Trends in the Extremes of Environments Associated with Severe U.S. Thunderstorms

ERWAN KOCH,[a,d] JONATHAN KOH,[a,d] ANTHONY C. DAVISON,[a] CHIARA LEPORE,[b] AND MICHAEL K. TIPPETT[c]

[a] *Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[b] *Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*
[c] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York*

(Manuscript received 30 October 2019, in final form 26 February 2020)

ABSTRACT: Severe thunderstorms can have devastating impacts. Concurrently high values of convective available potential energy (CAPE) and storm relative helicity (SRH) are known to be conducive to severe weather, so high values of PROD = $(\text{CAPE})^{1/2} \times \text{SRH}$ have been used to indicate high risk of severe thunderstorms. We consider the extreme values of these three variables for a large area of the contiguous United States over the period 1979–2015 and use extreme-value theory and a multiple testing procedure to show that there is a significant time trend in the extremes for PROD maxima in April, May, and August, for CAPE maxima in April, May, and June, and for maxima of SRH in April and May. These observed increases in CAPE are also relevant for rainfall extremes and are expected in a warmer climate but have not previously been reported. Moreover, we show that El Niño–Southern Oscillation explains variation in the extremes of PROD and SRH in February. Our results suggest that the risk from severe thunderstorms in April and May is increasing in parts of the United States where it was already high and that the risk from storms in February is increased over the main part of the region during La Niña years.

KEYWORDS: Extreme events; Thunderstorms; Statistical techniques; Southern Oscillation; Trends; Risk assessment

1. Introduction

Annual losses from severe thunderstorms in the United States have exceeded $10 billion in recent years (Saville 2019). In addition to economic losses, 2011 was marked by 552 deaths caused by tornadoes. These economic and human impacts are a strong motivation to study how and why U.S. thunderstorm activity varies from year to year and region to region. Two important aspects are trends potentially related to climate change or multidecadal variability, and modulation by El Niño–Southern Oscillation (ENSO). However, inadequacies in the length and quality of the thunderstorm data record present substantial challenges to addressing these questions directly (Verbout et al. 2006; Allen and Tippett 2015; Edwards et al. 2018).

In the United States, a severe thunderstorm is defined to be one that produces a tornado, hail greater than 1 in. (2.54 cm) in diameter, or wind gusts in excess of 50 kt ($1\text{ kt} \approx 0.51\text{ m s}^{-1}$). Supercell storms are responsible for a large fraction of severe thunderstorm reports [e.g., 79% of tornadoes according to Trapp et al. (2005)], even though only about 10% of thunderstorms are supercells (Doswell 2015), and a key element in forecasting severe thunderstorms is the prediction of where and when supercells will occur (Corfidi 2017). A supercell is a thunderstorm with a deep, long-lived rotating updraft (mesocyclone). The presence of buoyancy, that is, convective available potential energy (CAPE), and deep-layer vertical wind shear are important determinants for supercell development. In addition to the magnitude of the vertical shear, the angle between surface and upper-level winds is important for mesocyclone development and persistence. A key quantity is atmospheric helicity, which is computed relative to storm motion and is proportional to vertical wind shear and the amount of wind direction turning from the surface to upper levels (often 0–3 km).

Several recent studies of U.S. tornado reports have concluded that annual numbers of reliably observed tornadoes (i.e., those rated F/EF1 and greater on the Fujita or enhanced Fujita scale) show slight but statistically insignificant trends downward over time (Brooks et al. 2014; Elsner et al. 2015), whereas measures of tornado outbreaks or clusters show upward trends (Brooks et al. 2014; Elsner et al. 2015; Tippett et al. 2016). Changes in regional tornado activity have also been reported (Agee et al. 2016; Gensini and Brooks 2018), but there is less evidence for changes in hail and damaging straight-line wind, perhaps because of the poorer quality of the relevant databases.

In view of the limitations of the historical storm record, a valuable alternative is the analysis of meteorological environments associated with severe thunderstorms. As mentioned above, severe thunderstorms, especially supercell storms, are more likely in the presence of high values of CAPE and of certain measures of vertical wind shear (see, e.g., Brooks et al. 2003; Brooks 2013) such as storm relative helicity (SRH).

Denotes content that is immediately available upon publication as open access.

[d] These authors contributed equally to this paper.

Corresponding author: Erwan Koch, erwan.koch@epfl.ch

DOI: 10.1175/JCLI-D-19-0826.1

Weather forecasters have routinely used such quantities for two decades to interpret observations and the output of numerical weather prediction models (Johns et al. 1993; Rasmussen and Blanchard 1998; Doswell et al. 1996), and they are also useful in climatological studies, especially in areas outside the United States without extensive historical reports (Brooks et al. 2003). The environmental approach can also provide an indication of expected severe thunderstorm activity in a warmer climate based on climate projections that do not resolve thunderstorms explicitly (Trapp et al. 2009; Diffenbaugh et al. 2013). On time scales between weather forecasts and climate projections, this approach has provided a clearer picture of how ENSO modulates U.S. hail and tornado activity (Allen et al. 2015; Lepore et al. 2017).

Previous statistical studies of environments associated with severe thunderstorms have diagnosed relationships with ENSO based on monthly averages, which are indirect proxies for behavior on the time scale of weather. Similarly, Gensini and Brooks (2018) computed monthly accumulations of daily maxima of a significant tornado parameter. The submonthly environmental data used by Tippett et al. (2016) were aggregated on an annual and U.S.-wide basis. By contrast, the present work focuses on extremes of the environmental values rather than on monthly averages and presents results that are spatially and temporally resolved. The framework that we use is statistical extreme-value theory.

Gilleland et al. (2013) apply the conditional extreme-value framework of Heffernan and Tawn (2004) to the product $\mathrm{WS} \times W_{\max}$, where WS is a measure of wind shear and $W_{\max} = (2 \times \mathrm{CAPE})^{1/2}$, by conditioning on the 75th percentile of that variable computed across the spatial domain. This approach has the advantage of allowing the study of real spatial patterns under severe conditions, as opposed to approaches looking at pointwise maxima. They show some temporal variations in the mean simulated values from their model.

Mannshardt and Gilleland (2013) perform an unconditional univariate analysis in which they fit the generalized extreme-value (GEV) distribution to the annual maxima of $\mathrm{WS} \times W_{\max}$ and establish the existence of a time trend in the GEV location parameter. Heaton et al. (2011) consider three Bayesian hierarchical extreme-value models based on exceedances over a high threshold for $\mathrm{WS} \times W_{\max}$, their third model being based on a Poisson point process with a yearly time trend. Neither paper clarifies whether this trend is attributable to both CAPE and WS or only to one of them. Moreover, both articles consider trends in annual quantities and thus cannot detect month-specific features, and they do not account for multiple testing, a correction for which had earlier been applied by Gilleland et al. (2008). Finally, these papers consider only time as a covariate.

Our study covers a large part of the contiguous United States for individual months from 1979 to 2015 and we consider CAPE, SRH (0–3 km), and the combined variable $\mathrm{PROD} = (\mathrm{CAPE})^{1/2} \times \mathrm{SRH}$ separately. To motivate our use of PROD, we consider the discriminant line defined in Brooks et al. [2003, their Eq. (1)], which is one of the first thresholds used to distinguish low and high likelihoods of severe thunderstorm occurrence using a function of CAPE and vertical shear. This equation can be rewritten as $\mathrm{S6} \times \mathrm{CAPE}^{0.62} = 18.60$, where S6 is the 0–6-km shear. Replacing S6 with 0–3-km SRH and approximating the power 0.62 by 0.5 leads to a discriminant line of the form $\mathrm{SRH} \times (\mathrm{CAPE})^{1/2} = c$—that is, $\mathrm{PROD} = c$, where c is a real constant—and shows that high values of PROD can be expected to be indicative of an elevated risk of severe thunderstorms. PROD has already been used as a proxy for severe thunderstorms (e.g., Tippett et al. 2016), and the plot of Fig. 1 in Brooks et al. (2003) is little changed by replacing S6 with 0–3-km SRH (not shown). More generally, the product of CAPE and two shear-related variables (different or not), or equivalently its square root, is commonly used as an indicator of the likelihood of severe thunderstorm occurrence. For instance, the significant tornado parameter (STP) and the supercell composite parameter (SCP) involve the product of CAPE, S6, and 0–1-km SRH and the product of CAPE, S6, and 0–3-km SRH, respectively (e.g., Thompson et al. 2003).

To ensure the soundness of our results we carefully check the suitability of the GEV model and the use of time and ENSO as explanatory variables in its location parameter, and we account for multiple testing by implementing the false discovery rate procedure of Benjamini and Hochberg (1995). As stated in Gilleland et al. (2013, section 1 therein), in addition to studying PROD, it is insightful to consider its components separately. Furthermore, accounting for multiple testing is essential when testing many hypotheses simultaneously, as stressed by Gilleland et al. (2013, section 4 therein).

