AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079750

Special Section:
Bridging Weather and Climate: Subseasonal-to-Seasonal (S2S) Prediction

Key Points:
- Dynamical predictions of a simple lightning proxy give skillful forecasts of U.S. cloud-to-ground lightning
- Forecasts have statistically significant skill for lead times up to 15 days
- January weeks 3-4 forecasts of CG flash numbers and extent have statistically significant skill

Supporting Information:
- Supporting Information S1

Correspondence to:
M. K. Tippett,
mkt14@columbia.edu

Citation:
Tippett, M. K., & Koshak, W. J. (2018). A baseline for the predictability of U.S. cloud-to-ground lightning. *Geophysical Research Letters*, *45*, 10,719–10,728. https://doi.org/10.1029/2018GL079750

Received 24 JUL 2018
Accepted 18 SEP 2018
Accepted article online 21 SEP 2018
Published online 10 OCT 2018
Corrected 6 DEC 2018

This article was corrected on 6 DEC 2018. See the end of the full text for details.

A Baseline for the Predictability of U.S. Cloud-to-Ground Lightning

Michael K. Tippett[1] **and William J. Koshak**[2]

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, [2]Earth Science Branch, NASA Marshall Space Flight Center, Huntsville, AL, USA

Abstract The product of convective available potential energy and precipitation rate is used as a proxy for U.S. cloud-to-ground (CG) lightning. The proxy is computed from numerical weather prediction model outputs and thus provides a simple method for producing CG lightning threat forecasts. Here three types of forecasts are considered: number of CG flashes, spatial extent of CG lightning occurrence, and lightning/no lightning maps at 1° × 1° resolution. Forecasts have statistically significant skill for lead times up to 15 days depending on the type of forecast. Verification of the forecasts over the United States provides a predictability baseline against which other approaches can be compared.

Plain Language Summary Many studies of weather predictability have examined variables related to atmospheric circulation. Here we examine the predictability of cloud-to-ground (CG) lightning, which is the defining feature of thunderstorms. Forecasts for numbers of CG flashes and their spatial extent over the United States are constructed from the output of a state-of-the-art numerical weather prediction model and compared with observed CG lightning. The skill of the lightning forecasts is statistically significant for lead times up to 15 days, depending on the type of forecast, time of year, and region. The proxy-based forecasts provide a baseline for the predictability of thunderstorms against which other approaches can be compared.

1. Introduction

The question of how far in advance that weather can be predicted has been addressed using conceptual models for error growth as well as the observed skill of operational forecast systems (Lorenz, 1982; Simmons & Hollingsworth, 2002). With forecast system improvements, the limit of weather predictability has been shown to extend beyond 2 weeks although the precise limits depend on the physical quantity being predicted, its temporal and spatial resolution, and geographic location (Buizza & Leutbecher, 2015). Weather forecast predictability studies have often examined the predictability of quantities closely related to atmospheric circulation such as sea-level pressure or 500-hPa height. There have been relatively few studies of the predictability of other quantities such as precipitation or convection (Stern & Davidson, 2015; Zhu et al., 2013). The skill of monthly and seasonal forecasts of convective available potential energy (CAPE) has been examined in the context of severe thunderstorms (Jung & Kirtman, 2016; Lepore et al., 2018). The goal of this work is to establish a baseline for the predictability of lightning.

Lightning is the defining feature of thunderstorms, a weather phenomenon that occurs around the world. The most severe thunderstorms are associated with tornadoes, hail, and damaging winds, and they are responsible for property damage and loss of life. The frequency of severe thunderstorms in the United States is modulated on seasonal and subseasonal time scales by predictable signals that include the Madden-Julian Oscillation (MJO; Thompson & Roundy, 2013) and the El Niño-Southern Oscillation (ENSO; Allen et al., 2015). U.S. lightning activity has also been related to the MJO (Abatzoglou & Brown, 2009). However, most operational lightning-threat forecast products have lead times of at most a few days, and how the predictability of lightning compares to that of other quantities is unclear.

Lightning discharges can be classified as cloud-to-ground (CG) and intracloud. Lightning (intracloud, CG, or their sum) is represented in numerical weather prediction models with varying levels of detail. Current representations include explicit lightning physics (Fierro et al., 2013), dynamic evolution of lightning potential as measured by microphysical and dynamical processes (Lynn, 2017; Lynn et al., 2012), and diagnostics that

are computed from ice-phase hydrometeor fields in convection-allowing models (McCaul et al., 2009) and that are computed from other convective parameterization outputs in models with parameterized convection (Lopez, 2016). For the most part, the implementations of such formulations have been limited to forecast systems with fairly short lead times since nowcasts of lightning threat, especially ones with high spatial resolution, can be useful with lead times of an hour or less. For instance, the McCaul et al. (2009) diagnostic is used in the operational 3-km High-Resolution Rapid Refresh (Benjamin et al., 2016), which makes forecasts out to 36 hr, and the Lynn et al. (2012) scheme is used in the 4-km Weather Research and Forecast Model run by the National Severe Storms Laboratory (NSSL-WRF) for forecasts out to 36 hr.

Seasonal forecasts of lightning activity have been proposed (Dowdy, 2016; Muñoz et al., 2016), as well as projections of changes in lightning activity in future climates (Romps et al., 2014). These long-range forecasts and projections have either used purely statistical relations of climate modes such as ENSO with lightning, or statistical relations between lightning and predicted quantities such as CAPE. Romps et al. (2014) proposed the product of CAPE and precipitation as a proxy for the number of CG flashes and, based on its good performance during 2011, used it to interpret climate change projections. Tippett et al. (2018) found that the performance of a similar proxy over the United States was fairly good but varied by region, time of the year, and temporal averaging window. Proxy performance was better on shorter time scales and during the cool season, with some spatially varying biases. A similar proxy also performed well over Bangladesh (Dewan et al., 2018). While previous studies have computed this proxy using observations or reanalysis, here we compute the proxy from numerical weather prediction model outputs in the context of subseasonal forecasts whose lead times extend beyond those of short-range forecasts. The proxy-based forecasts are then verified against CG lightning observations to obtain an estimate for the predictability of lightning.

