0.25–0.27) while the 0900 and 1500 UTC SREFCT forecasts had mean CSIs of about 0.21 (0.20–0.22) and 0.22 (0.21–0.22), respectively. Notably, the day 2 performance peak exhibited by the HREFCT is equal to or greater than the day 1 peak of the SREFCT. The HREFCT demonstrated a minimum mean performance of about 0.18 (0.17–0.19) at 1000 UTC on day 2 (f46).

More notable differences are evident when comparing the diurnal and lead time variations of the SREFCT and HREFCT mean reliability error (Fig. 6b). After an initial spinup period in the first few forecast hours, the HREFCT maintained a reliability error within 5% of observations through 0300 UTC of day 2 (f39). Beyond that time, the guidance began to overforecast CG lightning probabilities by up to 10%–15%. This reduction in reliability is likely due in part

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

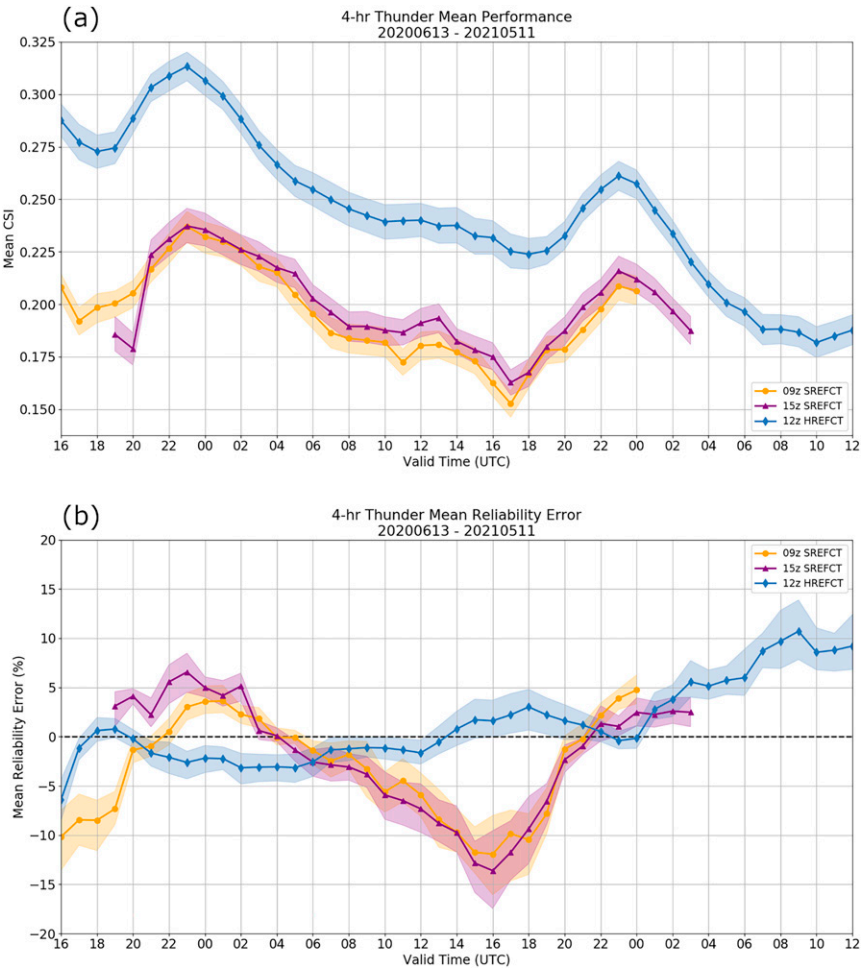

FIG. 6. A comparison of 0900 and 1500 UTC SREFCT and 1200 UTC HREFCT 4-h (a) mean performance and (b) mean reliability error as a function of lead time between 13 Jun 2020 and 11 May 2021. The shaded regions represent the 95% confidence intervals from 1000 bootstrapped samples. Forecast lead time increases to the right for both plots.

