ENSO and MJO Modulation of U.S. Cloud-to-ground Lightning Activity

Kelsey Malloy,[a] Michael K. Tippett,[a] William J. Koshak[b]

[a] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York*

[b] *Earth Science Branch, NASA Marshall Space Flight Center, Huntsville, Alabama*

Corresponding author: Kelsey Malloy, kmm2374@columbia.edu

ABSTRACT: Cloud-to-ground (CG) lightning substantially impacts human health and property. However, the relations between U.S. lightning activity and the Madden-Julian Oscillation (MJO) and El Niño-Southern Oscillation (ENSO), two predictable drivers of global climate variability, remain uncertain, in part because most lightning datasets have short records that cannot robustly reveal MJO- and ENSO-related patterns. To overcome this limitation, we developed an empirical model of 6-hourly lightning flash count over the contiguous U.S. (CONUS) using environmental variables (convective available potential energy and precipitation) and National Lightning Detection Network data for 2003–2016. This model is shown to reproduce the observed daily and seasonal variability of lightning over most of CONUS. Then, the empirical model was applied to construct a proxy lightning dataset for the period 1979–2021, which was used to investigate the summer MJO-lightning relationship at daily resolution and the winter-spring ENSO-lightning relationship at seasonal resolution. Overall, no robust relationship between MJO phase and lightning patterns was found when seasonality was taken into consideration. El Niño is associated with increased lightning activity over the Coastal Southeast U.S. during early winter, the Southwest during winter through spring, and the Northwest during late spring, whereas La Niña is associated with increased lightning activity over the Tennessee River Valley during winter.

SIGNIFICANCE STATEMENT: Cloud-to-ground lightning is dangerous for humans via direct strikes or through triggering wildfires, generating air pollution, etc. How lightning activity can be affected by climate remains unclear, and it is challenging to study their links because the data record for lightning is short. With the available lightning record, we developed a model that relates lightning flash counts over U.S. to environmental factors. This model well represents observed fluctuations in daily and seasonal lightning over most of the U.S. Because the model only needs environmental information to predict lightning flash counts, we were able to construct a longer record of predicted lightning based on the longer data record of environmental variables. With this dataset, we investigated the links between lightning and climate and found that the state of sea surface temperatures in the tropical Pacific (El Niño-Southern Oscillation) is linked to changes in U.S. lightning patterns in winter and spring.

1. Introduction

Cloud-to-ground (CG) lightning has large impacts on ecosystem health as well as human life and property (Koshak et al. 2015). In particular, CG lightning is a natural trigger for wildfires and can generate ground-level nitrogen oxides that lead to the production of tropospheric ozone (Kang et al. 2020). More generally, lightning characteristics, such as flash rate and polarity, have been linked to thunderstorm formation and structure (Goodman and MacGorman 1986; Branick and Doswell III 1992) as well as severe weather events, such as hail and tornadoes (MacGorman and Burgess 1994).

Although lightning has substantial impacts, relatively little is understood about its relation to large-scale climate variability. For instance, only a few studies have investigated the relationship between lightning and the El Niño-Southern Oscillation (ENSO), which is the dominant mode of predictable climate variability on seasonal and interannual timescales. Dowdy (2016) showed a statistical relation between ENSO and seasonal lightning activity in over 20% of the region 35°S–35°N in some seasons, and Muñoz et al. (2016) used ENSO and other sea surface temperature (SST) patterns to predict seasonal lightning activity in North Western Venezuela. LaJoie and Laing (2008) and Laing et al. (2008) found a spatially sporadic winter ENSO-lightning relationship over the Gulf Coast with an increase of lightning activity during El Niño in an 8-year dataset (1995–2002). Koehler (2020) also found a positive relation between El Niño and U.S. lightning activity

in a 26-year dataset (1993–2018). However, their use of an annual (January–December) ENSO index and contiguous U.S. (CONUS) totals precluded the resolution of spatial and temporal details of the ENSO-lightning relationship and mixed ENSO states from one winter to the next. Overall, an ENSO influence on U.S. lightning activity is plausible since ENSO explains a considerable proportion of U.S. precipitation variability (L'Heureux et al. 2015; Nigam and Sengupta 2021) and has been found to modulate the frequency of hail and tornadoes in winter and spring (Allen et al. 2015b; Koch et al. 2021; Tippett and Lepore 2021).

Likewise, few studies have examined the relation between lightning activity and the Madden-Julian Oscillation (MJO), which is the dominant mode of predictable climate variability on sub-seasonal timescales. Lightning density in MJO-associated tropical mesoscale convective systems evolves with MJO life cycle (Virts and Houze 2015). Abatzoglou and Brown (2009) linked June–September anomalous lightning activity to MJO phase in a 18-year dataset (1990–2007). Once again, modulation of U.S. lightning activity by the MJO is plausible since the MJO modulates U.S. precipitation (Becker et al. 2011) and may modulate spring and summer tornado activity (Thompson and Roundy 2013; Baggett et al. 2018; Kim et al. 2020), though this relationship is less robust than the ENSO-tornado relationship (Tippett 2018; Moore and McGuire 2020).

A challenge in investigating the relation between lightning and climate is that the ideal tools for studying climate relations — observational data and physics-based models — have their limitations. Lightning datasets often are relatively short or inhomogeneous, making it difficult to identify significant relationships between lightning activity and climate modes of variability as well as to discern trends due to a changing climate (Romps et al. 2014; Romps 2019; Finney et al. 2018). The most realistic physics-based representations of lightning are those used for weather, not climate timescales. In convection-permitting models, lightning is often parameterized based on cloud dynamics and microphysics (Price and Rind 1994; Michalon et al. 1999; Mansell et al. 2005; McCaul et al. 2009; Yair et al. 2010; Lynn et al. 2012; Fierro et al. 2013). However, these model simulations tend to be relatively short with few ensemble members, making it difficult to extract a robust ENSO or MJO signal in lightning activity compared to the model-based studies for ENSO- or MJO-modulated temperature and precipitation patterns.

Empirical models provide an alternative approach by predicting lightning characteristics based on environmental factors (Romps et al. 2014; Stolz et al. 2017). Romps et al. (2014) developed

a simple lightning proxy based on the product of convective available potential energy (CAPE), precipitation, and a scaling factor. This CAPE times precipitation (CP) proxy is attractive for many reasons. CAPE and precipitation are readily available in many reanalysis products and forecast/climate models (Saha et al. 2014; Vitart et al. 2017; Lepore et al. 2018; Pegion et al. 2019), and both CAPE (Jung and Kirtman 2016) and precipitation (Becker et al. 2014; DelSole et al. 2017) are forecasted with some skill on long-range timescales. Despite its simplicity, the CP proxy represents subseasonal CG lightning activity fairly well (Tippett and Koshak 2018). In addition, there is theoretical basis for how CAPE and precipitation might be modulated on various timescales. Also, the relationships between lightning, CAPE, and precipitation are well documented (Petersen and Rutledge 1998; Murugavel et al. 2012; Xu et al. 2013). For example, Romps et al. (2014) and Murugavel et al. (2012) estimated future increasing lightning trends in climate projections, using the corresponding CAPE and precipitation projections to explain these trends.

Yet the CP proxy has substantial limitations. While its simple calculation is attractive for practical purposes, a proportionality constant is the only tunable parameter. The CP proxy cannot capture land-ocean lightning contrasts (Romps et al. 2018), attesting to its missing physical processes. Consequently, its ability to capture CG lightning variability differs by region and season. For example, Tippett et al. (2019) analyzed CP proxy performance in representing CG lightning climatology and variability and found that CP proxy performance declined west of the Rocky Mountains and during the summer season. Other studies that analyzed the CP proxy in other regions of the world found that its performance varied by region or season, e.g., it under-performed during Indian monsoon season relative to the pre-monsoon season (Dewan et al. 2018) and under-performed in the plateau region of southwest China relative to other terrains (Zhao et al. 2022). However, the CP proxy may still be a useful tool when and where its performance is adequate.

The purposes of this study are to (1) develop a CG lightning proxy that generalizes its functional dependence on CAPE and precipitation, (2) evaluate the performance in capturing observed lightning variability on daily and seasonal timescales during the 2003–2016 period, and (3) demonstrate use of the CG lightning proxy in assessing the influence of climate variability on CG lightning activity during the longer 1979–2021 period. The data and methods used for constructing the lightning proxy and conducting the remaining analysis are detailed in Section 2. In Section 3,

the results are presented in three parts. First, the new lightning proxy is compared to lightning observations and the CP proxy for the 2003–2016 period. Then, two applications of the lightning proxy are explored: 1979–2021 daily proxy data is used to analyze the MJO-lightning relationship in summer, and 1979–2021 seasonal proxy data is used to analyze the ENSO-lightning relationship in winter through late spring. These example applications complement previous work using shorter records of lightning observations (e.g. Abatzoglou and Brown 2009; LaJoie and Laing 2008; Laing et al. 2008). The longer lightning proxy could provide further insight into these relationships. In addition, this analysis provides information about the potential for lightning prediction on long-range timescales from MJO and ENSO, two major sources of predictability on subseasonal-to-seasonal timescales. A summary of the results as well as a discussion of its connection to findings from previous literature and future work needed are given in Section 4.

2. Data and Methods

a. Data

CG lightning flash count data are taken from the National Lightning Detection Network (NLDN) for the 2003–2016 period during which the network is reliable and there are few missing days (Cummins and Murphy 2009; Koshak et al. 2015). There are nine days without data, which are excluded from analysis, and there is a notable increase in lightning counts during 2016 that may be from a change in the NLDN Total Lightning Processor (Nag et al. 2016). Lightning counts are summed over CONUS land for every 6 hours (at 00, 06, 12, and 18 UTC) and at a $1^\circ \times 1^\circ$ grid resolution. We use the total number of CG flashes, which is the sum of negative polarity flashes and positive polarity flashes with currents greater than 15 kA. The positive polarity threshold accounts for cloud pulses that are frequently misclassified as low-amplitude, positive CG lightning (Biagi et al. 2007; Cummins and Murphy 2009).

Mixed-layer CAPE (J kg^{-1}) and precipitation (mm day^{-1}) data are taken from North American Regional Reanalysis (NARR) for the 1979–2021 period (Mesinger et al. 2006). NARR provides data at a 3-hour resolution with 32-km native grid spacing, which we convert to a 6-hour and $1^\circ \times 1^\circ$ resolution. Mixed-layer CAPE in NARR is calculated from the mean equivalent potential temperature between the 0–180 hPa layer. Because NARR assimilates latent heating profiles, precipitation is better estimated in NARR than other global reanalyses (Bukovsky and Karoly

2007; Cui et al. 2017). Despite noted deficiencies of CAPE values in reanalysis (Gensini et al. 2014), CAPE biases are smaller in NARR than other reanalysis products (King and Kennedy 2019).

MJO is defined by the real-time multivariate (RMM) index (Wheeler and Hendon 2004), which uses the normalized combined empirical orthogonal function (EOF) of outgoing longwave radiation, 200-hPa zonal wind, and 850-hPa zonal wind between 15°S–15°N to construct eight phases of the MJO. We only consider active MJO phases when the RMM amplitude is greater than or equal to 1.

ENSO is defined by the Oceanic Niño Index (ONI). ONI is taken from the Climate Prediction Center and is computed by averaging the SST anomalies over the Niño 3.4 region (5°N–5°S, 120°–170°W), and a 3-month running mean is applied. El Niño is defined by an ONI greater than or equal to 0.5°C, and La Niña is defined by an ONI less than or equal to -0.5°C.

b. Poisson Regression

We fit the 6-hourly CG lightning flash counts for the 2003–2016 period using a Poisson regression (PR). In the PR model, the probability of the 6-hourly lightning flash count N taking on the values $n = 0, 1, 2, \ldots$ is:

$$P(N = n) = \frac{e^{-\mu}\mu^n}{n!}, \tag{1}$$

where μ is the expected value of N. The expected value μ depends log-linearly on the vector $\mathbf{x}$ of environmental predictors:

$$\mu \sim \exp(\mathbf{b}^T\mathbf{x}), \tag{2}$$

where $\mathbf{b}$ is the vector of corresponding regression coefficients. The regression coefficients are estimated by maximizing the log-likelihood from Eq. 1 of the observed lightning flash counts given the environmental variables. Here, the sample size is 3,676,016 which is the number of land grid points times the number of 6-hourly periods during 2003–2016. We include an intercept term b_0 in the regression as well as an offset term which accounts for the different areas of the 1°x1° grid boxes. Therefore, the expected lightning flash count μ in a grid box and within a 6-hourly interval is:

$$\mu = \exp[b_0 + \mathbf{b}^T\mathbf{x} + \log(\Delta\text{lat}\,\Delta\text{lon}\cos\phi)] \tag{3}$$

where Δlat and Δlon are the longitude and latitude spacings in degrees (both here are 1), respectively, and ϕ is the latitude.