2. Data and Methods

2.1. Data

CAPE and precipitation forecasts are produced by the NCEP Global Ensemble Forecast System (GEFS). The current GEFS version (v11) uses semi-Lagrangian dynamics with 64 sigma-pressure hybrid vertical levels (Zhou et al., 2017). Horizontal resolution is T574 (34 km) for the first eight forecast days and then decreases to T382 (70 km) for forecast days 9–35. Forecast day 1 of an integration with start time 0000 1 Jan is 1 January, and so on, to forecast day 35, which is 4 February. The GEFS retrospective forecasts used here were assembled in support of NOAA's Subseasonal Experiment (SubX), a Climate Testbed project focused on subseasonal predictability and predictions. Forecasts consist of integrations that start at 0000 UTC on Wednesday of each week 1999–2016. Here we use starts from 0000 UTC 1 Jan 2003 to 0000 UTC 28 Dec 2016, for a total of 731 forecasts. This period matches the availability of the lightning data used here (see below). Daily (UTC) averages of forecast CAPE and accumulated precipitation are provided for the SubX project at $1^\circ \times 1^\circ$ horizontal resolution. Each start has 11 ensemble members (1 control and 10 perturbed members generated using an ensemble Kalman filter). There are 21 members in the real-time SubX forecasts (not examined here), which began on 30 June 2017. Only ensemble means are considered here.

CG lightning flash data come from the National Lightning Detection Network (NLDN; Cummins & Murphy, 2009) and are the same that were used in Tippett et al. (2018). CG counts are summed on a $1^\circ \times 1^\circ$ grid at daily (UTC) resolution to match the GEFS SubX model data. We only use CG flash counts over contiguous U.S. (CONUS) land points. The NLDN data here cover the period 2003–2016 (5,114 days) during which the NLDN network is complete and fairly stable (Koshak et al., 2015). There are 9 days with missing NLDN data: 31 December 2003, 8 February 2004, 08 February 2006, 16 January 2007, 06 February 2007, 18 December 2007, 05 December 2008, 17 January 2009, and 16 January 2013, and those days are excluded from calculations. Forecasts of weekly and 2-week target periods are calculated using the forecast data that matches the available NLDN data. An increase noted in the number of positive CG flashes reported during 2016 may be due to a change in the NLDN Total Lightning Processor on 18 September 2015 (Nag et al., 2016). Negative and positive ($\geq$15 kiloampere; kA) polarity CG lightning flash counts are analyzed together. The threshold of 15 kA for positive polarity flashes accounts for the tendency of the NLDN to misclassify cloud pulses as low-amplitude, positive CG strokes (Biagi et al., 2007; Cummins & Murphy, 2009). Although the new NLDN Total Lightning Processor removes the 15-kA peak current limit for positive CG flashes, we apply a 15-kA filter to the positive CG flashes in 2015 and 2016 for consistency.

2.2. Methods

The forecast lightning proxy (CP) at each 1° × 1° grid point is defined to be proportional to the product of collocated daily values of CAPE and precipitation (Tippett et al., 2018). This definition differs from that of Romps et al. (2014) who formed the product of CONUS-averaged CAPE and precipitation, which precludes resolution of regional features. The constant of proportionality varies with lead time and is chosen so that the area-weighted CONUS average of forecast CP for each lead time matches the CONUS number of NLDN CG flashes over the period 2003–2016. The constant of proportionality varies with lead time but is the same for all forecasts and has no spatial or seasonal dependence. Forecasts of the daily number of CG flashes are simply the daily CONUS sums of forecast CP.

Forecasts of the daily number of CONUS 1°×1° grids with one or more CG flashes (number of grids with CG > 1; NGCG > 1) are constructed by counting the number of grids where forecast CP exceeds a specified threshold. The threshold varies with lead time and is chosen so that the forecasts of the daily number of CONUS 1° × 1° grids with one or more CG flashes are unbiased for each lead time over the entire forecast period. Like the constant of proportionality, the CP threshold used for predicting NGCG > 1 depends on the lead but has no spatial or seasonal dependence. Regional thresholds are used for forecasts of the regional number of grids with CP greater than one because regional biases in CP vary substantially (Tippett et al., 2018). No regionally varying constant of proportionality is used for the forecasts of the number of regionally aggregated CG flash number since forecast quality is measured using correlation, which is insensitive to scaling.

Skill of forecasts of numbers of CG flashes and NGCG > 1 are assessed using rank (Spearman) correlation, which is insensitive to outliers. Outliers are a concern since CG flash numbers vary over a large range. Except where noted, daily anomalies of forecasts and observations are used. For each lead time, a forecast climatology FC(T) is computed using two annual harmonics and has the form

$$\mathrm{FC}(T) = \exp\left(a_0 + a_1 \cos \nu T + b_1 \sin \nu T + a_2 \cos 2\nu T + b_2 \sin 2\nu T\right) , \tag{1}$$

where T is the target time in days, and $\nu = 2\pi/365.25$ days. The resulting climatology is annually periodic. Use of the exponential in the formulation guarantees that the climatology is positive. The five coefficients a_0, a_1, b_1, a_2, and b_2 are found by Poisson quasi-maximum likelihood estimation from the 731 forecasts. The observational climatology that are used to produce observed anomalies are computed in the same way and are fit to the data from the days that are forecast.

We assess regional skill at the level of NOAA climate regions (see supporting information Figure S1 and Karl & Koss, 1984) for forecasts of daily numbers of CG flashes and for forecasts of daily NGCG>1. Statistical association of GEFS forecasts of regional daily numbers of CG and daily NGCG > 1 are assessed using Pearson correlation because removal of the annual cycle can produce spurious rank correlations (not shown) for regions where lightning is rare.

Standard significance tests are used to determine the statistical significance at the 0.05 level of rank and Pearson correlations for daily and weekly forecasts. Standard significance tests are not suitable for weeks 3–4 forecasts because these forecasts are not independent. Forecast starts are weekly, and each weeks 3–4 forecast overlaps by 1 week with the previous and subsequent forecasts. This overlap has the effect of reducing the effective number of degrees of freedom and increases the correlation level required for statistical significance. Statistical significance of weeks 3–4 forecasts at the 0.05 level is determined by comparing forecast correlations to the 95th percentile of the correlation of no-skill forecasts. The no-skill forecasts are forecasts that are verified against 1,000 realizations of randomly permuted weeks 3–4 observations. The skill of weeks 3–4 forecasts is evaluated separately by start date month, and the randomly permuted observations come from the same month.