We find a significant time trend in the GEV location parameter for PROD maxima in April, May, and August (and to a lesser extent in June and December), in CAPE maxima in April, May, and June (and to a lesser extent in August, November, and January), and in SRH maxima in May (and to a lesser extent in April). The trends in CAPE maxima are striking because CAPE is expected to increase in a warming climate (Del Genio et al. 2007; Van Klooster and Roebber 2009) and are relevant to rainfall extremes (Lepore et al. 2015), but have not previously been observed over the United States. April and May are important months for PROD, as severe thunderstorms are frequent at this period. The corresponding time slope is positive in regions of the United States where severe thunderstorms are already common, which may have implications for risk assessment and management. Our study also reveals that ENSO can explain variation in the GEV location parameter for PROD and SRH maxima in February. The corresponding slope is negative over most of the region we consider, possibly suggesting an increased risk of high storm impacts in February during La Niña years. Our results differ from those of Heaton et al. (2011), Mannshardt and Gilleland (2013), and Gilleland et al. (2013), but are fairly consistent with those obtained by Gensini and Brooks (2018), who inter alia consider the numbers of tornado reports.

The remainder of the paper is organized as follows. Section 2 presents the data and a brief exploratory analysis. We describe our statistical approach and demonstrate its relevance in section 3. Section 4 details our main results, and section 5 summarizes our findings and discusses them.

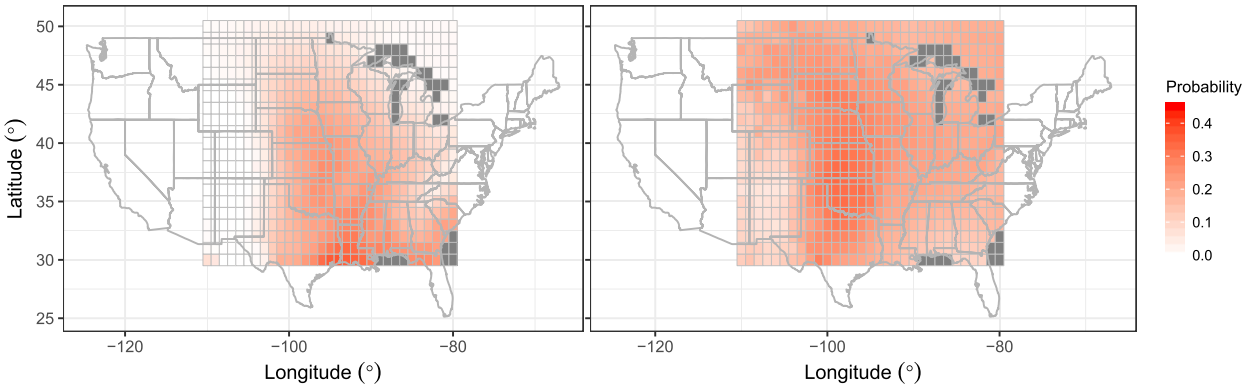

FIG. 1. Empirical pointwise probabilities of (left) 3-hourly CAPE exceeding 1400 J kg^{-1} and (right) SRH exceeding 170 m^2 s^{-2} for the entire period 1979–2015. Dark gray corresponds to grid points for which no observations are available.

2. Data and exploratory analysis

The data we investigate consist of 3-hourly time series of 0–180-hPa CAPE (J kg^{-1}) and 0–3-km SRH (m^2 s^{-2}) from 0000 UTC 1 January 1979 to 2100 UTC 31 December 2015. For consistency, we did not account for the data recorded on 29 February of leap years, and this does not affect our conclusions. The region covered is a rectangle over the contiguous United States from −110° to −80° longitude and from 30° to 50° latitude, and the resolution is 1° longitude and 1° latitude. These data constitute a coarse version of reanalysis data from the North American Regional Reanalysis (NARR); the original resolution is 32 km in longitude and 32 km in latitude (see, e.g., Mesinger et al. 2006). The region contains 651 grid points, with no data available for 32 grid points over the sea or lakes. Using these time series, we build 3-hourly time series of PROD = $(\text{CAPE})^{1/2} \times \text{SRH}$, measured in meters cubed per second cubed. As a physical covariate we use monthly values of the Niño-3.4 index (°C) from 1979 to 2015, taken from the ERSSTv5 dataset available on the NOAA Climate Prediction Center website.

Figure 1 shows the empirical pointwise probabilities that CAPE and SRH exceed thresholds corresponding to roughly the 90th percentile of each variable across the entire region. There is a clear north–south gradient for CAPE probabilities, while the regional spatial pattern for SRH suggests that the high values cluster toward the center of the region.

Figure 2 shows an increase in the exceedance probabilities for PROD at many grid points over the decades; a similar result is visible for SRH, but less so for CAPE. This increase is of interest for risk assessment, especially in regions with a high risk of severe thunderstorms. Figure 2 strongly suggests the presence of a temporal trend in the maxima, but there seems to be no geographical shift, notwithstanding the results of Gilleland et al. (2013).

The top left panel of Fig. 3 shows a positive correlation between PROD April maxima and time for many grid points, and the middle panels show a positive linear time trend for April maxima of PROD, CAPE, and SRH in the subregion indicated. The top right panel shows strong negative correlation between PROD February maxima and ENSO at many grid points, while the scatterplots in the bottom panels suggest that an increase in ENSO from −2° to 2°C is associated with a reduction in PROD and CAPE by a factor of roughly 2, with a more complex pattern for SRH. These analyses underscore the need to incorporate ENSO into our statistical modeling of maxima.

3. Method

a. Modeling of maxima

Risk assessment entails the estimation of return levels associated with very high return periods and of the probabilities of observing events so extreme that they have never occurred before. Extreme-value theory provides a solid framework for the extrapolation needed to perform these tasks for the maxima of PROD, CAPE, and SRH. Here we present the statistical background to the results in section 4; for further explanation and references see Coles (2001) or Davison and Huser (2015).

Let M_n denote the maximum of independent and identically distributed random variables $X_1, \ldots, X_n$. The extremal types theorem states that if there exist sequences $\{a_n\} > 0$ and $\{b_n\} \in \mathbb{R}$ such that $(M_n - b_n)/a_n$ has a nondegenerate limiting distribution as $n \to \infty$, then this distribution must be of generalized extreme-value form:

$$\text{GEV}_{\eta,\tau,\xi}(x) = \begin{cases} \exp\{-[1+\xi(x-\eta)/\tau]_+^{-1/\xi}\}, & \xi \neq 0, \\ \exp\{-\exp[-(x-\eta)/\tau]_+\}, & \xi = 0, \end{cases} \quad x \in \mathbb{R},$$

where ξ and η are real valued, $\tau > 0$, and, for any real a, $a_+ = \max(a, 0)$. If n is large enough, this theorem suggests using the approximation

$$\mathbb{P}(M_n \leq x) \approx \text{GEV}_{\eta,\tau,\xi}(x), \quad x \in \mathbb{R}, \tag{1}$$

for suitably chosen η, τ, and ξ, which are location, scale, and shape parameters. The latter defines the type of the distribution: $\xi > 0$, $\xi < 0$, and $\xi = 0$ correspond to the Fréchet, Weibull, and Gumbel types and allow very different statistical behaviors,

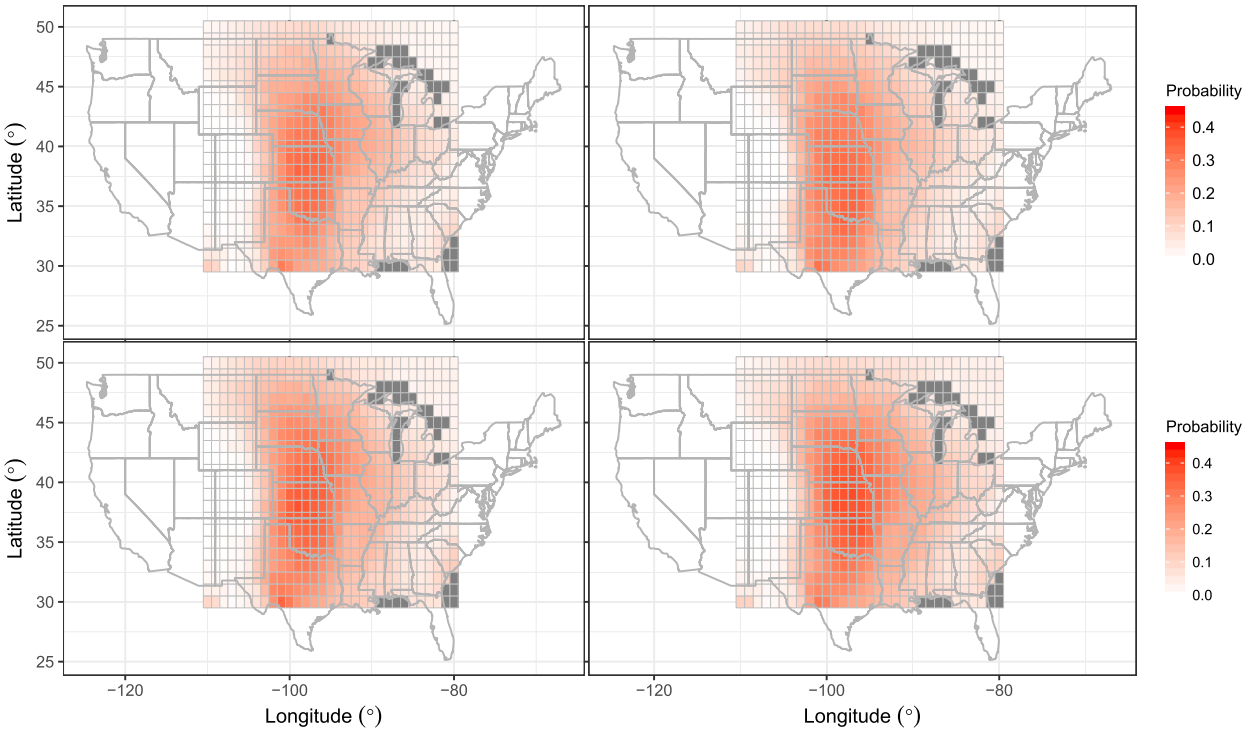

FIG. 2. Empirical pointwise probabilities of 3-hourly PROD exceeding 3300 $m^3 s^{-3}$ during the periods (top left) 1979–87, (top right) 1988–96, (bottom left) 1997–2005, and (bottom right) 2006–15.

with the first giving a heavy upper tail with polynomial decay, the second modeling bounded variables, and the third being an intermediate case, unbounded with an exponentially decaying upper tail.