Three functional forms for the dependence of μ on precipitation and CAPE are tested. PR Model 1 uses log(CAPE $\times$ Precipitation) as its predictor variable. PR Model 2 uses log(CAPE) and log(Precipitation) as its predictor variables. PR Model 3 uses log(CAPE), log(Precipitation), $(\log(\text{CAPE}))^2$, and $(\log(\text{Precipitation}))^2$ as its predictor variables. The three models are nested in the sense that PR Model 3 is a special case of PR Model 2, and PR Model 2 is special case of PR Model 1, which facilitate their comparison. Using log(CAPE) and log(Precipitation) as predictors means that μ has the following functional forms:

$$\text{PR Model 1: } \mu \sim (\text{CAPE} \times \text{Precipitation})^{b_1}$$

$$\text{PR Model 2: } \mu \sim (\text{CAPE})^{b_1} (\text{Precipitation})^{b_2}$$

$$\text{PR Model 3: } \mu \sim (\text{CAPE})^{b_1} (\text{Precipitation})^{b_2} \exp[b_3 (\log(\text{CAPE}))^2 + b_4 (\log(\text{Precipitation}))^2]$$

where $\mathbf{b} = \{b_1, b_2, ...\}$ are the corresponding coefficients for each predictor. We calculate deviance to determine goodness of fit, with a smaller deviance indicating a better fit. We perform leave-one-year-out jackknife resampling to quantify the uncertainty in the estimates of the regression coefficients. For example, we leave out 2003 and estimate the PR coefficients with the 2004-2016 data, then leave out 2004 and estimate the PR coefficients with the 2003 and 2005–2016 data, and so on. We use the 2.5 and 97.5 percentile of the resulting coefficients as 95% confidence intervals. The spatial and temporal rank correlations between observed and predicted flash counts are computed to evaluate performance.

After selecting and fitting a PR model, CAPE and precipitation data from 1979–2021 are used in the PR model as a proxy for 6-hourly lightning flash counts, which are aggregated to daily and seasonal resolution. For the MJO analysis, we calculate flash anomaly composites. For the ENSO analysis, we calculate the rank correlation between the 3-month running season flash anomaly (daily mean removed) and ONI. In addition, difference composites of lightning flashes are computed, which is the average of the lightning flashes during El Niño subtracted from the average of the lightning flashes during La Niña.

c. Statistical Significance

A bootstrapping test (with replacement) is used to assess the statistical significance of the temporal rank correlations.

A Wilcoxon rank sum test is used to assess the statistical significance of the difference between El Niño and La Niña composites. The Wilcoxon rank sum test procedure is non-parametric i.e., does not assume known distribution and essentially compares the medians of the composited sets.

We use a two-sided permutation test that preserves day of the year to calculate the statistical significance of the MJO composites. Each iteration of the permutation test scrambles the years of the MJO data while preserving day of the year. We perform 1000 permutations. The p-value of the composite is $2 \times \min(f, 1-f)$ where f is the fraction of permutations whose composites equal or exceed the observed composite. The design of the permutation test accounts for the seasonality and autocorrelation of the data.

The Wilcoxon rank sum and permutation tests are performed at each grid point. The false discovery rate (FDR) procedure from Benjamini and Hochberg (1995) is used to assess the statistical significance of the correlation and (difference) composite maps. The two-sided p-value is calculated at every land grid point and sorted from smallest to largest. Then, the sorted p-values are compared to the sequence α/S, $2\alpha/S$, $3\alpha/S$, ..., where S is the number of land grid points (S = 860) and α is the selected FDR, which we set at 10% here.

3. Results

a. PR Model Development and Performance

We begin by comparing the environmental dependence of the flash count data with that of CP and the three PR models. First, we examine their dependence on CAPE × Precipitation (which is proportional to CP), CAPE, and precipitation separately (Fig. 1). The flash density data binned by CAPE × Precipitation show an approximate power-law dependence (linear on a log-log plot; Fig. 1a black line). The slope of the binned flash density data on the log-log plot is less than one which differs substantially from that of CP (blue line) whose slope is equal to 1 by construction. The PR Model 1 with slope equal to 0.726 (Table 1; Fig. 1a orange line) better matches the flash count data. In the PR Model 2, the CAPE and precipitation exponents are fairly close to each other (0.684 and 0.754) to that of CAPE × Precipitation in the PR Model 1. The dependence of PR Model 2

on CAPE × Precipitation (green line) is nearly indistinguishable from that of PR Model 1 (orange line). With the addition of squared log terms, the PR Model 3 is able to match the departure of the flash data from pure power-law dependence on CAPE × Precipitation. Binning the observed flash data, CP, and the three PR models by CAPE and precipitation separately show similar behavior (Fig. 1b, c). The dependence of CP is mostly linear when binned by CAPE and precipitation separately but this behavior is not guaranteed by the functional form of CP. The mismatch between the dependence of CP and observed flash density on CAPE is larger than that for precipitation.

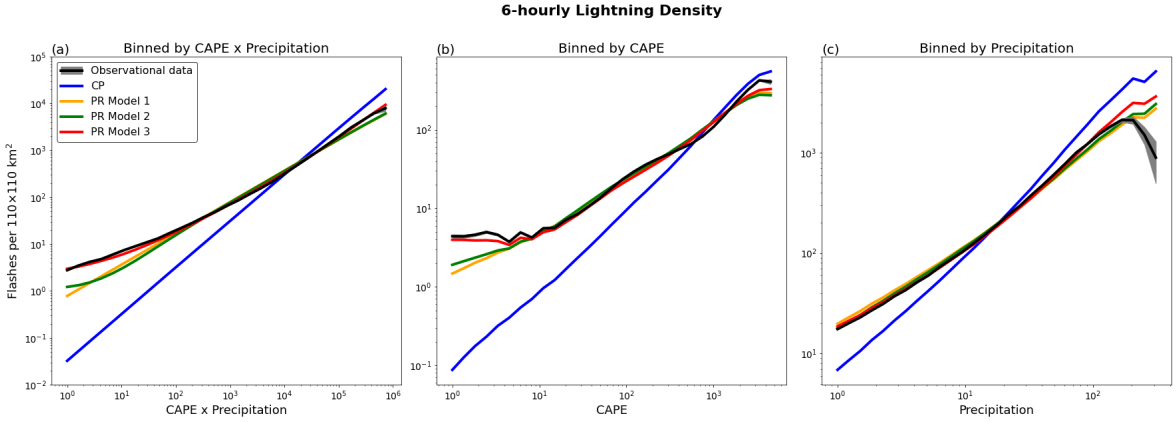


FIG. 1. Comparison of average 6-hourly lightning density (flashes per 110×110 km^2 i.e., 1°x1° grid box at the equator) with the CP proxy and PR models binned by (a) CAPE × Precipitation values, (b) CAPE values, and (c) Precipitation values. Observed data is shown with black line, with the gray shading indicating the expected range of 95% of the means. Values for the CP proxy and PR models are shown with the respective colored lines. See Table 1 for the variables and coefficients of the PR models tested.

TABLE 1. Summary of the Poisson regression models tested.

	PR Model 1		**PR Model 2**		**PR Model 3**	
Deviance	6.493x10^8		6.483x10^8		6.385x10^8	
Predictor	*Coef*	*95% Conf. Int.*	*Coef*	*95% Conf. Int.*	*Coef*	*95% Conf. Int.*
intercept	-0.911	[-0.930, -0.861]	-0.709	[-0.719, -0.680]	0.393	[0.380, 0.418]
log(CP)	0.726	[0.723, 0.730]	–	–	–	–
log(CAPE)	–	–	0.684	[0.683, 0.689]	0.250	[0.249, 0.260]
log(Precipitation)	–	–	0.754	[0.750, 0.760]	0.634	[0.631, 0.641]
log(CAPE)^2	–	–	–	–	0.0390	[0.0385, 0.0400]
log(Precipitation)^2	–	–	–	–	0.0327	[0.0323, 0.0335]

Figure 2 shows the joint dependence of the flash density, CP, and PR Model 3 on CAPE and precipitation. The isolines of the flash density data (Fig. 2a white lines) are not straight lines on the log-log axes, further indicating the departure of flash density data from pure power-law dependence on CAPE and precipitation, especially for low values of precipitation (i.e., CAPE matters more). On the other hand, the isolines of CP are linear. Consequently, CP underestimates flash density for low values of CAPE and precipitation and overestimates flash density for high values of CAPE and precipitation (Fig. 2b, d). The isolines of PR Model 3 better match the behavior of the observed isolines of the flash count data (Fig. 2c). Overall, the PR Model 3 better fits the observed dependence of flash counts on CAPE and precipitation than does the CP proxy, especially for extreme low or high values of CAPE and precipitation (Fig. 2e). The PR Model 3 for 6-hourly lightning is:

$$\mu = \exp[0.393 + 0.25\log(\text{CAPE}) + 0.039(\log(\text{CAPE}))^2 + 0.634\log(\text{Precipitation}) + 0.0327(\log(\text{Precipitation}))^2]\Delta\text{lat}\Delta\text{lon}\cos\phi \quad (4)$$

The remainder of the analysis, including all descriptions of PR predicted lightning, will use PR model 3.

Maps of the total lightning flashes for 2003–2016 show where lightning flashes occur most and least often. Observed lightning flash counts are highest along the Gulf Coast and Great Plains and are lowest along the West Coast (Fig. 3a). This overall pattern is reproduced by the CP proxy and PR model east of the Rocky Mountains (Fig. 3b,c). Both CP and the PR model underestimate flash counts over the Rocky Mountain states (west of 100°W) and overestimate flash counts over the West Coast and Upper Midwest (Fig. 3d,e). The largest bias in magnitude is over the West Coast, while the largest bias in spatial extent is over the Rocky Mountains. The bias of the PR model is reduced compared to that of CP (i.e., ratio closer to 1) over the Rocky Mountains and Upper Midwest (Fig. 3e). However, the PR overestimation of flash counts is increased over the West Coast and Pacific Northwest.