to predictability error in the spatial placement of convection at longer lead times, as well as reduced spread from the ensemble as only five members contribute to the probabilities after forecast hour 36. In contrast, the SREFCT generally remained within 5% of observations up to about 1000 UTC (0900 UTC SREF f25; 1500 UTC SREF f19) before underforecasting by up to 15% at 1600 UTC (0900 UTC SREF f31; 1500 UTC SREF f25). This strong diurnal signal in the SREFCT reliability error aligns with forecaster observations and past verification studies as discussed in the introduction. Of note, the HREFCT forecast reliability error did not exhibit a strong diurnal signal and was found to be statistically reliable on average through at least forecast hour 39.

Finally, the HREFCT and SREFCT forecast products were reviewed and compared for several case studies and in real-time SPC forecast operations. Anecdotally, the HREFCT generally produced spatially larger areas of thunder probabilities than the SREFCT, and these probabilities were frequently greater in magnitude. One example of this can be seen in Fig. 7, which shows a comparison of the HREFCT and SREFCT 4-h forecast for 1200–1600 UTC 12 April 2020. Both the HREFCT and SREFCT 4-h guidance correctly predicted the potential for lightning across the southern and central plains and the lower Mississippi Valley. However, the SREFCT probabilities peaked with a small area of 40% in northeast Texas, while the HREFCT painted a broad area of 70%–80% across parts of Arkansas and Louisiana. The SPC forecaster-created Thunderstorm Outlook for this time period included two regions of 70% across Arkansas, Louisiana, and Texas which more closely aligned with the HREFCT forecast. The HREFCT, SREFCT, and SPC Thunderstorm Outlook all failed to capture the observed lightning over parts of Georgia and Florida, although the HREFCT did have a few pockets of 5% probability in the vicinity. Scattered convection developed across much of Georgia and north Florida around 1500 UTC as a warm front lifted north across the region. Several HREF members accurately depicted the placement and timing of this convection, but the forecast 4-h $Z_{-10^{\circ}C}$ values were generally

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

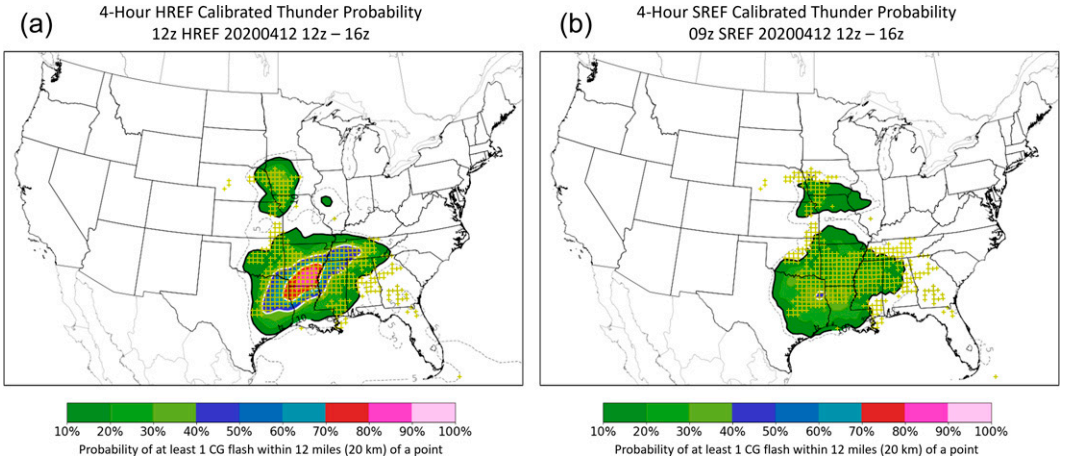

FIG. 7. (a) 1200 UTC HREFCT and (b) 0900 UTC SREFCT 4-h calibrated thunder forecasts for 1200–1600 UTC 12 Apr 2020. Yellow "+" symbols indicate grid points where there was at least one CG lightning flash detected during the valid forecast period.