The skill of 1°×1° CG lightning/no CG lightning forecasts is measured using metrics based on 2×2 contingency tables. Here we use the standard forecast quality measures: Threat Score (TS), Probability of Detection (POD), False Alarm Rate (FAR), and the Heidke Skill Score (HSS; Wilks, 2011). To assess their significance, we compare them to the 95% intervals of the scores of no-skill forecasts. The no-skill forecasts are forecasts that are verified against 1,000 randomly permuted observations. Permutations are constructed so that only the years are permuted and not the day of the year. This sampling strategy accounts for the seasonality of lightning activity.

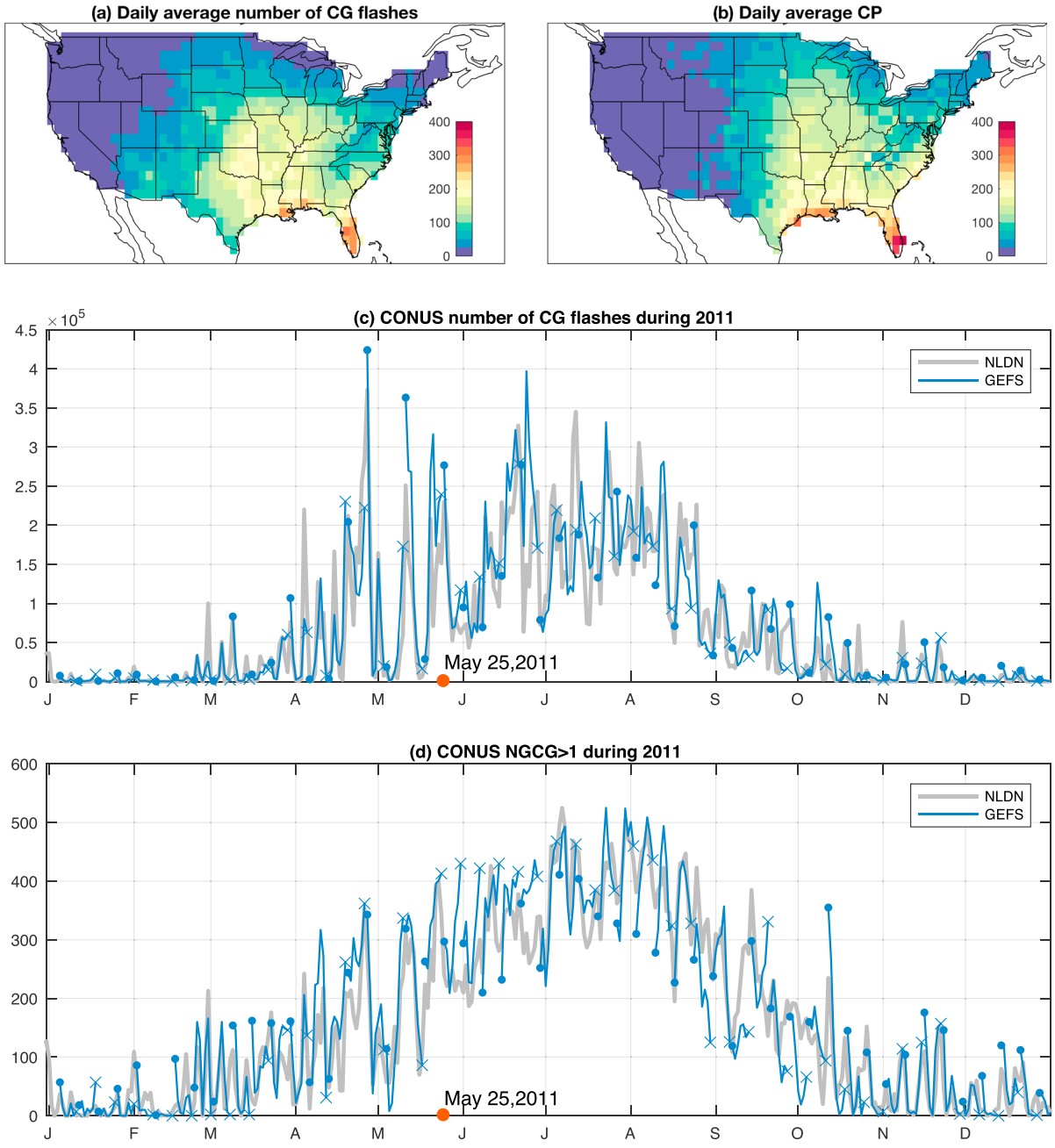

Figure 1. Average daily number of (a) CG flashes and (b) CP on forecast days 2003–2016. Number of (c) CG flashes and (d) grids with one or more CG flashes during 2011 (NGCG > 1; gray lines). Corresponding 7-day GEFS forecasts are shown as blue line segments with day 1 at the dot and day 7 at the x; 25 May is marked for reference. CG = cloud-to-ground; NGCG = number of grids with CG; GEFS = Global Ensemble Forecast System; NLDN = National Lightning Detection Network.

3. Results

The spatial pattern of the annual climatology of the daily CG flash rate matches fairly well that of forecast CP (all leads and starts) with an uncentered pattern correlation of 0.95, and a centered pattern correlation of 0.89 (Figures 1a and 1b). East of the Rockies, climatological values of CP tend to be too large to the north. Climatological values of CP are too low in west Texas, New Mexico, Arizona, Utah, and Colorado. The limited westward extension of CP compared to CG flash rate is also present when CP is computed using reanalysis (Tippett et al., 2018). One source of bias in CP is that it does not take into consideration differing rainfall processes and relations with lightning rate. For instances, the ratio of convective rainfall to lightning (rainfall yield) is relatively low in the arid southwestern United States where ice-based precipitation processes are dominant