One possible reason that the PR model performs worse than the CP proxy on the West Coast and Pacific Northwest is that the PR model coefficient for log(Precipitation) is greater than log(CAPE) (0.634 versus 0.25, respectively; cf. Table 1), whereas the CP proxy weights CAPE and precipitation equally. Climatologically, CAPE values are low across the western U.S., but the West Coast

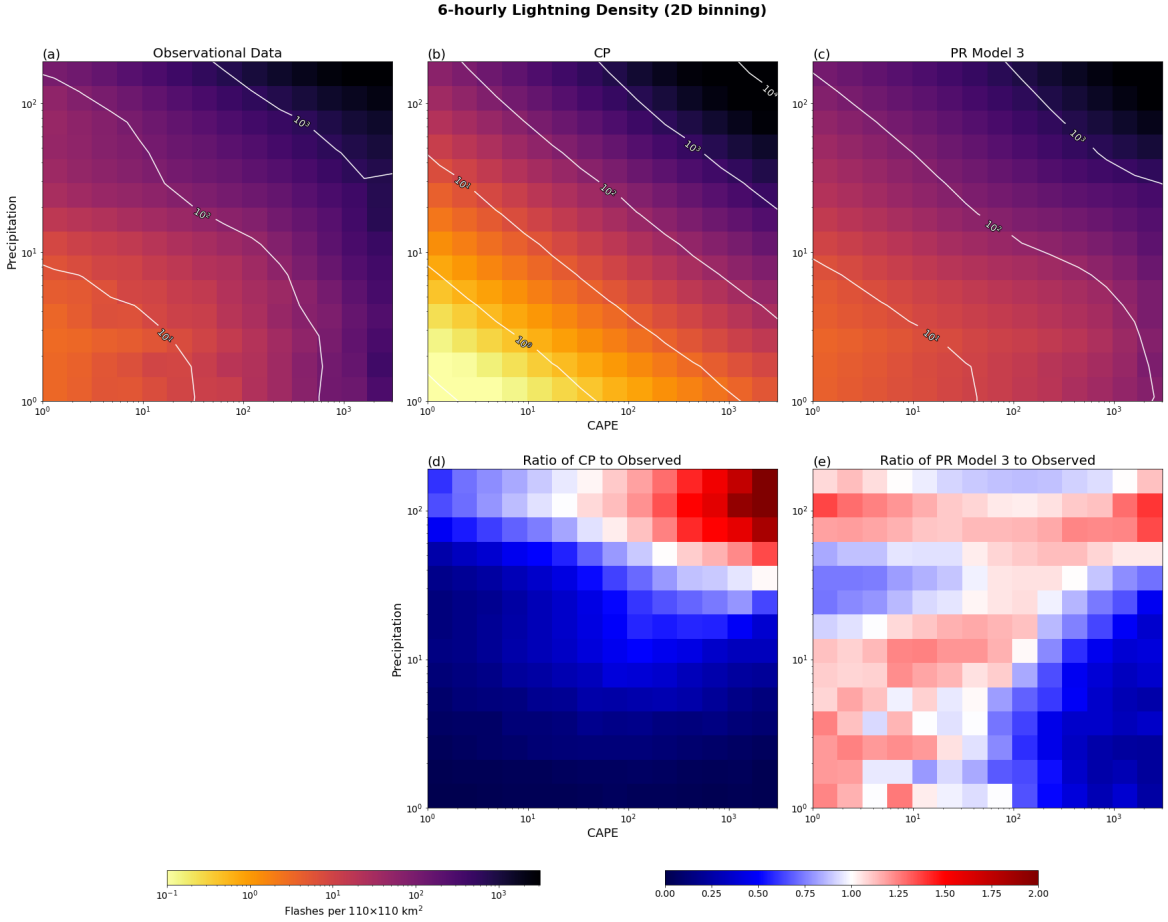
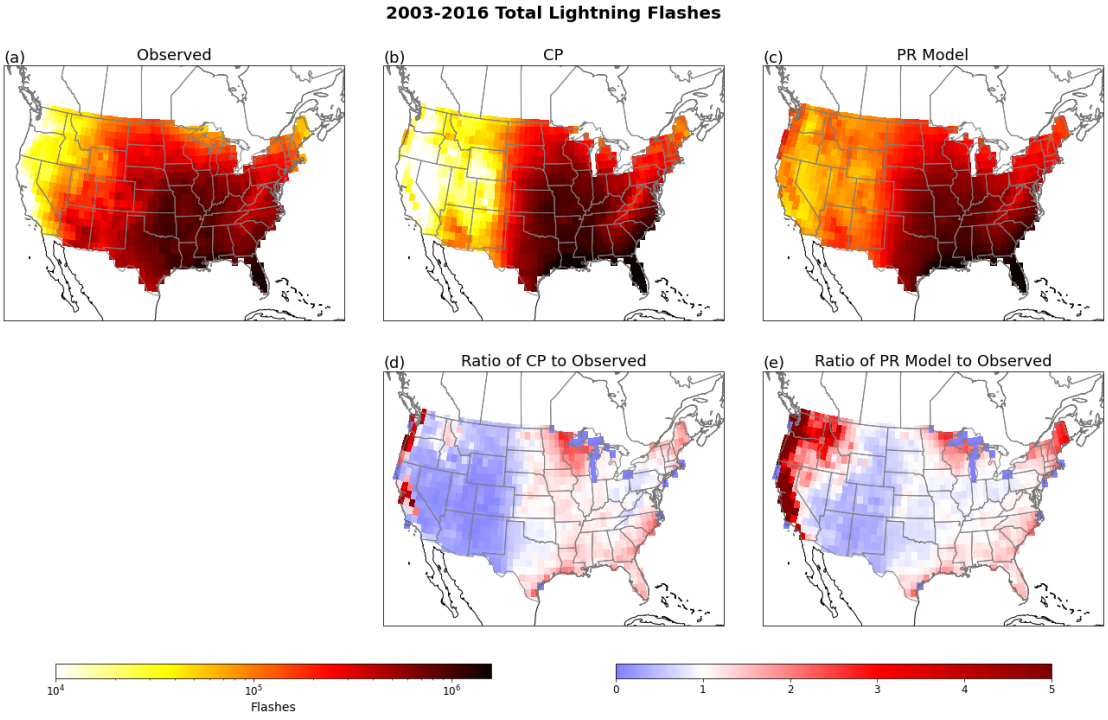


FIG. 2. Comparison of average 6-hourly lightning density values (flashes per 110×110 km^2 i.e., 1°×1° grid box at the equator) binned by CAPE and precipitation: the lightning density values binned by CAPE and precipitation for (a) observations, (b) CP proxy, and (c) PR Model 3, and the ratio of (d) CP proxy predicted lightning density to observed lightning density, and of (e) PR predicted lightning density to observed lightning density.

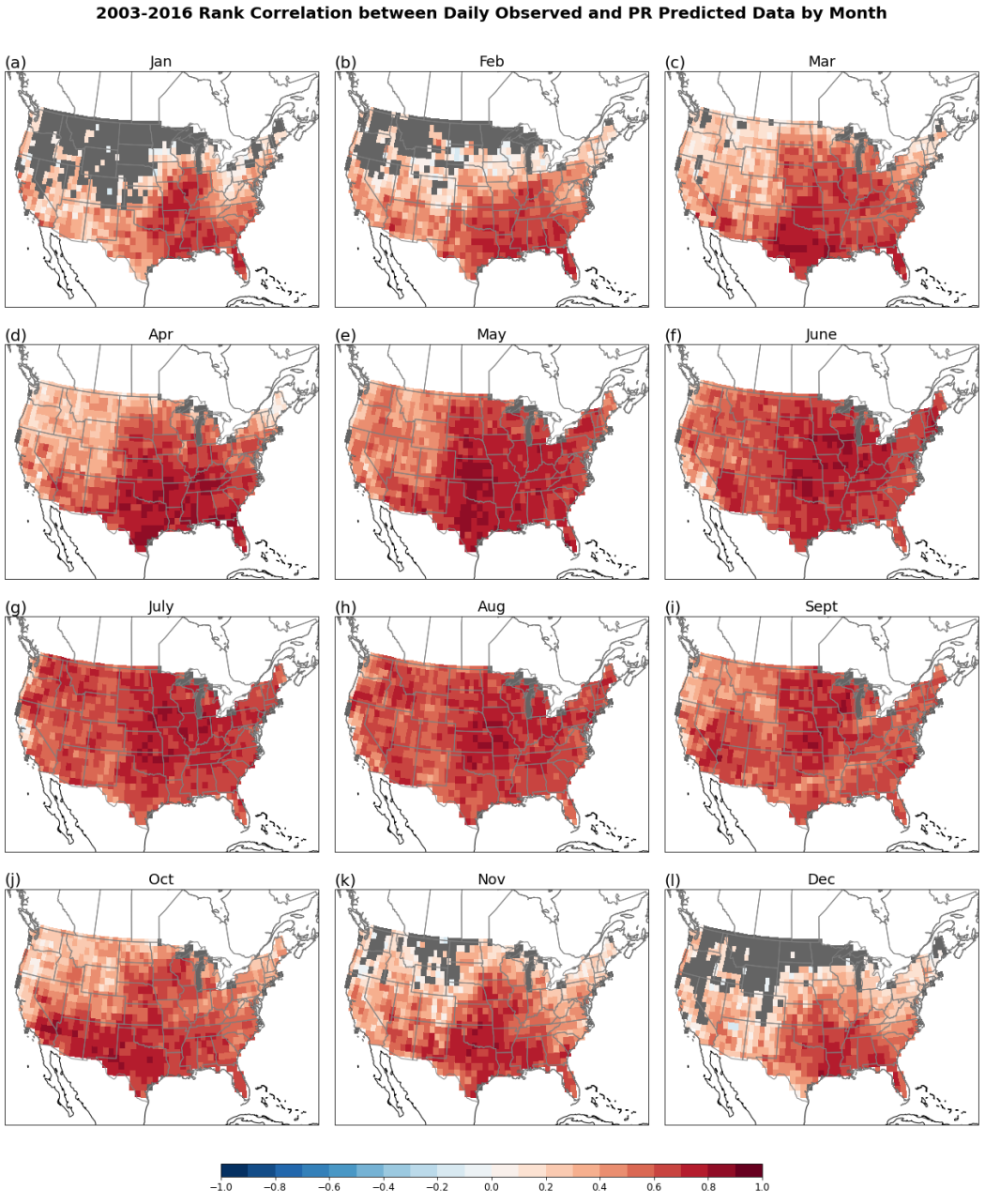
receives higher rainfall than the rest of the western U.S., resulting in the PR model predicting too high lightning flash density. When we trained the PR model with only West Coast (west of 120°W) data, the coefficient for log(Precipitation) decreased to 0.32 and the intercept coefficient decreased to -1.17, which confirms the differing regional sensitivity in the data of flash counts to precipitation. Lightning flash count depends less on precipitation on the West Coast and there are fewer flashes overall compared to the rest of the contiguous U.S., and, by using the same weight for precipitation everywhere, we overestimate lightning flashes over the West Coast. Due to the



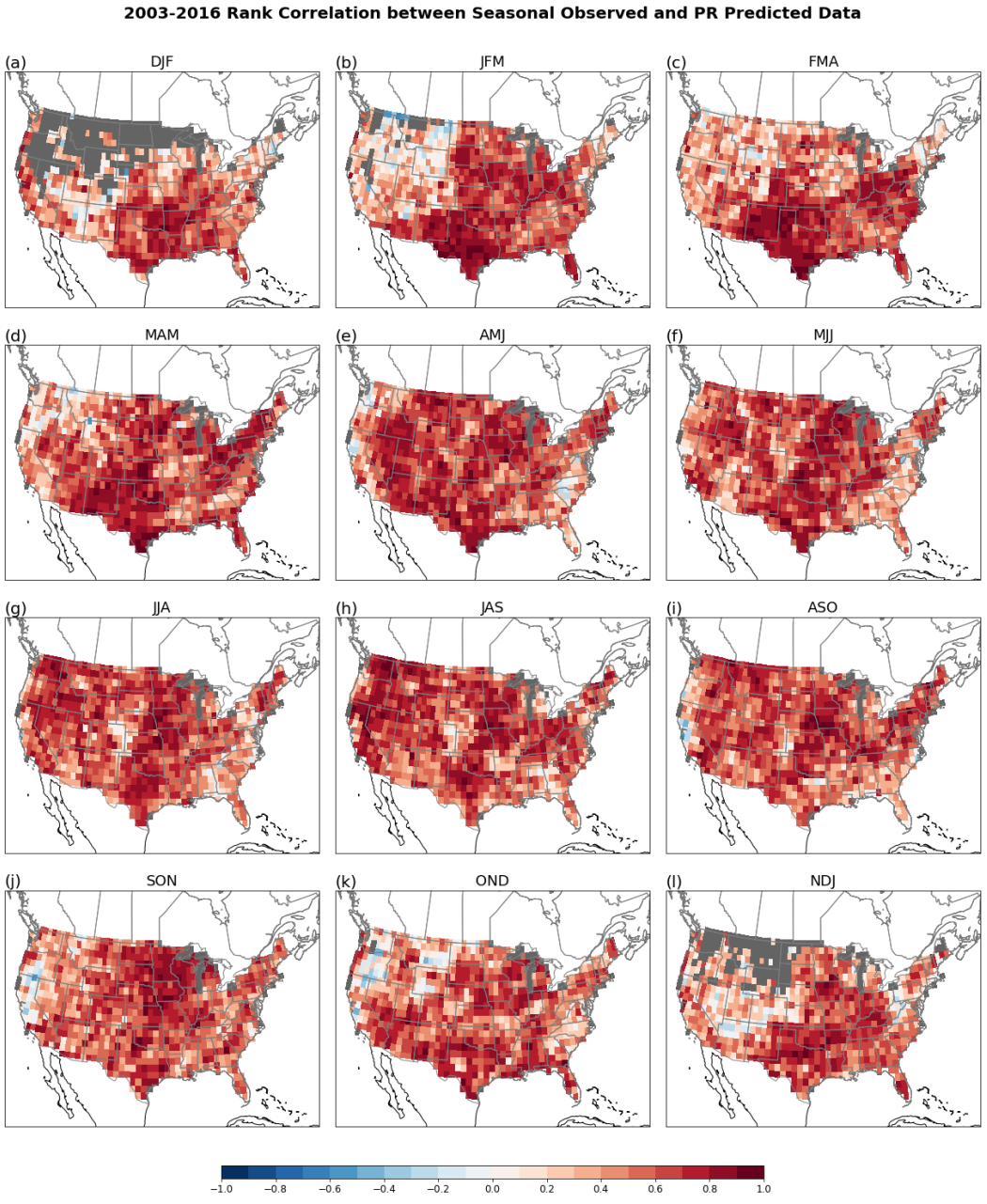
247 FIG. 3. (top) Number of lightning flashes between 2003–2016 in observations, using the CP proxy, and using
248 the PR model. (bottom) Ratio between CP proxy total lightning flashes and observed total lightning flashes, and
249 ratio between PR predicted total lightning flashes and observed total lightning flashes.