The GEV approximation for maxima remains valid if the variables are dependent, provided that distant extremes are "nearly independent," or, more formally, that Leadbetter's $D(u_n)$ condition is satisfied. We shall see below that near independence is credible for our time series, so it is plausible that (1) applies.

The results above provide a natural model for maxima of stationary sequences. To apply this model, we split the data into blocks of equal lengths and compute the maximum of each block. Assume that we have T blocks of length n and let $M_n^{(1)}, \ldots, M_n^{(T)}$ denote the corresponding maxima. If n is large enough, the distribution of the $M_n^{(t)}$ is approximately (1), and inference from fitting this model to the $M_n^{(t)}$ is commonly called the block maximum method. As noted in section 2, PROD, CAPE, and SRH maxima exhibit a time trend and/or a relation with ENSO for some months, and we can allow the GEV parameters to depend upon these variables. Figure 4 and results in section 4 show that the temporal or ENSO effects only appear for certain months. For instance, time trends for PROD, CAPE, and SRH are mainly present in April and May, April to June, and April and May, respectively. We therefore choose our blocks to be the months and study each month separately, fitting the models

$$M_n^{(t)} \sim \mathrm{GEV}_{\eta_{\mathrm{ti}}(t),\tau_{\mathrm{ti}},\xi_{\mathrm{ti}}}, \quad \eta_{\mathrm{ti}}(t) = \eta_{0,\mathrm{ti}} + \eta_{1,\mathrm{ti}} t, \quad t = 1, \ldots, T, \tag{2}$$

and

$$M_n^{(t)} \sim \mathrm{GEV}_{\eta_{\mathrm{en}}(t),\tau_{\mathrm{en}},\xi_{\mathrm{en}}}, \quad \eta_{\mathrm{en}}(t) = \eta_{0,\mathrm{en}} + \eta_{1,\mathrm{en}} \mathrm{ENSO}_t, \quad t = 1, \ldots, T, \tag{3}$$

where the subscripts ti and en refer to the dependence on time and on ENSO, respectively, $\eta_{0,\mathrm{ti}}$, $\eta_{1,\mathrm{ti}}$, $\eta_{0,\mathrm{en}}$, $\eta_{1,\mathrm{en}}$, ξ_{ti} and ξ_{en} are real valued, τ_{ti} and τ_{en} are positive, ENSO_t is the value of ENSO in that month for year t, and n is equal to 224, 232, 240, or 248, depending on the number of days in the month, because we have eight observations per day. Figure 3 suggests that effects of time and ENSO on maxima are roughly linear. For parsimony and owing to the satisfactory goodness-of-fit tests presented in section 3b, we incorporate the effects of time and ENSO only into the location parameter η and take the scale and shape parameters to be constant; it is generally unwise to allow the shape parameter to depend on a covariate owing to the large uncertainty of its estimate. The time trend induces nonstationarity between the blocks (i.e., across years) but does not violate the within-block stationarity assumption; see below. Figure 4 suggests that the time trend does not stem from a shift of seasonality.

We compute the monthly maximum for each month and a given grid point and thereby obtain the maxima $M_{248}^{(1)}, \ldots, M_{248}^{(37)}$ for January, say. We then fit the models in (2) and (3) by numerical maximum likelihood estimation for each month and grid point.

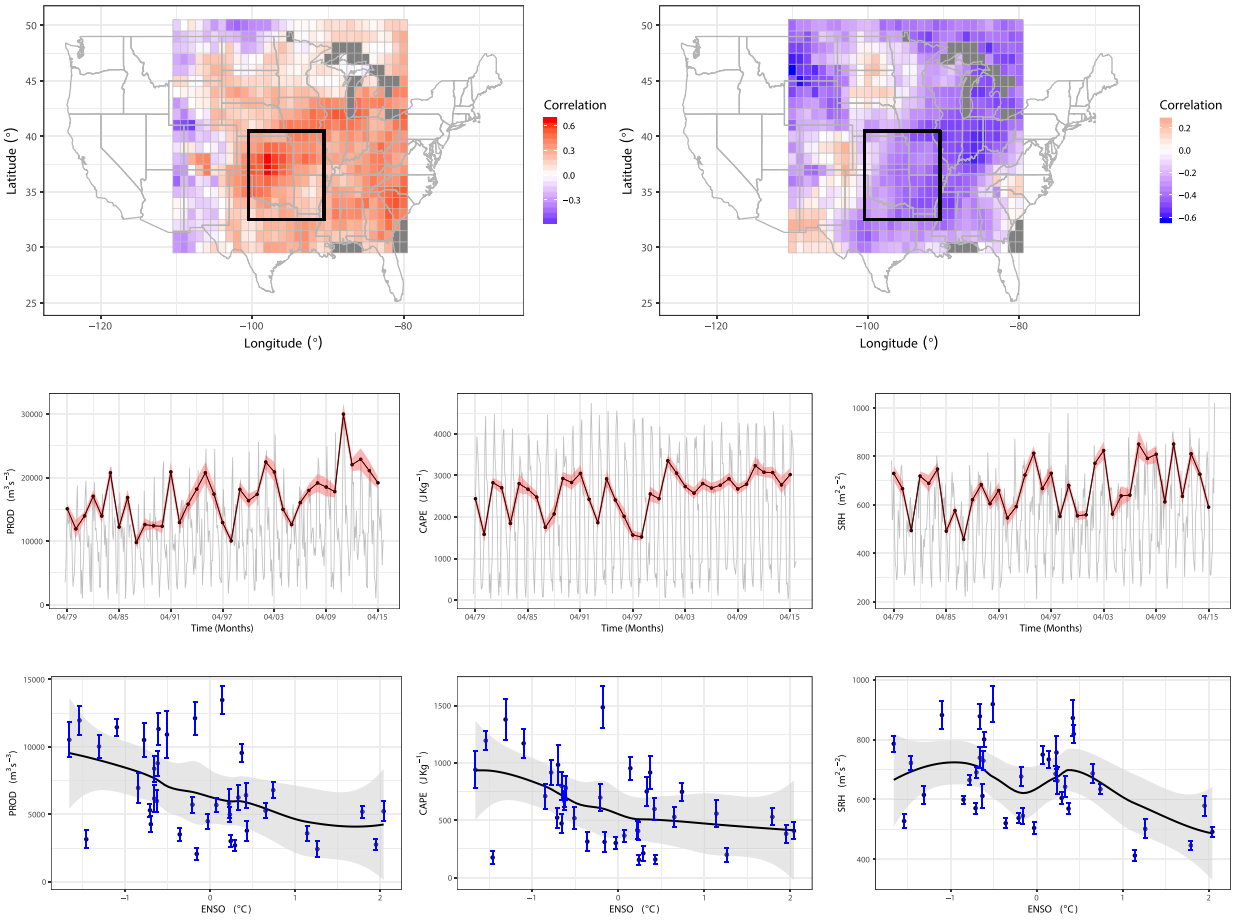

FIG. 3. Exploratory analysis for monthly maxima: (top left) The correlation map with time (in years from 1 to 37) for PROD April maxima, and (top right) the correlation map with ENSO for PROD February maxima. Also shown are (left) PROD, (center) CAPE, and (right) SRH analyses on a subregion indicated by the black-outlined rectangle drawn on the correlation maps, showing (middle) the region-averaged monthly maxima time series across all 444 months (gray lines), the region-averaged April maxima time series (black lines), and its 95% confidence interval bounds (red shaded region) (every point in the time series is the averaged maxima across all grid points in the subregion indicated before, for a particular month and a particular year) and (bottom) scatterplots of the region-averaged February maxima with ENSO, along with the 95% confidence interval bounds at each point, indicated by the blue whiskers (the black line represents the best fitted local regression trend estimate, with its 95% confidence interval bounds indicated by the shaded gray region).