<40 dBZ and 4-h QPF_{accum} values were <2 mm. Therefore, the reflectivity and precipitation thresholds of the HREFCT algorithm were not met and the resulting probabilities were <5% after calibration. This case highlights a potential limitation of the HREFCT to predict lightning in "dry" thunderstorm scenarios. In particular, the partial reliance of the HREFCT algorithm on QPF_{accum} may greatly limit probabilities when convection is forecast to produce <1 mm of precipitation in an hour or <2 mm over a 4-h period. Additional HREFCT products are currently in development to better account for dry thunderstorm scenarios which play an important role in fire weather forecasting.

The HREFCT guidance was implemented operationally on NCEP's Weather and Climate Operational Supercomputing System (WCOSS) on 11 May 2021 and is now being distributed by the NWS. The HREFv3 was also implemented operationally on the same day, replacing the NMMB member with an FV3 core and extending the HRRR to 48 forecast hours. Unfortunately, the timing of the HREFv3 implementation and limited retrospective and reforecast availability precluded a more in-depth analysis of the impacts of these changes to the HREFCT performance. That said, anecdotal observations and initial verification have not revealed any appreciable change in the guidance. At SPC, the HREFCT has now largely replaced the SREFCT in forecast operations as a "first guess" when generating Thunderstorm Outlooks and other thunderstorm forecast products.

4. Conclusions

A new suite of calibrated thunderstorm forecast products has been developed using a combination of HREF derived radar reflectivity, accumulated precipitation, and MU LI forecasts. These products have been shown to skillfully and reliably predict the probability of at least one CG lightning flash over a given 1-, 4-, or 24-h forecast period. Additionally, the HREFCT guidance has exhibited increased CSI and similar or improved reliability when compared to calibrated thunder guidance generated from the non-convection-allowing SREF. Anecdotally, the HREFCT generally produces spatially larger areas of thunder probabilities than the SREFCT, and these probabilities are frequently greater in magnitude. The new HREFCT guidance is also more consistently reliable with increasing lead time and does not exhibit strong diurnal variation in reliability error. This is in direct contrast to the SREFCT, which was shown to strongly underforecast thunder probabilities on average during the overnight and morning hours. Finally, the HREFCT was found to improve upon the SREFCT's performance at longer lead times, with the new guidance achieving a higher mean CSI in both the day 1 and day 2 forecast periods. Notably, the HREFCT's best performance during the day 2 forecast period was determined to be equal to or greater than that of the SREFCT on day 1.

After extensive testing, the HREFCT became operational within the NWS in May 2021, and the guidance is now being distributed by the NWS. Internally, SPC forecasters have largely begun to utilize the HREFCT in combination with or in replacement of the SREFCT when generating thunderstorm forecast products. As of this writing, the 1-h HREFCT is planned to be added to the NBMv4.1 in combination with the SREFCT, LAMP, and various MOS guidance to inform a blended thunderstorm probability product. Further development of the HREFCT is ongoing, with future updates expected to include the explicit prediction of CG lightning frequency, or the probability that the number of CG flashes will exceed some threshold value over a given forecast period. All HREFCT 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.

Acknowledgments. The authors wish to thank Nick Nauslar for his insights and early work into this project, as well as all SPC forecasters for their invaluable feedback. We also thank three anonymous reviewers whose

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

constructive suggestions undoubtedly improved the quality of this manuscript. The first author was provided support by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreements NA16OAR4320115 and NA21OAR4320204, U.S. Department of Commerce. Authors MDE, ILJ, and PTM completed this work as part of regular duties at the federally funded NOAA Storm Prediction Center. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