271 improved fit of the PR model to the observed data compared to the CP proxy, we drop the CP proxy
272 and focus on the PR model.

279 The degree to which the PR model matches observed variability varies depending on season
280 and region. We focus on the rank correlation between observed and PR predicted flash counts
281 at daily and seasonal time resolution because these are the relevant timescales for MJO and
282 ENSO influences, respectively. PR model performance for daily data is greatest east of the Rocky
283 Mountains during the warm season (April–September). Skill is worse over the Northwest U.S.,
284 Northern Plains, and Northeast U.S. during the cool season (October–March; Fig. 4a-c, k, l). These
285 regions of low skill are located where, climatologically, there are few lightning flashes (but more
286 than 10 lightning flashes occurred during the 14-year period in that month and so were not masked)
287 and low variance in lightning (cf. Fig. 3; Tippett et al. 2019). PR model performance for seasonal
288 totals (Fig. 5) is similar to that for daily data, but spatially the correlation skill is noisier due to



273 FIG. 4. 2003–2016 Grid point rank correlation between daily observed and PR predicted flash anomalies
274 grouped by month. Sample size is the number of days in month times 14 years. Gray grid points indicate where
275 10 or fewer observed lightning flashes occurred over the 14-year period for that particular month.



276 FIG. 5. 2003–2016 Grid point rank correlation between seasonal observed and PR predicted flash anomalies.
277 Sample size is 14 years. Gray grid points indicate where 30 or fewer observed lightning flashes occurred over
278 the 14-year period for that particular season.

the smaller sample size. Interestingly, despite the high rank correlation in daily summer flashes over the Southeast U.S., rank correlation for the summer seasonal totals for the Southeast U.S. is reduced. Overall, skill is low over the West Coast, Northwest U.S., and eastern U.S., especially during the winter (cf. Fig. 5a, b, k, l).

Seasonal and regional details of PR model skill are summarized in Tables 2 and 3. Some of these temporal rank correlation values are higher than would be expected from the spatial rank correlation maps for winter (cf. Fig. 4l vs. Table 2 Dec in top row) and are lower than would be expected from the spatial rank correlation maps for summer (cf. Fig. 4g vs. Table 2 July in top row), demonstrating that spatial aggregation can increase or decrease correlation. There is a statistically significant correlation between daily observed and PR predicted total lightning flash counts east of 120°W (CONUS excluding the West Coast) year-round (Table 2 top row). Daily PR predicted total flashes are highly correlated with daily observed total flashes during winter and spring in particular. There is a statistically significant correlation between seasonal observed and PR predicted total lightning flash counts east of 120°W during October through June, excluding the MAM season (Table 3 top row). In the warm season, correlations between daily observed and PR predicted total flashes are higher than correlations between seasonal observed and PR predicted total flashes (top rows of Table 2 Apr through Oct vs. Table 3 MAM through SON). During the warm season, when there are high day-to-day values and fluctuations in flash counts, temporal aggregation (via a seasonal summation) does not improve PR model performance. Conversely, in DJF and JFM, correlations between seasonal observed and PR predicted total flashes are higher than correlations between daily observed and PR predicted total flashes (top rows of Table 2 Jan and Feb vs. Table 3 DJF and JFM). Unlike during the warm season, temporal and spatial aggregation of flashes during the cool season benefits correlation skill.

Although the model fails to capture CONUS-wide seasonal flash count during some parts of the year, it still performs well for some regions after spatial aggregation. For example, seasonal PR predicted total CONUS flash counts in summer and fall are poorly correlated with observed total CONUS flash counts (Table 3 top row), but predicted total flash counts are well correlated with observed total flash counts over the Northwest and Southwest (Table 3 second and third row). Despite some large regional biases in lightning climatology, the variability is well captured by

TABLE 2. Rank correlations between observed and PR predicted daily total flashes within region, grouped by month. All (bolded) correlations are considered significant at the 1% level.

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
CONUS East of 120°W	**0.8**	**0.88**	**0.88**	**0.89**	**0.85**	**0.72**	**0.66**	**0.67**	**0.67**	**0.70**	**0.77**	**0.86**
Northwest [41-50°N, 102-125°W]	**0.59**	**0.63**	**0.69**	**0.60**	**0.77**	**0.80**	**0.86**	**0.85**	**0.72**	**0.60**	**0.69**	**0.69**
Southwest [30-41°N, 102-125°W]	**0.70**	**0.78**	**0.71**	**0.78**	**0.85**	**0.82**	**0.66**	**0.66**	**0.80**	**0.86**	**0.78**	**0.72**
Tennessee River Valley [33-42°N, 84-93°W]	**0.71**	**0.73**	**0.82**	**0.92**	**0.89**	**0.84**	**0.85**	**0.86**	**0.81**	**0.80**	**0.79**	**0.77**
Northeast [37-50°N, 65-90°W]	**0.63**	**0.57**	**0.72**	**0.84**	**0.86**	**0.86**	**0.88**	**0.85**	**0.80**	**0.70**	**0.61**	**0.64**
Southeast [24-37°N, 76-85°W]	**0.74**	**0.82**	**0.87**	**0.91**	**0.88**	**0.64**	**0.68**	**0.66**	**0.75**	**0.85**	**0.78**	**0.79**

TABLE 3. Rank correlations between observed and PR predicted seasonal total flashes within region. Bolded correlations are considered significant at the 10% level.

	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
CONUS East of 120°W	**0.91**	**0.95**	**0.82**	0.32	**0.51**	**0.64**	0.41	0.4	0.33	0.35	**0.76**	**0.81**
Northwest [41-50°N, 102-125°W]	**0.61**	0.0	0.03	0.0	0.42	**0.48**	**0.64**	**0.82**	**0.71**	**0.74**	0.31	**0.77**
Southwest [30-41°N, 102-125°W]	**0.54**	**0.68**	**0.87**	**0.84**	**0.94**	**0.5**	**0.63**	**0.51**	**0.56**	**0.54**	**0.89**	0.38
Tennessee River Valley [33-42°N, 84-93°W]	**0.85**	**0.92**	**0.64**	**0.44**	**0.58**	**0.77**	**0.58**	0.27	0.31	**0.44**	**0.52**	**0.75**
Northeast [37-50°N, 65-90°W]	**0.62**	**0.82**	**0.81**	**0.60**	0.42	0.13	0.04	0.16	0.30	**0.67**	**0.70**	**0.64**
Southeast [24-37°N, 76-85°W]	**0.59**	**0.78**	**0.63**	**0.46**	0.09	0.12	0.19	0.35	0.07	0.32	**0.66**	**0.73**

the model. This is especially true when using rank correlation, which is insensitive to systematic amplitude errors.

In brief, we determined that the PR model can generally reproduce many of the observed aspects of the climatology and variability of lightning. The MJO analysis during JJAS with daily data in the next subsection is appropriate given the PR model performance is good over these months and over all regions on this timescale. Similarly, ENSO analysis during DJF through MJJ is generally appropriate given PR model performance during these seasons. There are noted discrepancies, e.g., poor performance over Northwest during winter through midspring, that should be kept in mind.

b. MJO-Lightning Relationship

Motivated by previous work that found statistically significant relations between active MJO phases and CONUS lightning activity during JJAS (Abatzoglou and Brown 2009), we computed

maps of anomalous flash counts during active MJO phases for the period 2003–2016 (Supplementary Fig. S1). We found essentially no statistically significant flash count anomalies in any of the MJO phases. Although the period is relatively short, the number of active phase days in the analysis is not excessively small; phase 7 has the fewest active days (63), and phase 1 has the most (237). We conclude that any MJO signal in flash counts is too modest to be robustly detected in this sample.

We repeated the MJO composite analysis using the PR predicted values over the same 2003–2016 period and again found essentially no statistically significant anomalies in any of the MJO phases (Supplementary Fig. S2). Nevertheless, PR predicted flash anomaly composites show that the PR model is at least broadly capturing similar patterns to the observations on the days that go into the MJO composites. The average (over MJO phases) pattern correlation between the observed and PR predicted flash anomaly composites is 0.65.

The relatively good relation between flash count and PR composites in addition to the strong rank correlation of PR with daily flash counts during this time of year (Fig. 4) substantiates the strategy of using the PR model to examine the MJO-lightning relationship in the longer period 1979–2021 during which reanalysis data are available. Using a long period permits the detection of MJO signals that were too small to be detected in the 14-year lightning data set.

Despite the longer period, we also find for the 1979–2021 period that there are essentially no statistically significant PR flash anomalies in any of the MJO phases (Fig. 6), and anomalies are relatively small (at most, 0.1-0.2 standard deviations away from climatology in some regions). While some of the 1979–2021 PR predicted patterns are consistent with the 2003–2016 observed patterns (e.g., MJO phase 6 composites), the lack of statistical significance from the permutation test and FDR correction suggests these similarities might be due to chance rather than MJO variability.

These MJO-lightning results differ from Abatzoglou and Brown (2009), which found statistically significant relations between MJO phase and JJAS lightning counts. Like the study here, they used NLDN data at daily resolution aggregated to a $1^\circ \times 1^\circ$ grid. Some data and methodological differences might explain the differing results. For example, Abatzoglou and Brown (2009) used JJAS lightning data for the period 1990–2007 which differs from the 2003–2016 period used here. However, our PR modeled data analysis over longer period 1979–2021 also found no statistically significant associations. In addition, they assessed statistical significance using a Wilcoxon rank

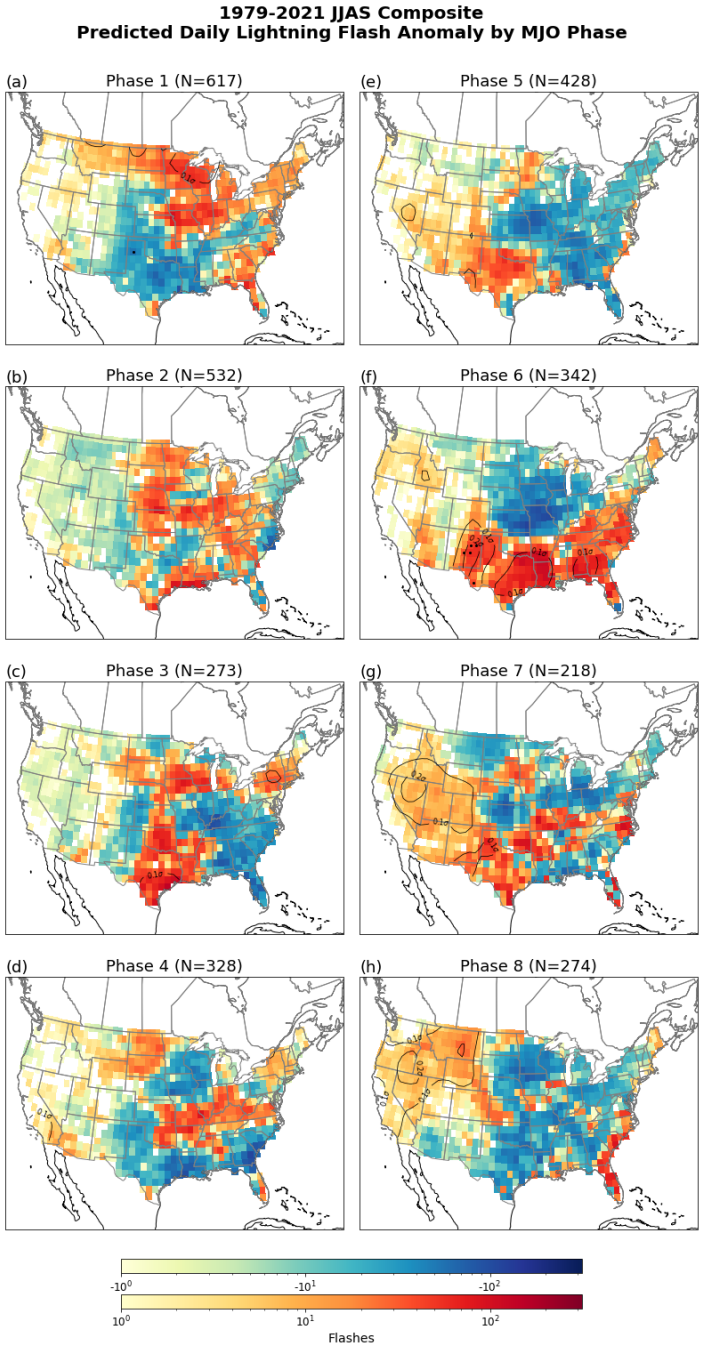


FIG. 6. 1979–2021 JJAS composites of PR predicted flash count anomalies by (active) MJO phase. Anomalies are with respect to a daily climatology. Black contours indicate anomalies that are 0.1 and 0.2 standard deviations away from climatology. Black stippling indicates statistically significant values.