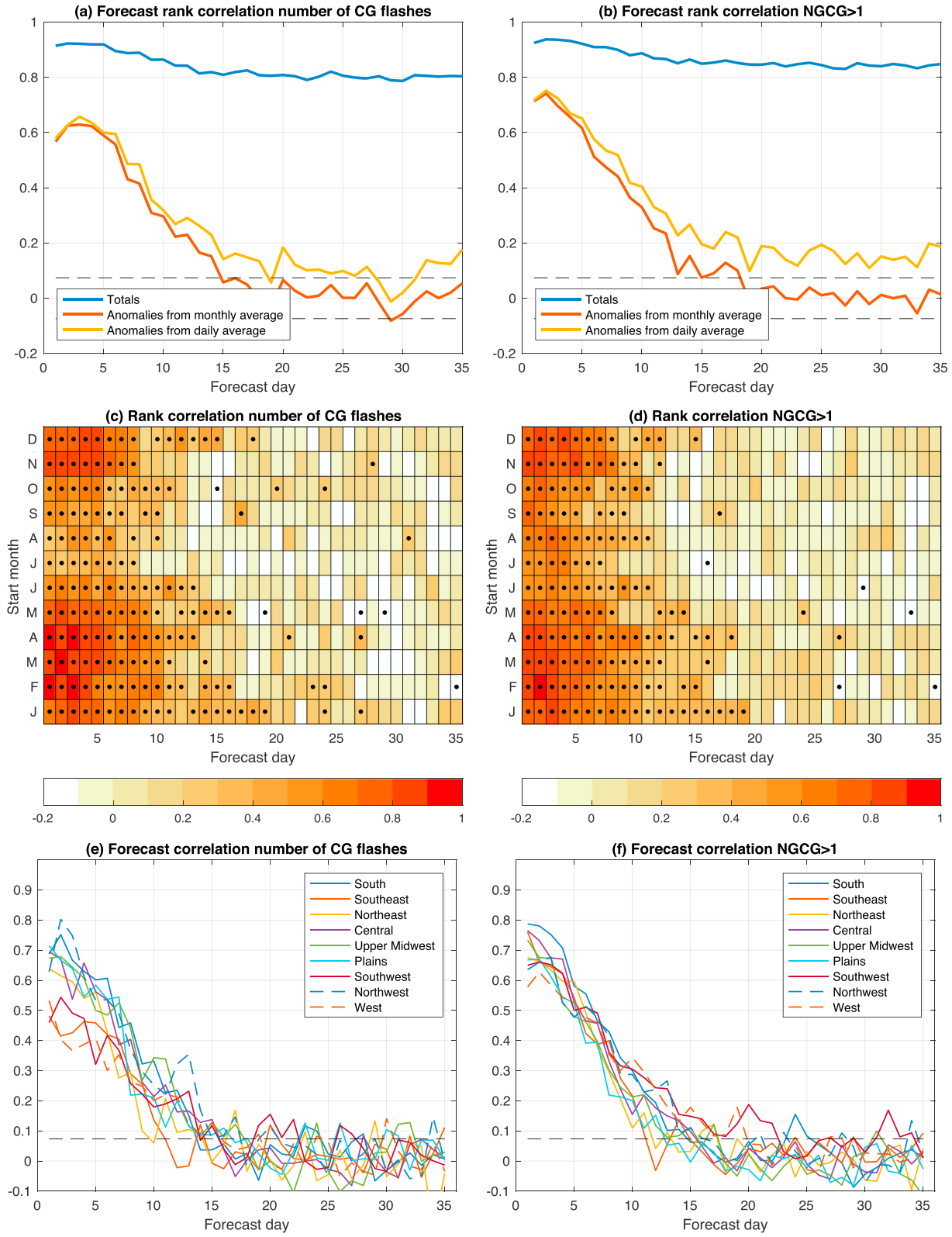

Figure 2. Rank correlation of total, anomalies with respect to a monthly climatology, and anomalies with respect to a daily climatology for all starts as a function of target day for forecasts of (a) the number of CG flashes and (b) the number of grids with one or more CG flashes (NGCG>1). Rank correlation of forecast daily anomalies as a function of start month and forecast day for (c) number of CG flashes and (d) number of grids with one or more CG flashes (NGCG > 1). Pearson correlation for forecasts of (e) the regional numbers of CG flashes and (f) the regional number of grids with one or more CG flashes (NGCG>1). CG = cloud-to-ground; NGCG = number of grids with CG.

(Petersen & Rutledge, 1998). Fuchs et al. (2015) hypothesized that storms with high cloud-base heights or shallow warm-cloud depths, which are less common east of the Rockies, have less warm-phase precipitation, more mixed-phase precipitation, and higher total lightning flash rates. Since CP uses a spatially uniform constant of proportionality, areas with lower rainfall yields have CP values that are too low compared to observations.

Visual inspection indicates that forecast values of CP track CONUS numbers of CG flashes (Figure 1c) and NGCG > 1 (Figure 1d) fairly well for forecast days 1–7 during 2011, the year that was analyzed in Romps et al. (2014), with the CP-based forecasts often capturing both the observed timing and amplitude. Romps et al. (2014) found using sounding and radar data that the product of CONUS-averaged CAPE and precipitation explained 77% of the variance (including annual and diurnal variation) in the 2011 time series of CONUS CG flashes. The number of CG flashes and NGCG > 1 show strong seasonality in 2011 with a clearly delineated active season from May through September. Other years show similar seasonality and correspondence between forecasts and CG data (not shown). Large daily variations are superimposed on the seasonal cycle in the observations and in the forecasts. The weekly frequency of forecasts starts has the consequence that forecasts sample lightning events at a limited number of discrete lead times. For instance, forecasts of 25 May 2011 are only available for forecast days 1, 8, and 15, which means that skill in predicting 25 May 2011 CG lightning cannot be assessed at intermediate leads. The fact that forecasts of a particular event are only available with leads time with intervals of 7 days presumably introduces sampling variability into skill verifications.

The forecast rank correlation of numbers of CG flashes and NGCG > 1 is greater than 0.8 at all leads when the annual cycle is not removed (blue curves in Figures 2a and 2b). This high level of correlation is in large part due to the forecasts matching well the observed annual cycle. The forecast rank correlation at the shortest leads is reduced to about 0.6 for forecasts of the number of CONUS flashes and to about 0.7 for forecasts of NGCG > 1 when the annual cycle is removed at either monthly or daily resolution (yellow and red curves in Figures 2a and 2b). This differing level of skill at the shortest leads suggests that the spatial extent of CG lightning occurrence is more predictable using the CP proxy than is the number of CG flashes. There is a small increase in the anomaly correlation over the first one or two forecast days that is perhaps due to model spinup. In addition to lowering the forecast correlation at the shortest leads, removal of the annual cycle results in forecast correlations that drop off steadily until about forecast day 15. When daily anomalies are computed with respect to a monthly climatology, forecast correlations are quite small but statistically significant for targets beyond forecast day 15. The likely reason for this behavior is that removal of a monthly climatology does not completely remove the impact of seasonality because some months experience a large seasonal cycle within the month. For instance, daily numbers of CG flashes are higher on average during the first half of August than during the second half. Correlations of forecast anomalies with respect to a daily climatology are slightly smaller and become statistically insignificant at about forecast day 15 for CONUS flash count forecasts and slightly longer for forecasts of NGCG > 1.