Data availability statement. SREF, HREFv2, and HREFv2.1 model data used in this study were provided to the SPC by NCEP's Environmental Modeling Center (EMC) and are archived at the National Severe Storms Laboratory (NSSL). Proprietary NLDN lightning observations were provided by Vaisala and may require licensing to access. All HREFCT and SREFCT forecasts presented in this study are archived internally at the SPC, and are publicly available on the SPC website (https://www.spc.noaa.gov/exper/href/?model=href&product=guidance_thunder_hrefct_004h and https://www.spc.noaa.gov/exper/sref/sref.php?run=2021082009&id=SREF_PROB_TRW_CALIBRATED_24HR__, respectively). Real-time HREFCT forecasts can also be accessed at https://nomads.ncep.noaa.gov/pub/data/nccf/com/spc_post/prod/. Archived HREF forecasts are available at https://data.nssl.noaa.gov/thredds/catalog/FRDD/HREF.html.

REFERENCES

Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932, https://doi.org/10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2.

Benjamin, S. G., 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.

Bergstra, J., and Y. Bengio, 2012: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, **13**, 218–305.

Berrocal, V. J., A. E. Raftery, and T. Gneiting, 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Wea. Rev.*, **135**, 1386–1402, https://doi.org/10.1175/MWR3341.1.

Bright, D. R., and J. S. Grams, 2009: Short Range Ensemble Forecast (SREF) calibrated thunder probability forecasts: 2007–2008 verification and recent enhancements. *Fourth Conf. on the Meteorological Applications of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc., 6.3, https://ams.confex.com/ams/89annual/techprogram/paper_148509.htm.

——, M. S. Wandishin, R. E. Jewell, and S. J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. *Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.3, https://ams.confex.com/ams/pdfpapers/84173.pdf.

Charba, J. P., F. G. Samplatsky, A. J. Kochenash, P. E. Shafer, J. E. Ghirardelli, and C. Huang, 2019: LAMP upgraded convection and total lightning probability and "potential" guidance for the conterminous United States. *Wea. Forecasting*, **34**, 1519–1545, https://doi.org/10.1175/WAF-D-19-0015.1.

Clark, A. J., W. A. Gallus Jr., and T. C. Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. *Mon. Wea. Rev.*, **135**, 3456–3473, https://doi.org/10.1175/MWR3467.1.

Craven, J. P., and Coauthors, 2018: Overview of National Blend of Models version 3.1. Part I: Capabilities and an outlook for future upgrades. *25th Conf. on Probability and Statistics*, Austin, TX, Amer. Meteor. Soc., 7.3, https://ams.confex.com/ams/98Annual/webprogram/Paper325347.html.

Cummins, K., M. Murphy, E. Bardo, W. Hiscox, R. Pyle, and A. Pifer, 1998: A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network. *J. Geophys. Res.*, **103**, 9035–9044, https://doi.org/10.1029/98JD00153.

Du, J., and Coauthors, 2014: NCEP regional ensemble update: Current systems and planned storm-scale ensembles. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction,* Atlanta, GA, Amer. Meteor. Soc., J1.4, https://ams.confex.com/ams/94Annual/webprogram/Paper239030.html.

EMC, 2021: HREFv3 official evaluation. Accessed 5 August 2021, https://www.emc.ncep.noaa.gov/users/meg/hrefv3/.

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.

Glahn, H. R., and D. L. Lowry, 1972: The use of model output statistics in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

Hamill, T. M., E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. National Blend of Models for statistical postprocessing of probability of precipitation and deterministic precipitation amount. *Mon. Wea. Rev.*, **145**, 3441–3463, https://doi.org/10.1175/MWR-D-16-0331.1.

Hughes, K., 2001: Development of MOS thunderstorm and severe thunderstorm forecast equations with multiple data sources. Preprints, *18th Conf. on Weather Analysis and Forecasting*, Fort Lauderdale, FL, Amer. Meteor. Soc., 191–195.

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.

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, https://doi.org/10.1175/2010WAF2222430.1.

Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the Global Forecast System at NMC. *Wea. Forecasting*, **6**, 425–435, https://doi.org/10.1175/1520-0434(1991)006<0425:RCIITG>2.0.CO;2.