sum test and did not account for multiple testing. We used a permutation test and a FDR control procedure. Accounting for multiple testing is important because when a statistical significance test is applied at many grid points with, for instance, $\alpha = 0.05$, then the null hypothesis is expected to be rejected 5% of the time when it is true (type I error), and the presence of spatial correlation leads to an even higher type I error rate (Wilks 2016). The FDR procedure used here is designed to avoid such type I errors. Moreover, its behavior here appears to be reasonable (i.e., not overly strict) since before the FDR procedure, on average, only 8% of the land points pass the statistical significance test at the nominal 5% level.

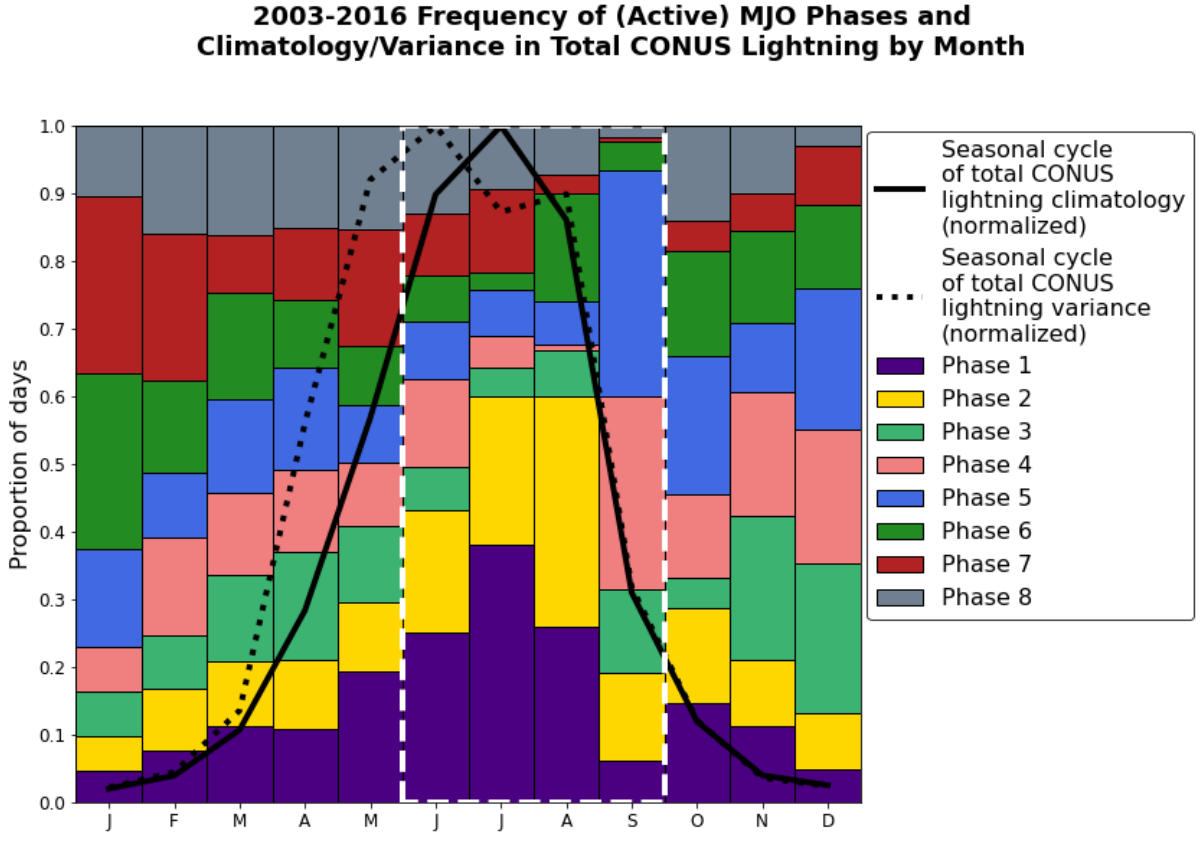


FIG. 7. 2003–2016 Proportion of days in each (active) MJO phase, with overlaid seasonal cycle of total CONUS lightning climatology (solid black) and variance (dotted black). White dashed box denotes season of interest for MJO-lightning analysis.

Of the possible explanations for the differing results, the application of the Wilcoxon rank sum test deserves attention. The Wilcoxon rank sum test addresses the question of whether two sets of samples come from the same distribution. In particular, the Wilcoxon rank sum test is designed to detect *stochastic dominance*. A random variable X has stochastic dominance over a random variable Y if $P(X > z) \geq P(Y > z)$ for all thresholds z and is strictly greater for some thresholds. In the application here, the Wilcoxon rank sum test decides whether the observed lightning flash counts are higher for a particular MJO phase compared to the other phases. This comparison assumes that all other factors are held fixed, and, when that is not the case, a confounding factor can produce an apparent relation in variables that are otherwise unrelated (DelSole and Tippett 2022). The annual cycle is the most common confounding factor in climate studies, and not accounting for it leads to detecting links in quantities whose only link is that they share annual cycle phasing. Fig. 7 shows that CONUS lightning count mean (solid black line) and variance (dashed black line) have strong seasonality, with highest values in summer. MJO phase frequency also presents seasonality (Fig. 7 color bars), especially in the 14-year period considered here. It is possible, for example, that the mostly negative JJAS flash anomalies in MJO phase 5 (Fig. 6e) might be due to phase 5 being most frequent in September when lightning activity is relatively low. We assessed the impact of failing to account for seasonality by applying the Wilcoxon rank sum test to data whose seasonality exactly matched that of the observed flash count data but which had no relation with MJO phase. We constructed such a dataset by randomly permuting the years of the summed CONUS lightning flash count data while maintaining day of the year. We then applied the Wilcoxon rank sum test with a significance level of $\alpha = 0.05$. In this situation the null hypothesis of no relation is true, and p-values < 0.05 constitute type I errors. We computed type I error rates for each MJO phase and repeated this procedure with 1000 random permutations of the years. Table 4 (first row) shows the type I error rates are far from their desired value of 5%. We note that removing the seasonal mean leaves the ranks unchanged and likewise the results of the rank sum test are the same. One strategy to account for seasonality is to remove the daily flash count climatology. Table 4 (second row) shows that removing the daily flash count climatology reduces the type I error rates but they remain far from their desired value of 5%. Seasonality of the lightning flash count variance as well as autocorrelation are possible reasons why removal of the daily climatology was inadequate. On the other hand, the permutation test which accounts

for seasonality and autocorrelation has type I error rates that are close to 5% (Table 4 third row). Hence, we conclude that the difference between previous results and ours regarding MJO-lightning links may be due at least in part to differences in statistical significance testing methodology.

TABLE 4. Type I Error Rate at the 0.05 significance level, which is Freq(p-value < 0.05), using the Wilcoxon rank sum test for the observed 2003–2016 MJO composites and randomized lightning flash data (top row) with seasonal mean removed, (middle row) with daily mean removed, and (bottom row) using the permutation test.

Phase	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Rank sum test, removed seasonal climatology	0.98	0.34	0.82	0.99	1.0	0.21	0.69	0.59
Rank sum test, removed daily climatology	0.18	0.24	0.20	0.16	0.10	0.25	0.26	0.26
Permutation test	0.05	0.05	0.04	0.04	0.06	0.04	0.06	0.05

c. ENSO-Lightning Relationship

We first examined ENSO influence on observed and PR predicted lightning activity for the 2003–2016 period during the six 3-month overlapping seasons DJF–MJJ. Because the period is short and contains few ENSO events (4 complete El Niño events and 5 complete La Niña events), statistical significance is not expected. Maps of rank correlation between ONI and observed lightning flash anomaly (Supplementary Figs. S3a-f) and maps of El Niño - La Niña difference composites (Supplementary Figs. S3g–l) show no statistical significance. Likewise, the PR predicted data show no statistically significant evidence for an ENSO signal during the period 2003–2016 in either rank correlations or difference composites (Supplementary Figs. S4a–f and S4g–l, respectively). Rank correlations between ONI and regionally summed lightning flash count anomalies for DJF–MJJ (Fig. S5) show: mostly positive correlations for the Northwest and statistical significance for FMA–MJJ, positive correlations for the Southwest and and statistical significance for DJF–MJJ, negative correlations for the Tennessee River Valley and statistical significance in FMA, and mostly negative values for the Coastal Southeast but no statistical significance.

The similarities between the observed and PR results for ENSO rank correlations (average pattern correlation 0.59) and for ENSO composites (average pattern correlation 0.64) further supports using the PR model to examine the ENSO-lightning relation in the longer period 1979–2021. In addition, rank correlations between ONI and the regionally summed lightning flash anomalies are generally consistent for all regions.