Forecast skill is not uniformly distributed throughout the year. Forecast correlations at short leads are noticeably lower for forecasts with starts during June–October, slightly more so for forecasts of flash counts than for forecasts of NGCG > 1 (Figures 2c and 2d). The lower level of skill for starts during the warm season likely reflects lower skill in predicting precipitation and CAPE then, as well as the poorer performance of the proxy itself during these months. The skill of U.S. precipitation forecasts tends to be lower during warm months when precipitation is dominated by small-scale convective events (Fritsch et al., 1986; Olson et al., 1995; Sukovich et al., 2014). The differing spatial scale may also be a reason for the correlation between daily CONUS numbers of CG flashes and corresponding CP values computed from reanalysis being lower during the months of June–October than in other months (Tippett et al., 2018). Forecast correlations are significant up to forecast day 10 for starts in most months. The threshold for statistical significance is higher when the forecasts are stratified by start month because the sample size is smaller by about a factor of 12.

There are notable regional variations in forecast skill as well (Figures 2e and 2f). Forecast anomaly correlations for CG flash numbers are initially highest in the northwest and south regions and lowest for the southeast, southwest, and west regions. The regional differences in correlation at the shortest lead are suggestive of problems with the proxy rather than lack of forecast skill. Monthly averages of CP computed using reanalysis do not match southeast observations well during the warm season and do not capture the annual cycle range in the southwest, and west regions (Tippett et al., 2018). Forecast anomaly correlations for CG flash numbers for the northeast drop fastest with increasing lead. There is less regional dependence in the forecast anomaly

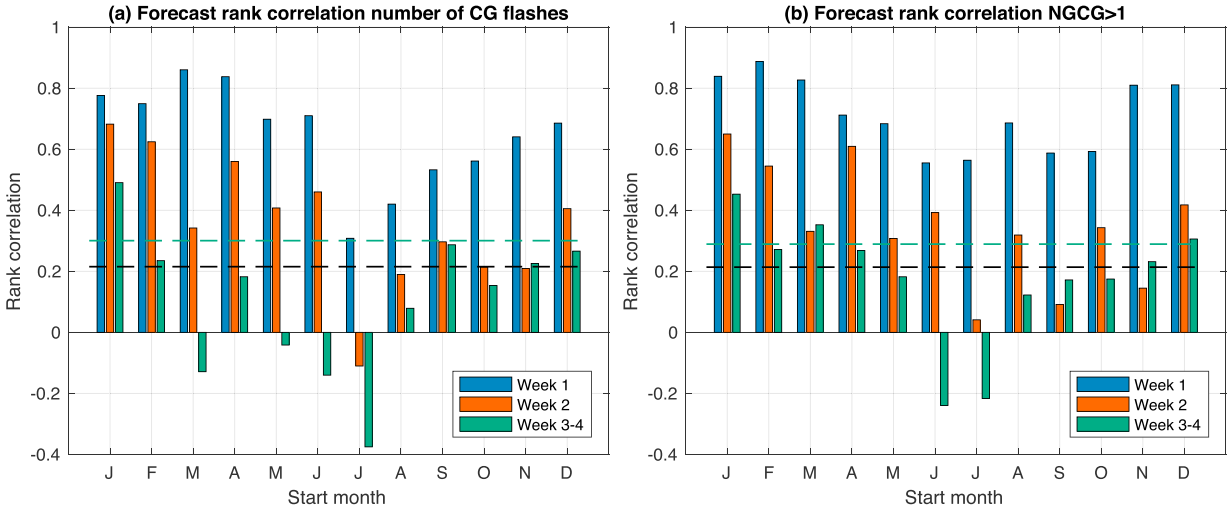

Figure 3. Forecast correlation for weeks 1, 2, and 3–4 (a) number of CG flashes and (b) number of grids with one or more CG flashes (NGCG > 1). Black dashed lines show the threshold for statistical significance of weeks 1 and 2 forecasts. The green dashed lines show the threshold for statistical significance of weeks 3–4 forecasts. CG = cloud-to-ground; NGCG = number of grids with CG.

correlation of NGCG > 1. Only the west region stands out as initially having lower skill than other regions, and again correlations drop fastest with increasing lead for the northeast region.

Averaging over target periods has the potential to increase forecast skill. The skill of week-1 forecasts of numbers of CG flashes and NGCG > 1 are statistically significant in all months but vary considerably by start month (Figure 3). The forecast correlations for NGCG > 1 tend to be higher than those for numbers of CG flashes. The lowest forecast correlations for week-1 forecasts of the number of CG flashes are in July and August, which is when the daily forecast performance is also poor. Week-1 forecasts of NGCG > 1 also have lower skill in July and August, but not to the same extent as forecasts of CG flash numbers, again indicating that the proxy-based predictability of lightning extent is greater than that of the number of flashes. Week-2 forecast correlations are highest for starts in January, February, and April, and there is less, if any, advantage of forecasts of NGCG > 1 over forecasts of the number of CG flashes. Weeks 3–4 forecast skill is only statistically significant for starts in January for forecasts of the number of flashes, and for starts in January, March, and December for forecasts of NGCG > 1. Weeks 3–4 forecast correlations are negative for starts in June and July. The higher skill of weeks 3–4 forecasts during the cool season could be due to the larger spatial scales, as well as the presence of stronger tropical forcings. DelSole et al. (2017) found that U.S. weeks 3–4 precipitation forecasts with January starts had higher skill than ones with July starts and that the predictable components of winter precipitation were related to ENSO and the MJO.