Recall that, provided the block size n is large enough, within-block stationarity and the $D(u_n)$ condition ensure the validity of (1) and hence allow us to consider the models in (2) and (3). To check the plausibility of these two properties, we consider the 3-hourly time series of PROD, CAPE, and SRH at 50 representative grid points. For each block (associated with a triplet grid point–month–year), we fit several autoregressive moving average (ARMA) processes to the corresponding time series, choose the fit that minimizes the Akaike information criterion, and use a Box–Pierce procedure to assess the independence of the corresponding residuals; we find no systematic departure from independence or stationarity. Often the residual distribution appears to lie in the Fréchet or Gumbel maximum domains of attraction, and Embrechts et al. (1997, section 5.5 therein) show that in such cases convergence of the maxima to the GEV limit occurs even for ARMA processes. Hence the time series of data within the months seem to satisfy both stationarity and the $D(u_n)$ condition. Choosing the months as blocks thus appears reasonable, as is confirmed by our analysis in the following section, whereas choosing seasons or years as blocks would mask many interesting features, and the sample size associated with day- or week-long blocks is too low for the GEV approximation in (1) to be reasonable.

b. *Assessment of fit*

At each grid point and for each month, we fit the models in (2) and (3) to the monthly maxima, as described in section 3a. We use the Kolmogorov–Smirnov statistic to assess the distributional proximity between these fitted models and the empirical distribution of the 37 observed monthly maxima. For PROD, CAPE, and SRH, in most months, the fit appears acceptable at the 5% level at all grid points. These good in-sample fits of the models in (2) and (3) for all variables are

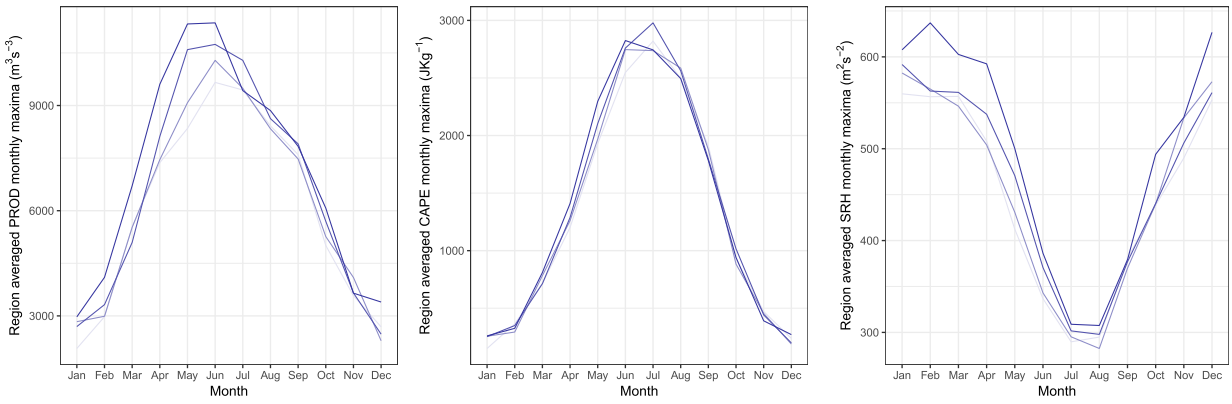

FIG. 4. Whole region–averaged monthly maxima of (left) PROD, (center) CAPE, and (right) SRH. The four lines colored from light blue to dark blue correspond to the time periods 1979–87, 1988–96, 1997–2005, and 2006–15, respectively.

confirmed by the quantile–quantile (QQ) plots, which are displayed for one grid point in Fig. 5.

However, these results do not take into account the fitting of the GEV distribution to the data, which systematically decreases the values of the Kolmogorov–Smirnov statistic. To make an informal allowance for this decrease, for each grid point i and month j we

1) fit the GEV distribution using the pooled observations from the eight grid points nearest to i to obtain $\hat{\eta}_{\mathrm{po}_i,j}$, $\hat{\tau}_{\mathrm{po}_i,j}$, and $\hat{\xi}_{\mathrm{po}_i,j}$ and then
2) use a Kolmogorov–Smirnov test to check the agreement between the "out sample" GEV fit with parameters $\hat{\eta}_{\mathrm{po}_i,j}$, $\hat{\tau}_{\mathrm{po}_i,j}$, and $\hat{\xi}_{\mathrm{po}_i,j}$, and the empirical distribution of the 37 observed monthly maxima at grid point i.

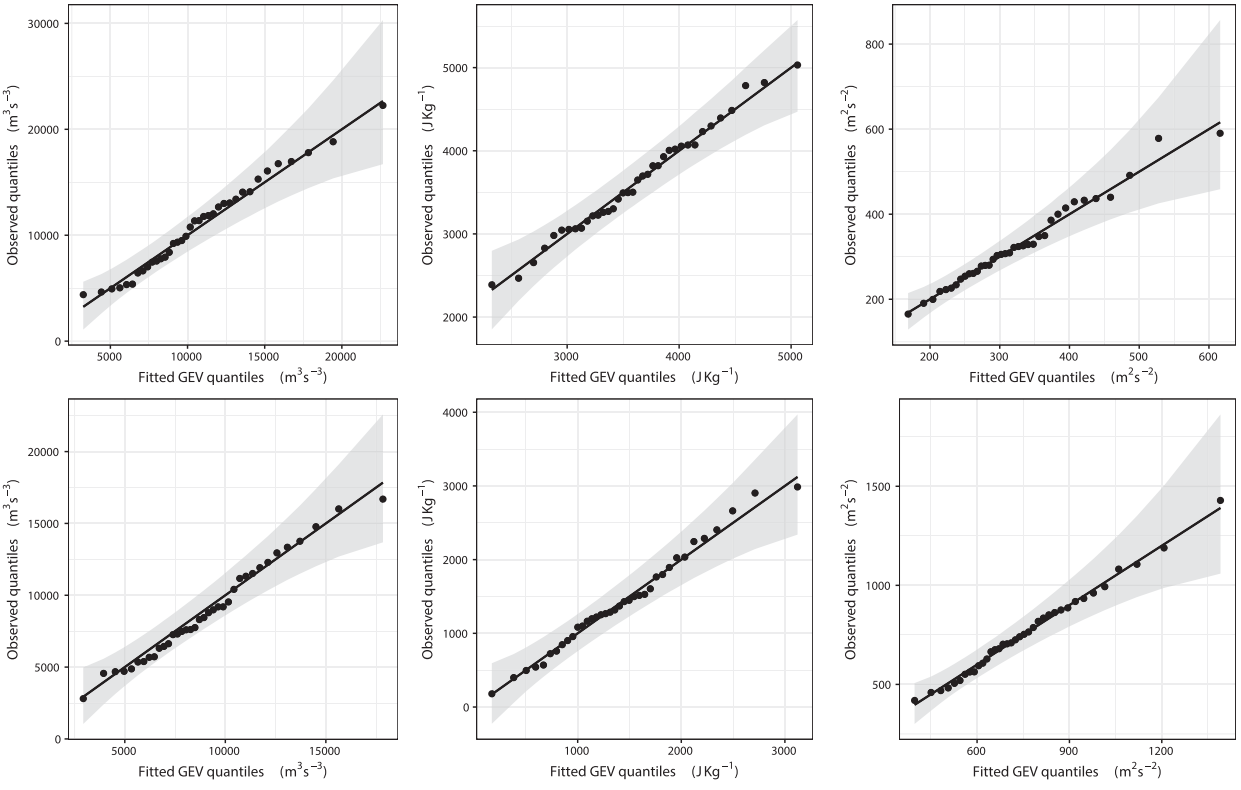

FIG. 5. Assessment of the in-sample fit of the models in (2) and (3), using QQ plots for (left) PROD, (center) CAPE, and (right) SRH maxima at the grid point whose center has coordinates 31°, −91° for (top) the model in (2) for May and (bottom) the model in (3) for February. The shaded regions indicate the 95% confidence bounds.

TABLE 1. Assessment of the out-sample fit of the GEV model: Number of rejections from our out-sample Kolmogorov–Smirnov test (at the 5% level and without accounting for multiple testing) for each variable and each month. For each of PROD, CAPE, and SRH, the first row gives the observed number of rejections and the second and third rows respectively provide the 5% and 95% quantiles of the empirical distributions of the number of rejections obtained from the simulation study.

Variable	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PROD	42	27	50	29	52	58	67	71	55	27	39	34
Sim PROD 5%	37	34	33	33	35	35	37	38	38	36	36	37
Sim PROD 95%	57	57	55	55	57	58	57	58	57	58	58	57
CAPE	59	33	48	34	60	64	71	90	67	43	58	74
Sim CAPE 5%	41	36	37	36	36	36	38	42	39	36	36	37
Sim CAPE 95%	59	58	60	56	57	58	61	63	61	60	58	59
SRH	36	23	24	21	22	42	42	34	35	26	24	36
Sim SRH 5%	36	36	34	35	34	36	36	36	34	36	36	34
Sim SRH 95%	60	59	59	53	57	57	57	58	58	58	61	57

We then repeat these two steps 100 times with data simulated from independent GEV models fitted at each grid point and compute the 5% and 95% quantiles of the empirical distribution of the number of rejections. Table 1 shows that, for all variables, the observed numbers of rejections are low compared to the number of grid points (619), especially as we did not account for multiple testing. Moreover, they are not hugely different from those obtained in the simulation study, although often slightly above the 95% quantile in the case of CAPE and slightly below the 5% quantile for SRH and PROD. These discrepancies may be explained by the substantial spatial dependence present in the data but not in the simulations. This analysis supports the use of the GEV model at grid points at which no data are available and thus goes beyond the initial goal of assessment of its fit.