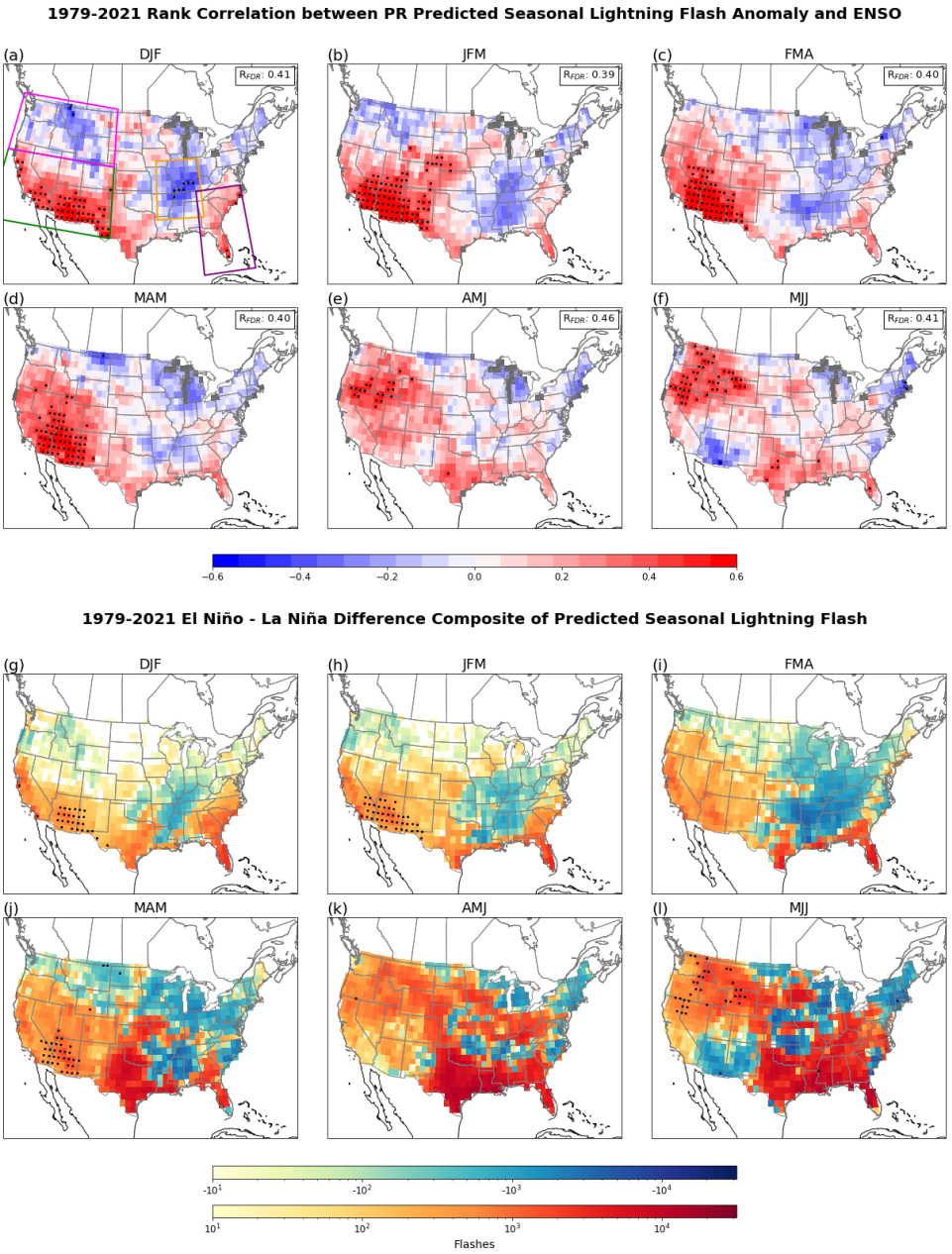
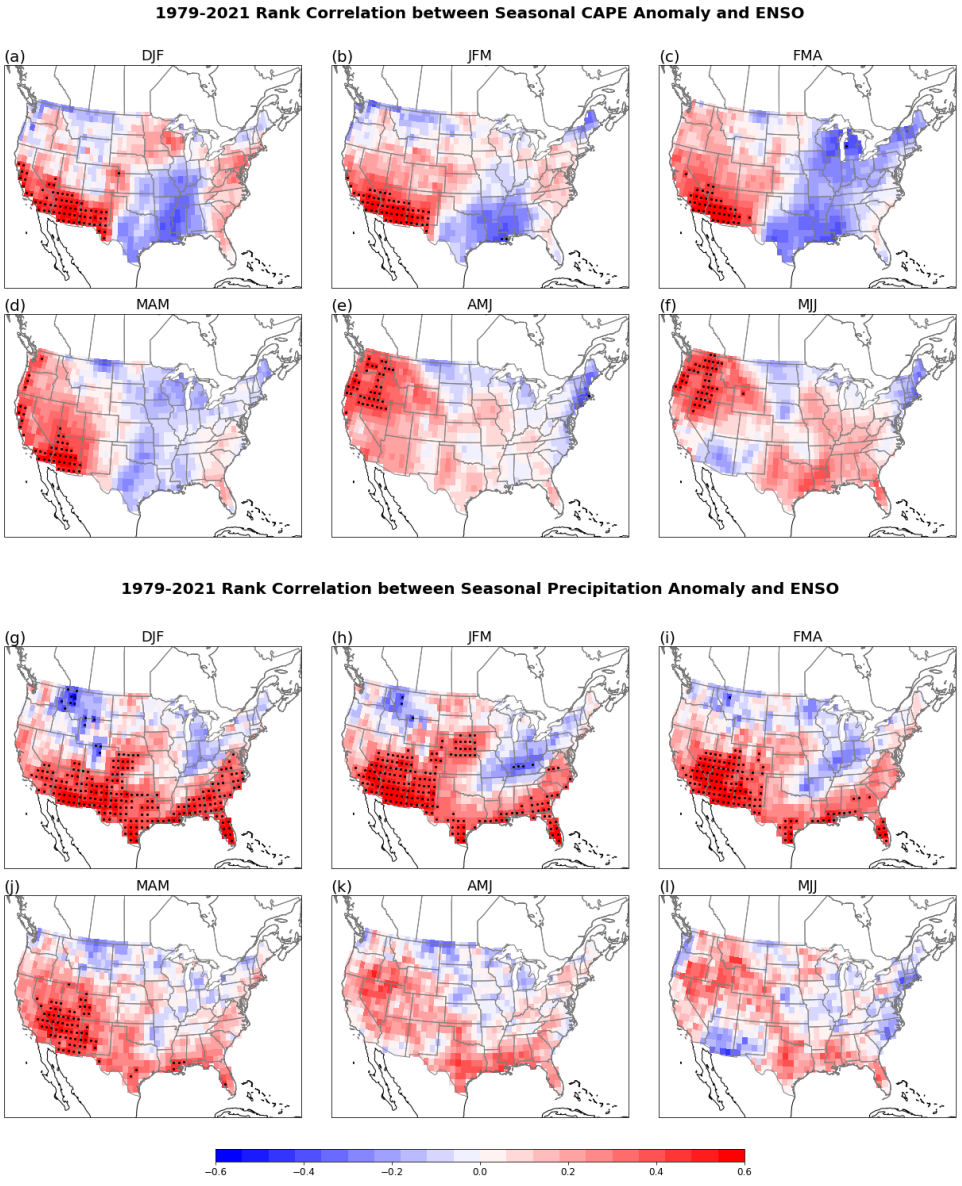


FIG. 8. (a-f) 1979–2021 Rank correlation between PR predicted seasonal lightning flash count anomalies and ONI. All insignificant correlations have amplitudes less than R_{FDR}, which is shown in top-right panel. (g–l) 1979–2021 difference composites (El Niño - La Niña) of PR predicted seasonal lightning flash count anomalies. Black stippling indicates statistically significant values. Domain boxes in panel (a) indicate regions of interest (see text for details).



439 FIG. 9. 1979–2021 Rank correlation between ONI and (a–f) seasonal CAPE anomaly and (g–l) seasonal
440 precipitation anomaly. Black stippling indicates statistically significant values.

441 We use rank correlation maps (Figs. 8a–f) and El Niño - La Niña difference composites (Fig.
442 8g–l) to examine the relation between ENSO and PR predicted lightning activity during winter
443 and spring over the longer period 1979–2021. The two measures are complementary, with rank
444 correlation showing the strength of the relationship and the El Niño - La Niña difference composites

showing their expected magnitude. Four regions stand out as having large-scale ENSO signals during some periods in winter and spring: Northwest, Southwest, Tennessee River Valley, and Coastal Southeast. The domains for these regions are displayed in Fig. 8a and described in Tables 2 and 3. The rank correlation maps indicate more regions with statistically significant links between ENSO and PR predicted lightning, likely because correlation utilizes the full dataset, as well as ONI amplitude, and is less sensitive to outliers than the composites. During the midwinter months, El Niño has a modest relationship with increased PR lightning activity over the Coastal Southeast (Fig. 8a), with expected ENSO-modulated PR lightning activity on the order of 10^2 to 10^3 flashes (~30-40% more than normal; Fig. 8g). During the winter through early spring months, El Niño is associated with increased PR lightning activity over the Southwest, and La Niña is associated with increased PR lightning activity over the Tennessee River Valley (Fig. 8a, b), though the signal over the Tennessee River Valley is statistically significant only for the DJF season. Because this season has fewer flashes climatologically (Tippett et al. 2019), the expected difference in ENSO-modulated PR lightning activity is on the order of 10^2 flashes (~30% more than normal; Fig. 8g, h), though locally differences are on the order of 10^3 flashes. During spring, El Niño is associated with increased PR lightning activity over the Southwest (Fig. 8c, d), and the difference in ENSO-modulated PR lightning activity is on the order of 10^2 in FMA (~40-50% more than normal; Fig. 8i) and 10^3 flashes in MAM (~50-60% more than normal; Fig. 8j). El Niño is associated increased PR lightning activity over the Northwest in the late spring (Fig. 8e, f). However, the difference composites do not indicate a clear positive signal related to El Niño (Fig. 8k, l), suggesting there might be asymmetries between the El Niño- and La Niña-modulated patterns of PR lightning activity in this region.

We explain ENSO influence on PR predicted lightning activity in these regions and seasons by examining the relationship between ENSO and the large-scale environmental variables in the PR model, CAPE and precipitation. The ENSO-precipitation relationship helps explain ENSO modulation during winter over the Coastal Southeast during midwinter (cf. Fig. 9g vs. Fig. 8a). The ENSO-CAPE relationship helps best explains ENSO modulation of lightning activity over the Tennessee River Valley, including its peak in winter through early spring (cf. Fig. 9a-c vs. Fig. 8a-c). Both CAPE and precipitation can be useful explanatory variables to describe ENSO

modulation patterns and seasonality over the Southwest (cf. Figs. 9a-d and 9g-j vs. Fig. 8a-d) and Northwest (cf. Figs. 9e, f and 9k, l vs. Fig. 8e, f).

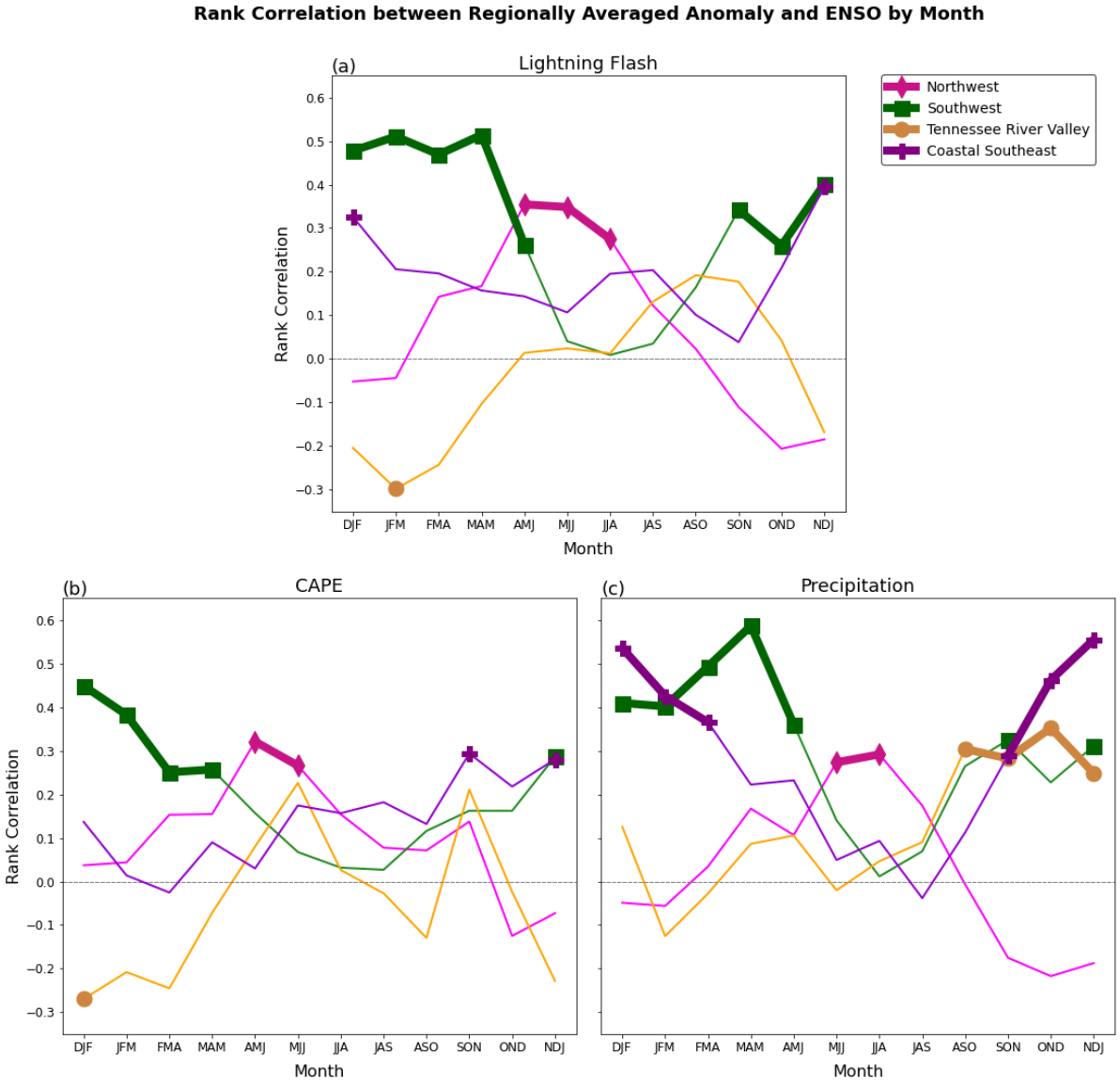


FIG. 10. 1979–2021 Rank correlation between ONI and regionally summed (a) PR predicted lightning flash anomaly, (b) CAPE anomaly, and (c) precipitation anomaly. Markers and any thicker lines between points indicate seasons when rank correlation is statistically significant at the 90% confidence level determined by bootstrapping method with 1000 iterations.