The skill of regional weeks 1, 2, and 3–4 forecasts has many similar features to that of the CONUS-level forecasts. The correlation of regional forecasts of week-1 numbers of CG flashes are statistically significant in all months and regions except in the northeast (October starts), northwest (March starts), and west (November starts); the correlations are barely significant in the southeast (July starts), and the west (March starts). The correlations of weeks 3–4 forecasts of regional numbers of CG flashes are statistically significant in the south (January, February, and April starts), central (November starts), upper midwest (February and September starts), southwest (March and May starts), northwest (April starts), and west (February, May, and December starts) regions. The skill of forecasts of regional NGCG > 1 is similar in structure, but week-1 correlations show less decline during warm month, especially in the south and southeast (Figure S3).

We use contingency table skill measures to compare how well forecasts are able to match the locations where one or more CG flashes are observed (Figure 4a). Probability of detection (POD) starts at over 65% and soon after forecast day 10 is at the level of a no-skill forecast that includes seasonality. The 95% interval of the no-skill forecast is greater at shorter leads because forecasts are spatially sharper at short leads and more diffuse at longer leads. The false alarm rate (FAR) starts at around 35% and rises to the no-skill rate of nearly 60% around forecast day 10. The threat score (TS) and Heidke skill score (HSS) show the same behavior with increasing lead time. Since these are forecasts of CG lightning occurrence rather than number of flashes, there is little

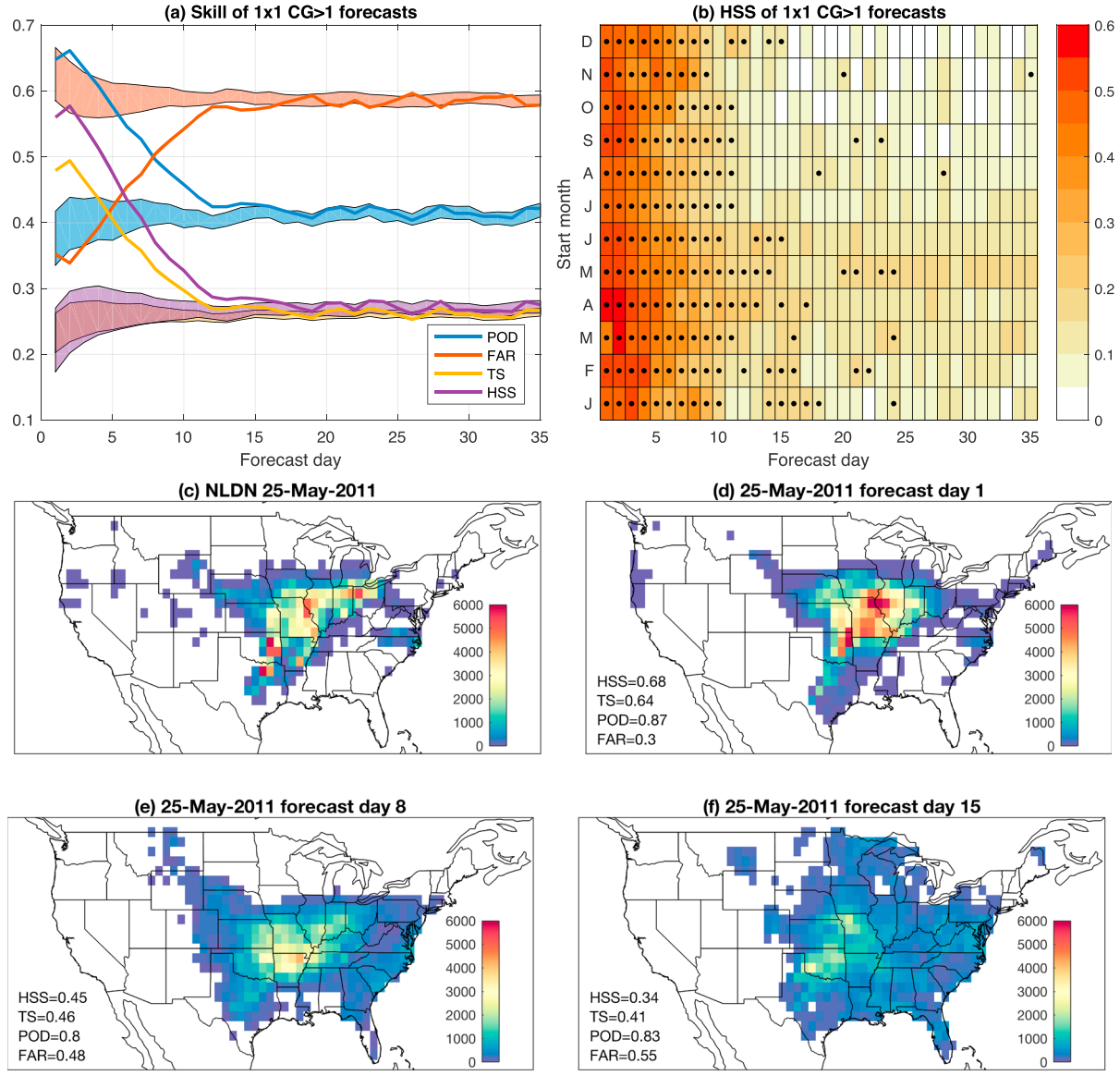

Figure 4. (a) POD, FAR, TS, and HSS as a function of forecast day for 1° × 1° forecasts of one or more CG flashes. Shading indicates the 95% interval of a no-skill forecast. (b) HSS as a function of start month and forecast day. Dots indicate HSS values that are greater than the 95th percentile of the HSS of a no-skill forecast; 25 May 2011 (c) NLDN CG lightning, and days (d) 1, (e) 8, and (f) 15 forecasts. Only CG values above 1, and CP values above the corresponding forecast thresholds are shaded. POD = Probability of Detection; FAR = False Alarm Rate; TS = Threat Score; HSS = Heidke Skill Score; CG = cloud-to-ground; NLDN = National Lightning Detection Network.

dependence of HSS values on start month (Figure 4b). HSS tends to be statistically insignificant after forecast day 10 in most months. We illustrate the character of the CP forecasts for 25 May 2011, when large numbers of CG flashes were observed over a wide area of the United States. The day-1 forecast for CG numbers is very similar in amplitude and spatial extent to the CG flashes recorded. The next forecast available for that target is the day-8 forecast which delineates less well the area with 1 or more CG flashes and has a higher FAR. At forecast day 15, the forecast amplitude is muted, and the area where lightning is forecast to occur is too large.