Because (2) and (3) fit the data adequately, we conclude that they provide suitable models for the monthly maxima of PROD, CAPE, and SRH.

c. *Testing procedure*

1) GENERAL

In section 4, we assess whether time and ENSO affect the GEV location parameter for PROD, CAPE, and SRH at each of the $m = 619$ grid points.

We first discuss the statistic used to test the significance of time and ENSO, respectively, in (2) and (3), at a single grid point. In the first case, we have to test the null and alternative hypotheses

$$H_0: \eta_{1,\mathrm{ti}} = 0 \quad \text{vs} \quad H_A: \eta_{1,\mathrm{ti}} \neq 0$$

by comparing the fits of the models

$$\mathrm{M}_0: \eta_{\mathrm{ti}}(t) = \eta_{0,\mathrm{ti}}, \quad \mathrm{M}_1: \eta_{\mathrm{ti}}(t) = \eta_{0,\mathrm{ti}} + \eta_{1,\mathrm{ti}} t, \quad t = 1, \ldots, 37,$$

and similarly for ENSO. We let $\ell_0(\mathrm{M}_0)$ and $\ell_1(\mathrm{M}_1)$ denote the maximized log-likelihoods for the models M_0 and M_1 and compute the signed likelihood ratio statistic

$$\tilde{T} = \mathrm{sgn}(\hat{\eta}_{1,\mathrm{ti}})\{2[\ell_1(\mathrm{M}_1) - \ell_0(\mathrm{M}_0)]\}^{1/2},$$

where $\mathrm{sgn}(\hat{\eta}_{1,\mathrm{ti}})$ is the sign of the estimated trend under model M_1; $\tilde{T}$ has an approximate standard Gaussian distribution under H_0, and the corresponding p value is $p = 2\Phi(-|\tilde{t}|)$, with $\tilde{t}$ being the observed value of $\tilde{T}$ and Φ denoting the standard Gaussian distribution function.

This test would be valid if applied at a single prespecified grid point, but we must make allowance for the facts that we shall perform the same test at m grid points and that spatial patterns in the effects of time and ENSO are likely to induce correlation among the p values for nearby grid points. We now discuss how to obtain a valid testing procedure despite these facts.

2) MULTIPLE TESTING

A popular approach for multiple testing in climatology is the field significance test of Livezey and Chen (1983), but unfortunately this gives little insight about where the results are significant, which is of high interest to us. The regression approach of DelSole and Yang (2011) has the same drawback. Among methods to identify which of the grid points have significant results are those, such as the Bonferroni procedure, that bound the probability that the number of falsely rejected null hypotheses exceeds unity. When the number of hypotheses is high, however, such methods are so stringent that their power is very low.

Benjamini and Hochberg (1995) introduce the false discovery rate (FDR), namely the expected proportion of incorrect rejections (false positives) out of all rejections, and propose a procedure to ensure that the FDR is below a given level q when performing multiple testing. Their approach, which we call the BH procedure, would reject H_0 at all grid points i such that $p_i \leq p_{(k)}$, where $p_{(1)} \leq \cdots \leq p_{(m)}$ are the ordered p values and

$$k = \max\left(i : p_{(i)} \leq q\frac{i}{m}\right).$$

The BH procedure ensures that the FDR is below qm_0/m, where m_0 denotes the unknown number of grid points at which H_0 is true, and is said to control the FDR at level qm_0/m.

Let m_A be the unknown number of grid points at which H_A is true and, for a chosen q, let S_q count the grid points at which a particular covariate is declared significant by the BH procedure. Then, because m_A is equal to the sum of the numbers of true positives and false negatives, we expect that

$$m_A \geq (1-q)S_q. \tag{4}$$

Because the BH procedure ensures that the FDR is not more than $qm_0/m = q(m - m_A)/m$, we may argue a posteriori that we have controlled the FDR at level

$$q^{(1)} = \frac{q[m-(1-q)S_q]}{m},$$

which entails that $m_A \geq [1 - q^{(1)}]S_q$. If we iterate this argument by defining

$$q^{(n+1)} = \frac{q\{m-[1-q^{(n)}]S_q\}}{m}, \quad n = 1, 2, \ldots,$$

we see that the effective level at which we have controlled the FDR is $q_{\lim} = \lim_{n\to\infty} q^{(n)}$. The limit is often well approximated after just a few iterations, and then we have

$$m_A \geq (1-q_{\lim})S_q. \tag{5}$$

The BH procedure was originally shown to be valid for independent test statistics, but Benjamini and Yekutieli (2001, their theorem 1.2) prove that it controls the FDR at level qm_0/m if the statistics have a certain form of positive dependence. Ventura et al. (2004) apply this procedure to simulations representative of climatological data and covering the range of correlation scales likely to be encountered in practice, and find empirically that it controls the FDR at level qm_0/m. Yekutieli and Benjamini (1999) and Benjamini and Yekutieli (2001) propose two modifications to account for more general dependence between the test statistics. The first is complex and does not much improve on the BH procedure, whereas the second is applicable for any dependence structure but has greatly reduced power, so Ventura et al. (2004) recommend the use of the BH procedure.

Test statistics and p values based on our data are clearly dependent, but as our data resemble those considered in Ventura et al. (2004), applying the BH procedure at level q should control the FDR at level qm_0/m. A more rigorous argument would show that the m components of our test statistic are asymptotically jointly Gaussian and that theorem 1.2 of Benjamini and Yekutieli (2001) can be applied, but this is outside the scope of the present paper.

4. Results

In this section we quantify the effects of time and ENSO in the GEV location parameter and study their significance, using $q = 0.05$ and $q = 0.2$, corresponding to control of the FDR at the nominal levels 5% and 20%. In each case we first discuss PROD, which is the main variable of interest for severe thunderstorm risk, and then consider CAPE and SRH.

We begin with the effect of time. Table 2 shows that many of the 619 grid points exhibit a significant time trend for PROD in April, May, and August (and to a lesser extent in June and December). In April, this number equals 313 at the 20% level, so (4) implies that at least 250 of these grid points indeed have a trend; with (5), this number rises to 278. Figure 6 indicates that, in April, the Northeast, a very wide Southeast corner, and the Southwest show significant time trends. In the first two regions, $\hat{\eta}_{1,\mathrm{ti}}$ is positive, corresponding to an increasing risk of severe thunderstorm impacts, particularly in already risky regions. Similar conclusions may be drawn from Fig. 7 in the case of May, although the Southeast is less prominent. The highest slope value corresponds to an annual increase of PROD maxima of about 3% of the corresponding PROD maximum recorded in 1979. Mannshardt and Gilleland (2013) and Heaton et al. (2011) do not find such a significantly positive time trend over the entire region most at risk, sometimes called tornado alley, nor do they find significantly positive trends in the Northeast of our region, whereas they find a significant positive trend toward the West. These differences probably arise because these earlier papers consider a less recent period (1958–99) and their product variable is slightly different from ours. Moreover, Mannshardt and Gilleland (2013) study annual rather than monthly maxima. The discrepancies with Heaton et al. (2011) may also be due to their analysis of threshold exceedances and a different, Bayesian hierarchical, approach. The evolution found by Gilleland et al. (2013) between the second (1979–92) and the third (1993–99) period is quite consistent with our trends in spring; for the other seasons, however, the results differ appreciably. There are also many dissimilarities in the changes between the first (1958–78) and the second (1979–92) periods, but the first period does not belong to the time range we consider. Gilleland et al. (2013) consider the mean simulated values conditional on the total amount of energy being large, and then not all grid point values need be extreme, whereas we analyze maxima at each grid point. Moreover, the trends we find account for the year-to-year variation, whereas the results presented in Gilleland et al. (2013) only allow changes to be assessed by comparing values for three successive periods of about 15 to 20 years. The positive time trends we detect in spring appear quite consistent with the results of Gensini and Brooks (2018), who use much more recent data than the papers previously described. The remaining differences, especially for Texas, may arise for the following reasons. First, as PROD is only an indicator of severe weather, there are necessarily discrepancies with results based on effective tornado reports. Second, PROD slightly differs from STP, so the corresponding results may differ somewhat. Furthermore, the findings of Gensini and Brooks (2018) about reports concern the total number of tornadoes per year, and those about STP are not based on the maxima of that variable.

With regard to CAPE, April, May, and June (and to a lesser extent, August, November, and January) show many grid points with a significant time trend. For April and May, Figs. 6 and 7 show significantly negative $\hat{\eta}_{1,\mathrm{ti}}$ in the West, contrasting with a significantly positive trend in the central region and the East. As pointed out by Trapp et al. (2009) and Diffenbaugh et al. (2013), a positive time trend for CAPE is expected in a context of climate change. However, to the best of our knowledge, an *observed* trend has not been previously reported in the literature.

For SRH, May and to a lesser extent April have many significantly positive grid points spread approximately uniformly except in a large Southwest corner in April and a large Southeast corner in May. The significance for PROD in April

TABLE 2. Number of grid points for which $\hat{\eta}_{1,\mathrm{ti}}$ and $\hat{\eta}_{1,\mathrm{en}}$ are significant for PROD, CAPE, and SRH maxima for each month (first 12 rows); number of grid points for which $\hat{\eta}_{1,\mathrm{ti}}$ is significant for PROD, CAPE, and SRH maxima residuals (res) after accounting for the relation with ENSO (next six rows); and number of grid points for which $\hat{\eta}_{1,\mathrm{en}}$ is significant for PROD, CAPE, and SRH maxima residuals after accounting for the relation with time (bottom six rows). We have accounted for multiple testing using the BH procedure with the values of q displayed.