As a summary, we took each 3-month season of (PR predicted) regionally summed lightning flash, CAPE, or precipitation anomaly for the Northwest, Southwest, Tennessee River Valley, or Coastal Southeast and correlated it with ENSO (Fig. 10), which describes the major takeaways for the 1979–2021 ENSO-lightning analysis:

- Over the Northwest, there is a positive correlation between ENSO and lightning activity - hence El Niño (La Niña) is associated with increased (decreased) lightning activity - that peaks in late spring and early summer (Fig. 10a magenta line). This is explained by El Niño-related (La Niña-related) increases (decreases) in both CAPE and precipitation over the region (Fig. 10b, c magenta lines).

- Over the Southwest, there is a positive correlation between ENSO and lightning activity - hence El Niño (La Niña) is associated with increased (decreased) lightning activity - that peaks in the winter and spring (Fig. 10a green line). This is best explained by El Niño-related (La Niña-related) increases (decreases) in both CAPE and precipitation over the region (Fig. 10b, c green lines).

- Over the Tennessee River Valley, there is a negative correlation between ENSO and lightning activity - hence La Niña (El Niño) is associated with increased (decreased) lightning activity - that peaks during winter (Fig. 10a gold line). This is best explained by La Niña-related (El Niño-related) increases (decreases) in CAPE over the region (Fig. 10b gold line).

- Over the Coastal Southeast, there is a positive correlation between ENSO and lightning activity - hence El Niño (La Niña) is associated with increased (decreased) lightning activity - that peaks in winter (Fig. 10a purple line). This is best explained by El Niño-related (La Niña-related) increases (decreases) in precipitation over the region (Fig. 10c purple line).

The weak ENSO signal in PR lightning activity over the Tennessee River Valley and Coastal Southeast can perhaps be attributed to the contrasting ENSO signals in CAPE and precipitation. For instance, when the ENSO-PR lightning link over the Tennessee River Valley is highest in late winter, the ENSO-CAPE correlation is negative and the ENSO-precipitation correlation is positive over the region.

4. Summary and Discussion

Despite its substantial impacts, relatively little is known about the relation of lightning with predictable large-scale climate variability. In this study, we modeled 2003–2016 lightning activity over the U.S. using Poisson regression (PR), relating lightning flash counts data to CAPE and precipitation values. The PR model more accurately captured the dependence of flash counts on the environmental factors compared to previous models. In particular, the PR model with the best performance used log(CAPE), log(CAPE)^2, log(Precipitation), and log(Precipitation)^2 terms as predictors. Correlations between the observed and PR predicted lightning flash counts at daily and seasonal scales show that the PR model overall represents daily and seasonal variability well. We did not detect a robust MJO-lightning or ENSO-lightning relationship for the short 2003–2016 period, but we found observed and PR predicted patterns were similar. Therefore, with the PR model, we constructed a lightning proxy for longer period 1979–2021. For the 1979–2021 period, we found no evidence that the lightning activity is related to active MJO phases during JJAS. However, ENSO was associated with anomalous PR lightning flash patterns. La Niña is associated with increased lightning activity over the Tennessee River Valley during winter, and El Niño is associated with increased lightning activity over the Coastal Southeast during winter, the Southwest during winter and spring, and the Northwest during late spring and early summer. In terms of the magnitude of lightning flash anomalies during ENSO phases, most differences are on the order of 10^2 more (or fewer) flashes per season except for some localized differences on the order of 10^3 more flashes, e.g., Montana and Wyoming in early summer. The PR lightning proxy also provides insight into how ENSO influences CAPE and precipitation patterns, including the seasonal evolution of the ENSO-CAPE and ENSO-precipitation link. As a result, ENSO may be a potentially useful predictor for winter through late spring lightning activity, valuable for protecting public safety as well as for understanding increased risk of lightning-caused wildfires and thunderstorm-related impacts.

The PR approach has its limitations. For example, western U.S, especially West Coast, lightning climatology was not well represented. It has been found that lightning along the West Coast involves different mechanisms/processes than lightning over the rest of the contiguous U.S. (Zipser 1994; Holle et al. 2016). The West Coast has a high number of positive CG flashes that are not associated with severe weather (Zajac and Rutledge 2001; Koshak et al. 2015; Medici et al. 2017)

but might be associated with greater ice clouds (Fuchs et al. 2015; Mülmenstädt et al. 2015). The PR model does not take into account other important environmental factors, e.g., wet-bulb temperature (Koshak et al. 2015), warm cloud depth (Stolz et al. 2017), mid-level humidity (Westermayer et al. 2017), cloud top height (Finney et al. 2018), etc., which may limit its representation of lightning across many climates and terrains. Noting the western U.S. deficiencies, fitting the data separately shows a systematically different base rate and dependence on CAPE and precipitation that might be due to differing precipitation/lightning processes or missing physics. We did find that we could improve lightning climatology and skill by adding lifted condensation level (LCL) as a predictor in the PR model (not shown), reducing the biases over the western U.S. by accounting for the lower cloud bases, warm rain processes, and/or greater precipitation efficiency in the western U.S. (Mülmenstädt et al. 2015; Fuchs et al. 2015; Tippett et al. 2019). Regardless, the PR model with the LCL predictor did not impact nor improve representation of lightning variability and, by extension, the MJO- or ENSO-lightning relationship, motivating the choice to exclude the LCL in the model for the purpose of this study. While PR predicted lightning flash counts of the PR model during 2003–2016 were comparable to observations, it is not guaranteed that the PR model would perform well outside this data-fitting period, especially if there were long-term trends involved (Stainforth et al. 2007; Camargo et al. 2014). Overall, MJO-lightning and ENSO-lightning results were qualitatively similar when using the CP proxy predicted lightning data (not shown).

Because of the high variability of lightning and its strong seasonality, the PR model performs better at daily resolution in the warm season and at seasonal resolution in the cool season (cf. Tables 2 and 3). During seasons and in regions with high lightning variability, temporal and/or spatial aggregation does not improve PR model performance. This dependence of skill on timescale may reflect that during climatologically active times of the year, aggregation over a season requires accurate representation over many days, whereas during climatologically inactive times of the year, aggregation over a season requires accurate representation over fewer days. Likewise, the dependence of skill on region may reflect that over climatologically active locations, spatial aggregation requires accurate representation over many grid points. The timescale of climate drivers as well as season or region of interest should be considered when using the lightning proxy. For example, the lightning proxy may performs better for shorter (i.e., subseasonal) climate timescales, especially in the warm season. However, overall, ENSO and MJO spatial patterns of PR predicted lightning

look similar to the ENSO and MJO spatial patterns of observed lightning during 2003–2016 (cf. Supplementary Figs. 1–4), suggesting that the PR model can capture large-scale, low-frequency variability in lightning activity and is still useful in investigating climate links and patterns.

The lack of statistically significant MJO-lightning relations in our study is different from the MJO-lightning patterns found in Abatzoglou and Brown (2009). Methodological differences may play a role. Here, we demonstrated that failing to account for seasonality in MJO phase frequency and lightning climatology resulted in type I errors (rejecting the null hypothesis when it is true) in as many as 100% of cases. This behavior may provide an explanation for why the results here differ from previous ones which did not take seasonality into consideration. The lack of a clear relation between the MJO and lightning has implications for subseasonal prediction of lightning activity and its indirect impacts (e.g., wildfire ignition).

Over the eastern U.S., the link between ENSO and lightning might be expected to be similar to that between ENSO and severe convective storm activity. For instance, there is a documented relationship between La Niña and increased frequency of tornado and hail events over the Tennessee River Valley in late winter and early spring (Allen et al. 2015a; Koch et al. 2021; Tippett and Lepore 2021), which can be associated with ENSO-related CAPE patterns. However, the ENSO signals in lightning are generally weaker than the ENSO signals in severe thunderstorms. We hypothesize that the contrasting ENSO signals in CAPE and precipitation may be reducing the ENSO signal in lightning, especially over the Tennessee River Valley. On the other hand, for severe thunderstorms, additional environmental factors become relevant, such as storm relative helicity (SRH). ENSO influences patterns of SRH and CAPE similarly i.e., La Niña increases both CAPE and SRH in the region (Lepore et al. 2018; Koch et al. 2021), and the ENSO signal in SRH is stronger than (convective) precipitation in climate forecast models, which can overpower any opposing precipitation signals in the region.

The NLDN is one of the most reliable datasets of lightning observations available, but future work should consider, when available, the re-classified NLDN data, which estimates 2015–2022 lightning counts with improved detection methods (personal communications, Koshak 2023). Future work should also consider how this PR model could be used to understand how other climate modes of variability may influence lightning activity as well as for identifying trends. Global lightning trends are not well understood, and there is general disagreement on whether lightning activity

will increase or decrease with a warming climate (Price and Rind 1994; Finney et al. 2018; Romps 2019). Some studies have reported that U.S. CG lightning has increased (Romps et al. 2014; Romps 2019), including over the Great Plains where severe thunderstorms are frequent (Villarini and Smith 2013). However, these studies use datasets with short records, so there would be benefit from a comprehensive analysis over a longer period of time and over the entire U.S.

Acknowledgments. The authors gratefully acknowledge Vaisala Inc. for providing the NLDN data used in this study as part of the Marshall Mentored Project (MMP, ID #02). The work by co-author Koshak was supported by the Precipitation and Lightning Work Package for the Internal Science Funding Model (ISFM) project Lightning as an Indicator of Climate under NASA Headquarters (Dr. Jack Kaye and Dr. Lucia Tsaoussi), which supports NASA's participation in the National Climate Assessment (NCA). The authors thank Paul E. Roundy and two anonymous reviewers for their constructive feedback, which improved the paper.

Data availability statement. Proprietary NLDN data were provided by Vaisala Inc. and is available for purchase at `https://www.vaisala.com/en/digital-and-data-services/lightning`. NARR data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA from their website `https://psl.noaa.gov/data/gridded/data.narr.html`. MJO RMM data are provided by the Centre for Australian Weather and Climate Research from their website `http://www.bom.gov.au/climate/mjo/`. ONI data are provided by NOAA/CPC at `https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php`.

References

Abatzoglou, J. T., and T. J. Brown, 2009: Influence of the Madden–Julian oscillation on summertime cloud-to-ground lightning activity over the continental United States. *Monthly weather review*, **137 (10)**, 3596–3601.

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015a: An empirical model relating US monthly hail occurrence to large-scale meteorological environment. *Journal of Advances in Modeling Earth Systems*, **7 (1)**, 226–243.

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015b: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geoscience*, **8 (4)**, 278–283.

Baggett, C. F., K. M. Nardi, S. J. Childs, S. N. Zito, E. A. Barnes, and E. D. Maloney, 2018: Skillful subseasonal forecasts of weekly tornado and hail activity using the Madden-Julian Oscillation. *Journal of Geophysical Research: Atmospheres*, **123 (22)**, 12–661.

Becker, E., H. v. den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *Journal of Climate*, **27 (15)**, 5891–5906.

Becker, E. J., E. H. Berbery, and R. W. Higgins, 2011: Modulation of cold-season US daily precipitation by the Madden-Julian oscillation. *Journal of Climate*, **24 (19)**, 5157–5166.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, **57**, 289–300.

Biagi, C. J., K. L. Cummins, K. E. Kehoe, and E. P. Krider, 2007: National lightning detection network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003–2004. *Journal of Geophysical Research: Atmospheres*, **112 (D5)**.

Branick, M. L., and C. A. Doswell III, 1992: An observation of the relationship between supercell structure and lightning ground-strike polarity. *Weather and Forecasting*, **7 (1)**, 143–149.

Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *Journal of Hydrometeorology*, **8 (4)**, 837–846.

Camargo, S. J., M. K. Tippett, A. H. Sobel, G. A. Vecchi, and M. Zhao, 2014: Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model. *Journal of Climate*, **27 (24)**, 9171–9196.

Cui, W., X. Dong, B. Xi, and A. Kennedy, 2017: Evaluation of reanalyzed precipitation variability and trends using the gridded gauge-based analysis over the CONUS. *Journal of Hydrometeorology*, **18 (8)**, 2227–2248.

Cummins, K. L., and M. J. Murphy, 2009: An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the US NLDN. *IEEE transactions on electromagnetic compatibility*, **51 (3)**, 499–518.

DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of week-3–4 average temperature and precipitation over the contiguous United States. *Journal of Climate*, **30 (10)**, 3499–3512.

DelSole, T. M., and M. K. Tippett, 2022: *Statistical Methods for Climate Scientists*. Cambridge University Press, https://doi.org/10.1017/9781108659055.

Dewan, A., E. T. Ongee, M. Rafiuddin, M. M. Rahman, and R. Mahmood, 2018: Lightning activity associated with precipitation and CAPE over Bangladesh. *International Journal of Climatology*, **38 (4)**, 1649–1660.

Dowdy, A. J., 2016: Seasonal forecasting of lightning and thunderstorm activity in tropical and temperate regions of the world. *Scientific reports*, **6 (1)**, 1–10.

Fierro, A. O., E. R. Mansell, D. R. MacGorman, and C. L. Ziegler, 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 (7)**, 2390–2415.

Finney, D. L., R. M. Doherty, O. Wild, D. S. Stevenson, I. A. MacKenzie, and A. M. Blyth, 2018: A projected decrease in lightning under climate change. *Nature Climate Change*, **8 (3)**, 210–213.

Fuchs, B. R., and Coauthors, 2015: Environmental controls on storm intensity and charge structure in multiple regions of the continental United States. *Journal of Geophysical Research: Atmospheres*, **120 (13)**, 6575–6596.

Gensini, V. A., T. L. Mote, and H. E. Brooks, 2014: Severe-thunderstorm reanalysis environments and collocated radiosonde observations. *Journal of Applied Meteorology and Climatology*, **53 (3)**, 742–751.

Goodman, S. J., and D. R. MacGorman, 1986: Cloud-to-ground lightning activity in mesoscale convective complexes. *Monthly Weather Review*, **114 (12)**, 2320–2328.

Holle, R. L., K. L. Cummins, and W. A. Brooks, 2016: Seasonal, monthly, and weekly distributions of NLDN and GLD360 cloud-to-ground lightning. *Monthly Weather Review*, **144 (8)**, 2855–2870.

Jung, E., and B. P. Kirtman, 2016: Can we predict seasonal changes in high impact weather in the United States? *Environmental Research Letters*, **11 (7)**, 074 018.

Kang, D., R. Mathur, G. A. Pouliot, R. C. Gilliam, and D. C. Wong, 2020: Significant ground-level ozone attributed to lightning-induced nitrogen oxides during summertime over the mountain west states. *NPJ climate and atmospheric science*, **3 (1)**, 1–7.

Kim, D., S.-K. Lee, and H. Lopez, 2020: Madden–Julian oscillation–induced suppression of northeast Pacific convection increases US tornadogenesis. *Journal of Climate*, **33 (11)**, 4927–4939.

King, A. T., and A. D. Kennedy, 2019: North American supercell environments in atmospheric reanalyses and RUC-2. *Journal of Applied Meteorology and Climatology*, **58 (1)**, 71–92.

Koch, E., J. Koh, A. C. Davison, C. Lepore, and M. K. Tippett, 2021: Trends in the extremes of environments associated with severe US thunderstorms. *Journal of Climate*, **34 (4)**, 1259–1272.

Koehler, T. L., 2020: Cloud-to-ground lightning flash density and thunderstorm day distributions over the contiguous United States derived from NLDN measurements: 1993–2018. *Monthly Weather Review*, **148 (1)**, 313–332.

Koshak, W. J., K. L. Cummins, D. E. Buechler, B. Vant-Hull, R. J. Blakeslee, E. R. Williams, and H. S. Peterson, 2015: Variability of CONUS lightning in 2003–12 and associated impacts. *Journal of Applied Meteorology and Climatology*, **54 (1)**, 15–41.

Laing, A., M. LaJoie, S. Reader, and K. Pfeiffer, 2008: The influence of the El Niño–Southern Oscillation on cloud-to-ground lightning activity along the Gulf Coast. Part II: Monthly correlations. *Monthly Weather Review*, **136 (7)**, 2544–2556.

LaJoie, M., and A. Laing, 2008: The influence of the El Niño–Southern Oscillation on cloud-to-ground lightning activity along the Gulf Coast. Part I: Lightning climatology. *Monthly Weather Review*, **136 (7)**, 2523–2542.

Lepore, C., M. K. Tippett, and J. T. Allen, 2018: CFSv2 monthly forecasts of tornado and hail activity. *Weather and Forecasting*, **33 (5)**, 1283–1297.

Lynn, B. H., Y. Yair, C. Price, G. Kelman, and A. J. Clark, 2012: Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Weather and Forecasting*, **27 (6)**, 1470–1488.

L'Heureux, M. L., M. K. Tippett, and A. G. Barnston, 2015: Characterizing ENSO coupled variability and its impact on North American seasonal precipitation and temperature. *Journal of Climate*, **28 (10)**, 4231–4245.

MacGorman, D. R., and D. W. Burgess, 1994: Positive cloud-to-ground lightning in tornadic storms and hailstorms. *Monthly Weather Review*, **122 (8)**, 1671–1697.

Mansell, E. R., D. R. MacGorman, C. L. Ziegler, and J. M. Straka, 2005: Charge structure and lightning sensitivity in a simulated multicell thunderstorm. *Journal of Geophysical Research: Atmospheres*, **110 (D12)**.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting*, **24 (3)**, 709–729.

Medici, G., K. L. Cummins, D. J. Cecil, W. J. Koshak, and S. D. Rudlosky, 2017: The intracloud lightning fraction in the contiguous United States. *Monthly Weather Review*, **145 (11)**, 4481–4499.

Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87 (3)**, 343–360.

Michalon, N., A. Nassif, T. Saouri, J. Royer, and C. Pontikis, 1999: Contribution to the climatological study of lightning. *Geophysical research letters*, **26 (20)**, 3097–3100.

Moore, T. W., and M. P. McGuire, 2020: Tornado-days in the United States by phase of the Madden–Julian oscillation and global wind oscillation. *Climate Dynamics*, **54 (1-2)**, 17–36.

Mülmenstädt, J., O. Sourdeval, J. Delanoë, and J. Quaas, 2015: Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals. *Geophysical Research Letters*, **42 (15)**, 6502–6509.

Muñoz, Á., J. Díaz-Lobatón, X. Chourio, and M. Stock, 2016: Seasonal prediction of lightning activity in north western Venezuela: Large-scale versus local drivers. *Atmospheric Research*, **172**, 147–162.

Murugavel, P., S. Pawar, and V. Gopalakrishnan, 2012: Trends of convective available potential energy over the Indian region and its effect on rainfall. *International Journal of Climatology*, **32 (9)**, 1362–1372.

Nag, A., M. J. Murphy, and J. A. Cramer, 2016: Update to the US national lightning detection network. *24th International Lightning Detection Conference and 6th International Lightning Meteorology Conference*.

Nigam, S., and A. Sengupta, 2021: The full extent of El Niño's precipitation influence on the United States and the Americas: The suboptimality of the Niño 3.4 SST index. *Geophysical Research Letters*, **48 (3)**, e2020GL091 447.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, **100 (10)**, 2043–2060.

Petersen, W. A., and S. A. Rutledge, 1998: On the relationship between cloud-to-ground lightning and convective rainfall. *Journal of Geophysical Research: Atmospheres*, **103 (D12)**, 14 025–14 040.

Price, C., and D. Rind, 1994: Possible implications of global climate change on global lightning distributions and frequencies. *Journal of Geophysical Research: Atmospheres*, **99 (D5)**, 10 823–10 831.

Romps, D. M., 2019: Evaluating the future of lightning in cloud-resolving models. *Geophysical Research Letters*, **46 (24)**, 14 863–14 871.

Romps, D. M., A. B. Charn, R. H. Holzworth, W. E. Lawrence, J. Molinari, and D. Vollaro, 2018: CAPE times P explains lightning over land but not the land-ocean contrast. *Geophysical Research Letters*, **45 (22)**, 12–623.

Romps, D. M., J. T. Seeley, D. Vollaro, and J. Molinari, 2014: Projected increase in lightning strikes in the United States due to global warming. *Science*, **346 (6211)**, 851–854.

Saha, S., and Coauthors, 2014: The NCEP climate forecast system version 2. *Journal of climate*, **27 (6)**, 2185–2208.

Stainforth, D. A., M. R. Allen, E. R. Tredger, and L. A. Smith, 2007: Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365 (1857)**, 2145–2161.

Stolz, D. C., S. A. Rutledge, J. R. Pierce, and S. C. van den Heever, 2017: A global lightning parameterization based on statistical relationships among environmental factors, aerosols, and convective clouds in the TRMM climatology. *Journal of Geophysical Research: Atmospheres*, **122 (14)**, 7461–7492.

Thompson, D. B., and P. E. Roundy, 2013: The relationship between the Madden–Julian oscillation and US violent tornado outbreaks in the spring. *Monthly Weather Review*, **141 (6)**, 2087–2095.

Tippett, M. K., 2018: Robustness of relations between the MJO and US tornado occurrence. *Monthly Weather Review*, **146 (11)**, 3873–3884.

Tippett, M. K., and W. J. Koshak, 2018: A baseline for the predictability of US cloud-to-ground lightning. *Geophysical Research Letters*, **45 (19)**, 10–719.

Tippett, M. K., and C. Lepore, 2021: ENSO-based predictability of a regional severe thunderstorm index. *Geophysical Research Letters*, **48 (18)**, e2021GL094 907.

Tippett, M. K., C. Lepore, W. J. Koshak, T. Chronis, and B. Vant-Hull, 2019: Performance of a simple reanalysis proxy for US cloud-to-ground lightning. *International Journal of Climatology*, **39 (10)**, 3932–3946.

Villarini, G., and J. A. Smith, 2013: Spatial and temporal variability of cloud-to-ground lightning over the continental US during the period 1995–2010. *Atmospheric research*, **124**, 137–148.

Virts, K. S., and R. A. Houze, 2015: Variation of lightning and convective rain fraction in mesoscale convective systems of the MJO. *Journal of the Atmospheric Sciences*, **72 (5)**, 1932–1944.

Vitart, F., and Coauthors, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, **98 (1)**, 163–173.

Westermayer, A., P. Groenemeijer, G. Pistotnik, R. Sausen, and E. Faust, 2017: Identification of favorable environments for thunderstorms in reanalysis data. *Meteorologische Zeitschrift*, **26 (1)**, 59–70.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, **132 (8)**, 1917–1932.

Wilks, D., 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, **97 (12)**, 2263–2273.

Xu, W., R. F. Adler, and N.-Y. Wang, 2013: Improving geostationary satellite rainfall estimates using lightning observations: Underlying lightning–rainfall–cloud relationships. *Journal of applied meteorology and climatology*, **52 (1)**, 213–229.

Yair, Y., B. Lynn, C. Price, V. Kotroni, K. Lagouvardos, E. Morin, A. Mugnai, and M. d. C. Llasat, 2010: Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields. *Journal of Geophysical Research: Atmospheres*, **115 (D4)**.

Zajac, B. A., and S. A. Rutledge, 2001: Cloud-to-ground lightning activity in the contiguous United States from 1995 to 1999. *Monthly Weather Review*, **129 (5)**, 999–1019.

Zhao, P., H. Xiao, C. Liu, Y. Zhou, X. Xu, and K. Hao, 2022: Evaluating a simple proxy for climatic cloud-to-ground lightning in Sichuan Province with complex terrain, Southwest China. *International Journal of Climatology*, **42 (7)**, 3909–3927.

Zipser, E. J., 1994: Deep cumulonimbus cloud systems in the tropics with and without lightning. *Monthly weather review*, **122 (8)**, 1837–1851.