4. Summary and Conclusions

Studies of weather predictability have found that large-scale variables such as sea-level pressure and 500 hPa heights are predictable beyond the traditional 2-week predictability limit (Buizza & Leutbecher, 2015; Simmons & Hollingsworth, 2002). Fewer studies have examined the predictability of precipitation or quantities related to convection. Here we have examined the predictability of CG lightning, an aspect of

thunderstorm activity with important societal impacts. While predictability is not the same as forecast skill, the forecast skill of a particular forecast system provides a lower limit for predictability. Forecast skill is a lower limit because forecast system improvements have the potential to reveal higher levels of predictability. Here we have assessed forecasts of U.S. CG lightning using a lightning proxy constructed from the output of the GEFS. The lightning proxy used here is the product of CAPE and precipitation, and previous studies have shown that this proxy is proportional to observed numbers of U.S. CG flashes well when the proxy is computed using observations (Romps et al., 2014) and reanalysis (Tippett et al., 2018).

We examined three types of forecasts. The first type are forecasts of the number of CG flashes aggregated over the CONUS, and over NOAA climate regions which consist of several states. Aggregating all starts, forecasts of the daily numbers of CG flashes have skill that is statistically significant until about forecast day 15, both at the CONUS and regional levels, except for the northeast and southeast regions where skill drops below the level for statistical significance at about forecast day 10. The relatively poor performance in these two regions is notable because substantial CG lightning activity occurs in these regions. Forecasts of the daily number of CONUS CG flashes have lower skill for starts during the warm period June–October, when skill drops below the level of statistical significance around forecast day 10. Forecasts of week-1 CONUS CG numbers have statistically significant skill for starts in all months, but week-2 forecast correlations are quite low for starts during July–November. Forecasts of weeks 3–4 CONUS CG numbers only have statistically significant skill for starts in January and have negative skill for starts in June and July.

The second type of forecast is that of the number of $1° \times 1°$ grids with one or more CG flashes recorded, denoted NGCG > 1, and again aggregated over the CONUS, and over NOAA climate regions. Overall, forecast skill at the shortest leads is higher for forecasts of NGCG > 1 than for forecasts of the number of CG flashes, but both lose statistical significance at about the same lead time. The reduced forecast skill during the warm period June–October is less pronounced for forecasts of NGCG > 1 than for forecasts of the numbers of CG flashes. This difference suggests that the spatial extent of CG lighting is more predictable by this proxy than is the CG lighting rate, which is also consistent with the skill of week-1 forecasts. The third type of forecast are daily maps at $1° \times 1°$ spatial resolution showing where one or CG flashes are forecast to occur. Forecast skill values saturate at the level of a no-skill forecast around forecast day 10 and do not show indications of seasonal dependence.

The success of these simple proxy-based forecasts means that higher levels of forecast skill would be expected from more realistic and sophisticated models that include explicit lightning physics and permit convection (Fierro et al., 2013) or that diagnose lightning flash densities from the output of convective parameterization schemes (Lopez, 2016). Nevertheless, the analysis here provides an initial baseline for the predictability of U.S. CG lightning. We have assessed average forecast skill, and an outstanding question is if situations with above-average skill or below-average skill can be identified in advance. A related question is whether predictability limits can be extended using either probabilistic or multimodel approaches. Whether these results have any implication for the long-range prediction of severe thunderstorms is unclear since the favorability of conditions for severe thunderstorm occurrence depends on additional factors such as vertical wind shear that are not part of the lightning proxy.

Acknowledgments
SubX data are available at http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/. We acknowledge the agencies that support the SubX system, and we thank the climate modeling groups (Environment Canada, NASA, NOAA/NCEP, NRL and University of Miami) for producing and making available their model output. NOAA/MAPP, ONR, NASA, and NOAA/NWS jointly provided coordinating support and led development of the SubX system. M. K. Tippett acknowledges the support of NOAA's Modeling, Analysis, Predictions and Projections (MAPP) Program, grant NA16OAR4310145. We are also thankful for the support from NASA Program NNH14ZDA001N-INCA (Climate Indicators and Data Products for Future National Climate Assessments; Jack Kaye and Lucia Tsaoussi, NASA Headquarters). The authors gratefully acknowledge Vaisala Inc. for providing the NLDN data used in this study. The authors thank Eugene W. McCaul Jr and two anonymous reviewers for their helpful comments.

References

Abatzoglou, J. T., & Brown, T. J. (2009). Influence of the Madden–Julian oscillation on summertime cloud-to-ground lightning activity over the continental United States. *Monthly Weather Review*, *137*, 3596–3601. https://doi.org/10.1175/2009MWR3019.1

Allen, J. T., Tippett, M. K., & Sobel, A. H. (2015). Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geoscience*, *8*, 278–283. https://doi.org/10.1038/ngeo2385

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, *144*, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Biagi, C. J., Cummins, K. L., Kehoe, K. E., & Krider, E. P. (2007). National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *Journal of Geophysical Research*, *112*, D05208. https://doi.org/10.1029/2006JD007341

Buizza, R., & Leutbecher, M. (2015). The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, *141*, 3366–3382. https://doi.org/10.1002/qj.2619

Cummins, K. L., & Murphy, M. J. (2009). An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN. *IEEE Transactions on Electromagnetic Compatibility*, *51*, 499–518. https://doi.org/10.1109/TEMC.2009.2023450

DelSole, T., Trenary, L., Tippett, M. K., & Pegion, K. (2017). Predictability of week 3-4 average temperature and precipitation over the contiguous United States. *Journal of Climate*, *30*, 3499–3512. https://doi.org/10.1175/JCLI-D-16-0567.1

Dewan, A., Ongee, E. T., Rafiuddin, M., Rahman, M. M., & Mahmood, R. (2018). Lightning activity associated with precipitation and CAPE over Bangladesh. *International Journal of Climatology*, *38*, 1649–1660. https://doi.org/10.1002/joc.5286

Dowdy, A. J. (2016). Seasonal forecasting of lightning and thunderstorm activity in tropical and temperate regions of the world. *Scientific Reports*, *6*(20), 874. https://doi.org/10.1038/srep20874

Fierro, A. O., Mansell, E. R., MacGorman, D. R., & Ziegler, C. L. (2013). The implementation of an explicit charging and discharge lightning scheme within the WRF-ARW model: Benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. *Monthly Weather Review*, *141*, 2390–2415. https://doi.org/10.1175/MWR-D-12-00278.1