Variable	Covariate	q	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PROD	Time	0.05	7	0	1	41	36	0	0	36	2	0	0	22
	Time	0.2	40	0	4	313	203	81	13	148	23	0	0	98
	ENSO	0.05	0	58	10	0	0	1	0	0	0	0	0	1
	ENSO	0.2	1	172	26	0	3	3	0	0	0	0	0	1
CAPE	Time	0.05	37	13	28	109	60	89	18	55	4	0	30	1
	Time	0.2	92	37	73	268	273	206	75	133	35	40	134	16
	ENSO	0.05	15	0	0	0	0	2	2	0	0	0	1	1
	ENSO	0.2	27	11	21	0	0	3	16	14	0	1	6	13
SRH	Time	0.05	0	1	0	7	43	2	1	7	0	0	0	0
	Time	0.2	15	44	4	138	230	14	50	45	6	0	0	27
	ENSO	0.05	0	255	0	0	1	0	0	0	0	0	0	0
	ENSO	0.2	3	384	59	18	4	0	8	7	4	1	0	82
PROD res	Time	0.05	7	0	2	30	88	0	0	41	2	0	0	38
	Time	0.2	50	16	6	274	221	86	21	137	18	0	2	100
CAPE res	Time	0.05	35	20	15	87	96	89	25	59	9	0	19	2
	Time	0.2	88	46	51	219	267	223	91	139	54	41	120	29
SRH res	Time	0.05	0	0	0	7	38	2	1	7	0	0	0	0
	Time	0.2	20	1	6	126	241	7	46	41	1	0	0	60
PROD res	ENSO	0.05	1	66	8	0	0	1	0	0	0	0	0	7
	ENSO	0.2	1	178	26	0	49	3	0	0	0	0	0	33
CAPE res	ENSO	0.05	1	0	0	0	0	3	5	1	0	0	0	2
	ENSO	0.2	21	38	0	1	0	4	17	16	0	0	1	21
SRH res	ENSO	0.05	0	209	0	0	4	0	0	0	1	0	0	0
	ENSO	0.2	1	359	20	38	14	0	3	7	2	1	0	63

and May comes from both CAPE and SRH. Figures 6 and 7 suggest that the significant positive time trend in the riskiest part of the United States stems mainly from CAPE in April and from SRH in May. Overall, we found no specific driver for any specific season: CAPE seems to drive PROD in January, April, August, November, and December, whereas SRH seems to drive it in February, May, June, and September. For March, July, and October, there is no clear driver. Anyway, trying to relate the behavior of PROD maxima with that of CAPE and SRH maxima has its limitations. Indeed, the maximum of PROD may not equal the product of the square root of CAPE maximum and the maximum of SRH, as their maxima may not coincide.

We now comment on the effect of ENSO. For PROD, Table 2 reveals that many grid points exhibit a significant relation in February. Figure 8 indicates that $\hat{\eta}_{1,\mathrm{en}}$ is negative at those and that the main regions concerned are the Northeast, the South Central area, and the Northwest; we expect higher PROD maxima during La Niña years in these regions. The highest slope absolute value corresponds to a decrease of PROD maxima per unit of ENSO of about 10% of the corresponding basic level of PROD maximum.

There is no strikingly significant result for CAPE, although Allen et al. (2015) found ENSO signals in CAPE seasonal averages for winter and spring, not accounting for multiple testing.

For SRH, Fig. 8 shows that almost all grid points in the region exhibit significance in February, apart from a strip in the North and a tiny diagonal strip in the Southeast corner. The estimate $\hat{\eta}_{1,\mathrm{en}}$ is highly negative in most of the region but very positive in the extreme Southeast, with a very rapid change in sign, presumably due to proximity with the Gulf of Mexico. There is a significant negative relation in regions at risk of thunderstorms or large-scale storms, for which SRH plays an essential role. The risk of large impacts may increase during La Niña years. A relationship between seasonal averages of SRH and ENSO in winter was noticed by Allen et al. (2015). Finally, Fig. 8 suggests that CAPE contributes more than SRH to PROD in terms of significance, although the relation with ENSO is more pronounced for SRH than for CAPE.

We also consider the residuals of PROD, CAPE, and SRH maxima after accounting for ENSO or temporal effects. For instance, if we observe a time trend, the idea of considering the residuals after accounting for ENSO is to determine whether the time trend is explained by ENSO. This allows us to determine whether the time and ENSO effects are ''independent.''

In the case of PROD, Table 2 shows that removing ENSO does not much decrease the number of grid points exhibiting a significant time trend; there is a slight decrease for April but a small increase for some other months. Accounting for the time trend, on the other hand, can slightly increase the number

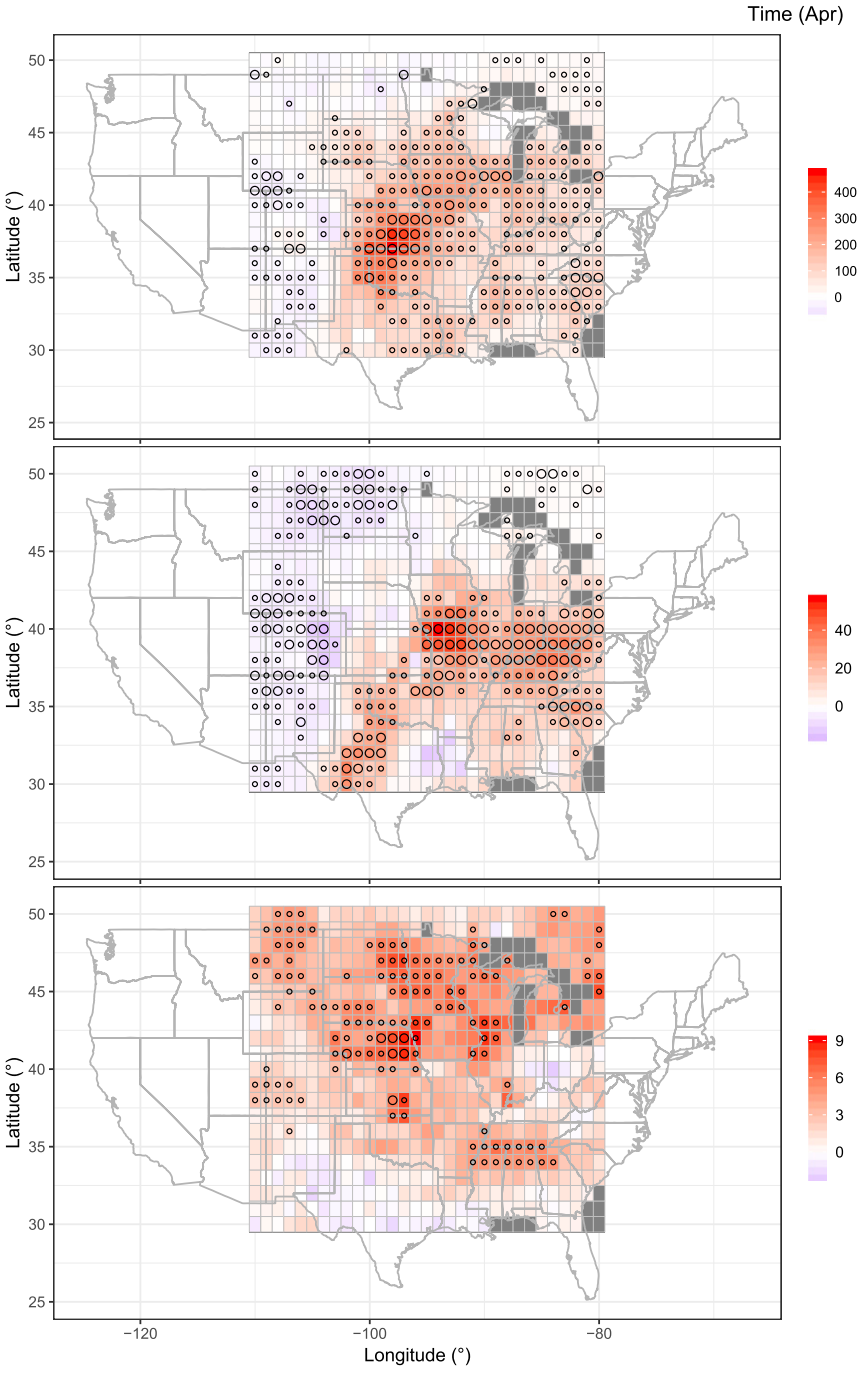

FIG. 6. Values and significance of the slope $\hat{\eta}_{1,ti}$ for (top) PROD ($m^3 s^{-3} yr^{-1}$), (middle) CAPE (J $kg^{-1} yr^{-1}$), and (bottom) SRH ($m^2 s^{-2} yr^{-1}$) maxima in April. Large and small circles indicate significance (after accounting for multiple testing using the BH procedure) at any level not lower than 5% and 20%, respectively. Dark gray corresponds to grid points for which no observations are available.

of grid points showing a significant relation with ENSO. For CAPE, removing ENSO decreases the number of grid points exhibiting a significant time trend for March, but there is a slight increase for other months, whereas accounting for time slightly decreases the number of grid points showing a significant relation with ENSO in January and March only, with a slight increase in other months. Regarding SRH, removing ENSO decreases the number of grid points exhibiting a