Kann, A., C. Wittmann, Y. Wang, and X. Ma, 2009: Calibrating 2-m temperature of limited-area ensemble forecasts using high-resolution analysis. *Mon. Wea. Rev.*, **137**, 3373–3387, https://doi.org/10.1175/2009MWR2793.1.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC

Maloney, J., E. Engle, P. Shafer, and G. Wagner, 2009: The NMM MOS replacement for the Eta MOS. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 6A.3, https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154270.htm.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasting with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.*, **77**, 2637–2649, https://doi.org/10.1175/1520-0477(1996)077<2637:IIQPFW>2.0.CO;2.

——, T. L. Black, D. W. Plummer, and J. H. Ward, 1990: Eta model precipitation forecasts for a period including Tropical Storm Allison. *Wea. Forecasting*, **5**, 483–493, https://doi.org/10.1175/1520-0434(1990)005<0483:EMPFFA>2.0.CO;2.

NWS, 2021: Service change notice 21–38. Accessed 5 August 2021, https://www.weather.gov/media/notification/scn21-38hiresw_v8_hrefaaa.pdf.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the high-resolution Rapid Refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, https://doi.org/10.1175/WAF-D-14-00118.1.

Raftery, A., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

Reap, R. M., and D. S. Foster, 1979: Automated 12–36 hour probability forecasts of thunderstorms and severe local storms. *J. Appl. Meteor.*, **18**, 1304–1315, https://doi.org/10.1175/1520-0450(1979)018<1304:AHPFOT>2.0.CO;2.

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.

——, 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.

Rogers, E., and Coauthors, 2005: The NCEP North American Mesoscale Modeling System: Final Eta model/analysis changes and preliminary experiments using the WRF-NMM. *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction*, Washington, DC, Amer. Meteor. Soc., 4B.5, https://ams.confex.com/ams/WAFNWP34BC/techprogram/paper_94707.htm.

——, and Coauthors, 2017: Upgrades to the NCEP North American Mesoscale (NAM) system. Bluebook Rep., 2 pp., http://wmc.meteoinfo.ru/bluebook/uploads/2017/docs/05_Rogers_Eric_mesoscale_modeling.pdf.

Saunders, C. P. R., 1993: A review of thunderstorm electrification processes. *J. Appl. Meteor.*, **32**, 642–655, https://doi.org/10.1175/1520-0450(1993)032<0642:AROTEP>2.0.CO;2.

Shafer, P. E., and H. E. Fuelberg, 2008: A perfect prognosis scheme for forecasting warm-season lightning over Florida. *Mon. Wea. Rev.*, **136**, 1817–1846, https://doi.org/10.1175/2007MWR2222.1.

——, and D. E. Rudack, 2015: Development of a MOS thunderstorm system for the ECMWF model. *Seventh Conf. on the Meteorological Applications of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc., 2.1, https://ams.confex.com/ams/95Annual/webprogram/Paper261494.html.

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.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

SPC, 2021a: About the SPC. Accessed 2 August 2021, https://www.spc.noaa.gov/misc/aboutus.html.

——, 2021b: SPC products. Accessed 2 August 2021, https://www.spc.noaa.gov/misc/about.html#Thunderstorm%20Outlooks.

Stough, S., E. M. Leitman, J. L. Peters, and J. Correia Jr., 2012: On the role of Storm Prediction Center products in decision making leading up to hazardous weather events. *11th Annual AMS Student Conf. and Career Fair*, New Orleans, LA, Amer. Meteor. Soc., https://ams.confex.com/ams/92Annual/webprogram/Paper204254.html.

Tew, M. A., A. Stern, A. Horwitz, K. Gilbert, and D. T. Myrick, 2016: National Blend of Models: Transformational forecast change by the National Weather Service. *Fourth Symp. on Building a Weather-Ready Nation*, New Orleans, LA, Amer. Meteor. Soc., 6.2, https://ams.confex.com/ams/96Annual/webprogram/Paper286228.html.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/15/22 05:24 AM UTC