Fritsch, J. M., Kane, R. J., & Chelius, C. R. (1986). The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *Journal of Applied Meteorology and Climatology*, *25*, 1333–1345. https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2

Fuchs, B. R., Rutledge, S. A., Bruning, E. C., Pierce, J. R., Kodros, J. K., Lang, T. J., et al. (2015). Environmental controls on storm intensity and charge structure in multiple regions of the continental United States. *Journal of Geophysical Research: Atmospheres*, *120*, 6575–6596. https://doi.org/10.1002/2015JD023271

Jung, E., & Kirtman, B. P. (2016). Can we predict seasonal changes in high impact weather in the United States? *Environmental Research Letters*, *11*, 074018. https://doi.org/10.1088/1748-9326/11/7/074018

Karl, T. R., & Koss, W. J. (1984). Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983. In *Historical climatology series 4-3* (p. 38). Asheville, NC: National Climatic Data Center.

Koshak, W. J., Cummins, K. L., Buechler, D. E., Vant-Hull, B., Blakeslee, R. J., Williams, E. R., & Peterson, H. S. (2015). Variability of CONUS lightning in 2003–12 and associated impacts. *Journal of Applied Meteorology and Climatology*, *54*, 15–41. https://doi.org/10.1175/JAMC-D-14-0072.1

Lepore, C., Tippett, M. K., & Allen, J. T. (2018). CFSv2 monthly forecasts of tornado and hail activity. *Weather Forecasting*, *33*, 1283–1297. https://doi.org/10.1175/WAF-D-18-0054.1

Lopez, P. (2016). A lightning parameterization for the ECMWF Integrated Forecasting System. *Monthly Weather Review*, *144*, 3057–3075. https://doi.org/10.1175/MWR-D-16-0026.1

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*, 505–513. https://doi.org/10.1111/j.2153-3490.1982.tb01839.x

Lynn, B. H. (2017). The usefulness and economic value of total lightning forecasts made with a dynamic lightning scheme coupled with lightning data assimilation. *Weather Forecasting*, *32*, 645–663. https://doi.org/10.1175/WAF-D-16-0031.1

Lynn, B. H., Yair, Y., Price, C., Kelman, G., & Clark, A. J. (2012). Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Weather Forecasting*, *27*, 1470–1488. https://doi.org/10.1175/WAF-D-11-00144.1

McCaul, E. W., Goodman, S. J., LaCasse, K. M., & Cecil, D. J. (2009). Forecasting lightning threat using cloud-resolving model simulations. *Weather Forecasting*, *24*, 709–729. https://doi.org/10.1175/2008WAF2222152.1

Muñoz, Á. G., Díaz-Lobatón, J., Chourio, X., & Stock, M. J. (2016). Seasonal prediction of lightning activity in North Western Venezuela: Large-scale versus local drivers. *Atmospheric Research*, *172–173*, 147–162. https://doi.org/10.1016/j.atmosres.2015.12.018

Nag, A., Murphy, M. J., & Cramer, J. A. (2016). Update to the U.S. National Lightning Detection Network. In *24th International Lightning Detection Conference & 6th International Lightning Meterology Conference*. San Diego, CA.

Olson, D. A., Junker, N. W., & Korty, B. (1995). Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Weather Forecasting*, *10*, 498–511. https://doi.org/10.1175/1520-0434(1995)010<0498:EOYOQP>2.0.CO;2

Petersen, W. A., & Rutledge, S. A. (1998). On the relationship between cloud-to-ground lightning and convective rainfall. *Journal of Geophysical Research*, *103*, 14,025–14,040. https://doi.org/10.1029/97JD02064

Romps, D. M., Seeley, J. T., Vollaro, D., & Molinari, J. (2014). Projected increase in lightning strikes in the United States due to global warming. *Science*, *346*, 851–854. https://doi.org/10.1126/science.1259100

Simmons, A. J., & Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *128*, 647–677.

Stern, H., & Davidson, N. E. (2015). Trends in the skill of weather prediction at lead times of 1–14 days. *Quarterly Journal of the Royal Meteorological Society*, *141*, 2726–2736. https://doi.org/10.1002/qj.2559

Sukovich, E. M., Ralph, F. M., Barthold, F. E., Reynolds, D. W., & Novak, D. R. (2014). Extreme quantitative precipitation forecast performance at the Weather Prediction Center from 2001 to 2011. *Weather Forecasting*, *29*, 894–911. https://doi.org/10.1175/WAF-D-13-00061.1

Thompson, D. B., & Roundy, P. E. (2013). The relationship between the Madden-Julian oscillation and U.S. violent tornado outbreaks in the spring. *Monthly Weather Review*, *141*, 2087–2095. https://doi.org/10.1175/MWR-D-12-00173.1

Tippett, M. K., Lepore, C., Koshak, W. J., Chronis, T., & Vant-Hull, B. (2018). Performance of a simple proxy for U.S. cloud-to-ground lightning. *Weather Forecasting*. https://arxiv.org/abs/1810.00930

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences: An Introduction*. New York: Academic Press.

Zhou, X., Zhu, Y., Hou, D., Luo, Y., Peng, J., & Wobus, R. (2017). Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Weather Forecasting*, *32*, 1989–2004. https://doi.org/10.1175/WAF-D-17-0023.1

Zhu, H., Wheeler, M. C., Sobel, A. H., & Hudson, D. (2013). Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, *142*, 1556–1569. https://doi.org/10.1175/MWR-D-13-00222.1

Erratum

In the originally published version of this article, the reference Ashraf, D., Ongee, E. T., Rafiuddin, M., Rahman, Md. M., & Rezaul, M.(2017). Lightning activity associated with precipitation and CAPE over Bangladesh. *International Journal of Climatology*, 38, 1649–1660. https://doi.org/10.1002/joc.5286, should have been published as Dewan, A., Ongee, E. T., Rafiuddin, M., Rahman, M. M., & Mahmood, R. (2018). Lightning activity associated with precipitation and CAPE over Bangladesh. *International Journal of Climatology*, 38, 1649–1660. https://doi.org/10.1002/joc.5286. This reference and its citation in the article text have since been corrected, and the present version may be considered the authoritative version of record.