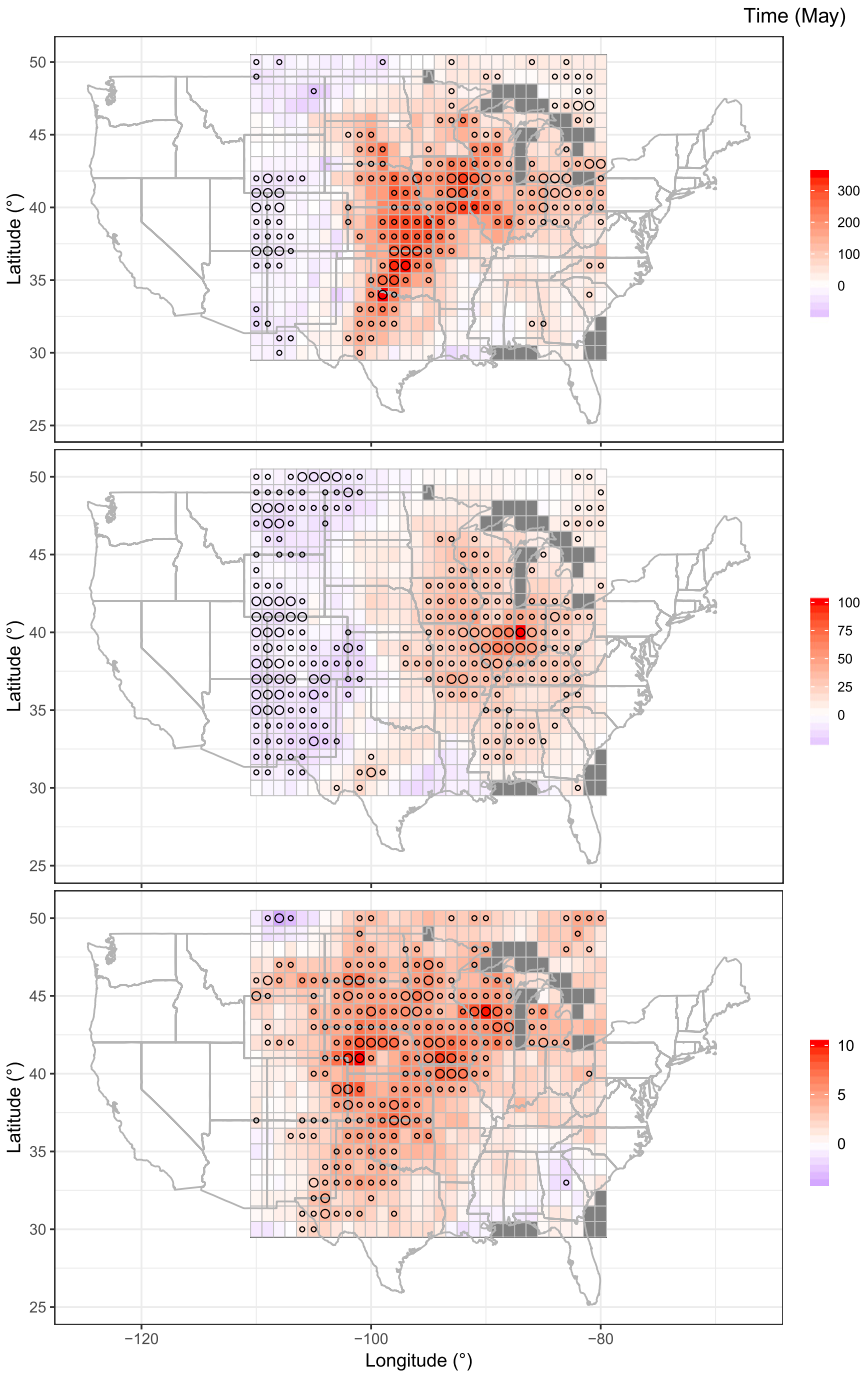

FIG. 7. As in Fig. 6, but for May.

significant time trend in February but there is little impact for other months. The conclusions are similar when accounting for the time trend and studying the ENSO effect. The maps of the residuals (not shown) indicate that when removing a covariate has little impact on the number of grid points at which the relation with the other covariate is significant, it has almost no impact on their positions either. In summary, the effects of time and ENSO appear ''independent'' except for CAPE in January and March and SRH in February.

5. Conclusions

This article quantifies the effects of time and ENSO on the distribution of monthly maxima of PROD, CAPE, and SRH, which are highly relevant to the risk of severe thunderstorms.

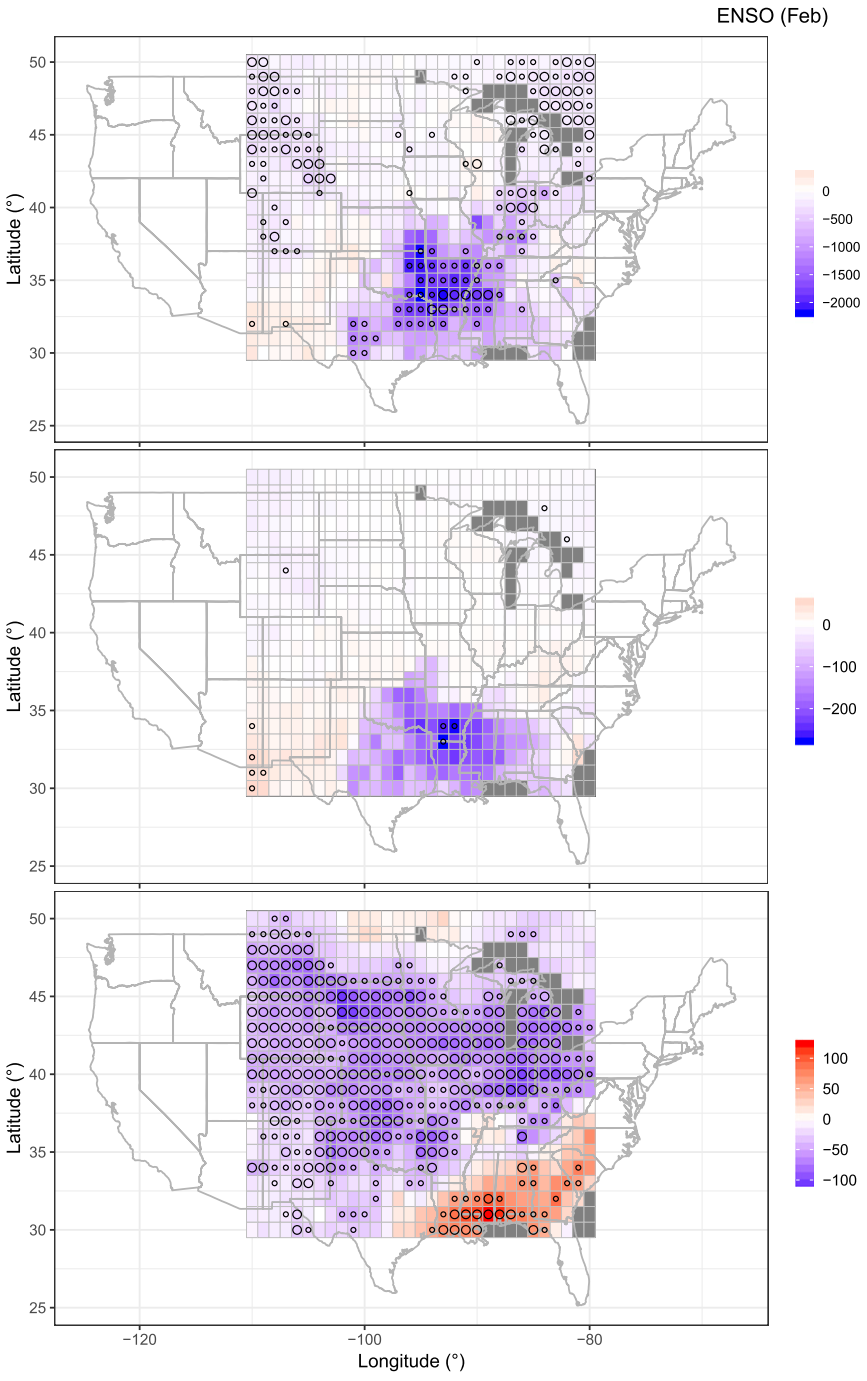

FIG. 8. As in Fig. 6, but for ENSO coefficient $\hat{\eta}_{1,\text{en}}$ for (top) PROD ($\text{m}^3\,\text{s}^{-3}\,^\circ\text{C}^{-1}$), (middle) CAPE ($\text{J kg}^{-1}\,^\circ\text{C}^{-1}$), and (bottom) SRH ($\text{m}^2\,\text{s}^{-2}\,^\circ\text{C}^{-1}$) maxima in February.

The use of the GEV distribution appears justified in our setting. After allowance for multiple testing we detect a significant time trend in the GEV location parameter for PROD maxima in April, May, and August, CAPE maxima in April, May, and June, and SRH maxima in April and May. The observed upward time trend for CAPE, although expected in a warming climate, has not been reported before. April and May are prominent for PROD, as severe thunderstorms are common at this period, and the corresponding trend is positive in parts of the United States where the risk is already high, which may have important consequences. We also find ENSO to be a good covariate in the GEV location parameter for PROD and SRH maxima in February. The corresponding relationship is negative over most of the region we consider, suggesting that the

risk of storm impacts in February increases during La Niña years. Our results differ from those of Heaton et al. (2011), Mannshardt and Gilleland (2013), and Gilleland et al. (2013), but are quite consistent with those obtained by Gensini and Brooks (2018), perhaps in part because these authors consider a period similar to ours, more recent than in the earlier studies.

We investigate the effects of time and ENSO on the marginal (at each grid point) extremal behavior of PROD, CAPE, and SRH. Quantifying the potential impacts of these covariates on the local spatial extremal dependence of these variables would also be useful for risk assessment. Modeling the extremal dependence between CAPE and SRH might also be informative.

Although concurrently high values of environments such as CAPE and SRH are conducive to severe weather, they do not guarantee that severe thunderstorms will occur. The degree to which changes in environmental characteristics result in changes in thunderstorm properties is also uncertain (Hoogewind et al. 2017). Hence, an interesting issue is the precise implication of an increase of PROD (or SRH) maxima in terms of risk. PROD can be seen as a proxy for the probability of severe thunderstorm occurrence, so it is natural to think that PROD maxima may be aligned with the maximum number of severe thunderstorms daily. If so, then those days with the highest values of PROD would tend to also have the most severe thunderstorm impacts. Better insight into this potential relationship would be valuable.

Acknowledgments. We are grateful to the reviewers for their helpful comments. The work was supported by the Swiss National Science Foundation (project 200021_178824).

Data availability statement. NARR data were downloaded from the Research Data Archive (RDA) at the National Center for Atmospheric Research (NCAR) Computational and Information Systems Laboratory (http://rda.ucar.edu/datasets/ds608.0/). NCAR is sponsored by the National Science Foundation. The ERSSTv5 dataset is from the NOAA Climate Prediction Center (https://www.cpc.ncep.noaa.gov/data/indices/ersst5.nino.mth.81-10.ascii).

REFERENCES

Agee, E., J. Larson, S. Childs, and A. Marmo, 2016: Spatial redistribution of U.S. tornado activity between 1954 and 2013. *J. Appl. Meteor. Climatol.*, **55**, 1681–1697, https://doi.org/10.1175/JAMC-D-15-0342.1.

Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10**, 3, https://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/149.

——, ——, and A. H. Sobel, 2015: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nat. Geosci.*, **8**, 278–283, https://doi.org/10.1038/ngeo2385.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57B**, 289–300, https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.

——, and D. Yekutieli, 2001: The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, **29**, 1165–1188, https://projecteuclid.org/euclid.aos/1013699998.

Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129–138, https://doi.org/10.1016/j.atmosres.2012.04.002.

——, J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94, https://doi.org/10.1016/S0169-8095(03)00045-0.

——, G. W. Carbin, and P. T. Marsh, 2014: Increased variability of tornado occurrence in the United States. *Science*, **346**, 349–352, https://doi.org/10.1126/science.1257460.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 209 pp., https://doi.org/10.1007/978-1-4471-3675-0.

Corfidi, S., 2017: Forecasting severe convective storms. *Oxford Research Encyclopedia of Climate Science*, Oxford University Press, https://doi.org/10.1093/acrefore/9780190228620.013.59.

Davison, A. C., and R. Huser, 2015: Statistics of extremes. *Annu. Rev. Stat. Appl.*, **2**, 203–235, https://doi.org/10.1146/annurev-statistics-010814-020133.

Del Genio, A. D., M.-S. Yao, and J. Jonas, 2007: Will moist convection be stronger in a warmer climate? *Geophys. Res. Lett.*, **34**, L16703, https://doi.org/10.1029/2007GL030525.

DelSole, T., and X. Yang, 2011: Field significance of regression patterns. *J. Climate*, **24**, 5094–5107, https://doi.org/10.1175/2011JCLI4105.1.

Diffenbaugh, N. S., M. Scherer, and R. J. Trapp, 2013: Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proc. Natl. Acad. Sci. USA*, **110**, 16 361–16 366, https://doi.org/10.1073/pnas.1307758110.

Doswell, C. A., III, 2015: Mesoscale meteorology: Severe storms. *Encyclopedia of Atmospheric Sciences*, 2nd ed., Elsevier, 361–368, https://doi.org/10.1016/B978-0-12-382225-3.00366-2.

——, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Edwards, R., J. T. Allen, and G. W. Carbin, 2018: Reliability and climatological impacts of convective wind estimations. *J. Appl. Meteor. Climatol.*, **57**, 1825–1845, https://doi.org/10.1175/JAMC-D-17-0306.1.

Elsner, J. B., S. C. Elsner, and T. H. Jagger, 2015: The increasing efficiency of tornado days in the United States. *Climate Dyn.*, **45**, 651–659, https://doi.org/10.1007/s00382-014-2277-3.

Embrechts, P., T. Mikosch, and C. Klüppelberg, 1997: *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, 648 pp., https://doi.org/10.1007/978-3-642-33483-2.

Gensini, V. A., and H. E. Brooks, 2018: Spatial trends in United States tornado frequency. *npj Climate Atmos. Sci.*, **1**, 38, https://doi.org/10.1038/S41612-018-0048-2.

Gilleland, E., M. Pocernich, H. E. Brooks, B. G. Brown, and P. Marsh, 2008: Large-scale indicators for severe weather. UCAR Doc., 16 pp., https://ral.ucar.edu/~ericg/GillelandEtAl2008.pdf.

——, B. G. Brown, and C. M. Ammann, 2013: Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, **24**, 418–432, https://doi.org/10.1002/env.2234.

Heaton, M. J., M. Katzfuss, S. Ramachandar, K. Pedings, E. Gilleland, E. Mannshardt-Shamseldin, and R. L. Smith, 2011: Spatio-temporal models for large-scale indicators of extreme weather. *Environmetrics*, **22**, 294–303, https://doi.org/10.1002/env.1050.

Heffernan, J. E., and J. A. Tawn, 2004: A conditional approach for multivariate extreme values (with discussion). *J. Roy. Stat. Soc.*, **66B**, 497–546, https://doi.org/10.1111/j.1467-9868.2004.02050.x.

Hoogewind, K. A., M. E. Baldwin, and R. J. Trapp, 2017: The impact of climate change on hazardous convective weather in the United States: Insight from high-resolution dynamical downscaling. *J. Climate*, **30**, 10 081–10 100, https://doi.org/10.1175/JCLI-D-16-0885.1.

Johns, R. H., J. M. Davies, and P. W. Leftwich, 1993: Some wind and instability parameters associated with strong and violent tornadoes: 2. Variations in the combinations of wind and instability parameters. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, Vol. 79, Amer. Geophys. Union, 583–590, https://doi.org/10.1029/GM079p0583.

Lepore, C., D. Veneziano, and A. Molini, 2015: Temperature and CAPE dependence of rainfall extremes in the eastern United States. *Geophys. Res. Lett.*, **42**, 74–83, https://doi.org/10.1002/2014GL062247.

——, M. K. Tippett, and J. T. Allen, 2017: ENSO-based probabilistic forecasts of March–May US tornado and hail activity. *Geophys. Res. Lett.*, **44**, 9093–9101, https://doi.org/10.1002/2017GL074781.

Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59, https://doi.org/10.1175/1520-0493(1983)111<0046:SFSAID>2.0.CO;2.

Mannshardt, E., and E. Gilleland, 2013: Extremes of severe storm environments under a changing climate. *Amer. J. Climate Change*, **2**, 47–61, https://doi.org/10.4236/ajcc.2013.23A005.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164, https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2.

Saville, G., 2019: Managing severe thunderstorm risk: Looking beyond the weather forecast. Willis Towers Watson Blog Post, accessed 26 February 2020, https://www.willistowerswatson.com/en-GB/Insights/2017/04/managing-severe-thunderstorm-risk-looking-beyond-the-weather-forecast.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Tippett, M. K., C. Lepore, and J. E. Cohen, 2016: More tornadoes in the most extreme U.S. tornado outbreaks. *Science*, **354**, 1419–1423, https://doi.org/10.1126/science.aah7393.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from squall lines and bow echoes. Part I: Climatological distribution. *Wea. Forecasting*, **20**, 23–34, https://doi.org/10.1175/WAF-835.1.

——, N. S. Diffenbaugh, and A. Gluhovsky, 2009: Transient response of severe thunderstorm forcing to elevated greenhouse gas concentrations. *Geophys.Res. Lett.*, **36**, L01703, https://doi.org/10.1029/2008GL036203.

Van Klooster, S. L., and P. J. Roebber, 2009: Surface-based convective potential in the contiguous United States in a business-as-usual future climate. *J. Climate*, **22**, 3317–3330, https://doi.org/10.1175/2009JCLI2697.1.

Ventura, V., C. J. Paciorek, and J. S. Risbey, 2004: Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *J. Climate*, **17**, 4343–4356, https://doi.org/10.1175/3199.1.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93, https://doi.org/10.1175/WAF910.1.

Yekutieli, D., and Y. Benjamini, 1999: Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plann. Inference*, **82**, 171–196, https://doi.org/10.1016/S0378-3758(99)00041-5.