

Environmental Covariate Representation of Seasonal U.S. Tornado Frequency

VITTORIO A. GENSI NI

Department of Geographic and Atmospheric Sciences, Northern Illinois University, DeKalb, Illinois

LELYS BRAVO DE GUENNI

Division of Statistics, Northern Illinois University, DeKalb, Illinois

(Manuscript received 21 November 2018, in final form 25 March 2019)

ABSTRACT

The significant tornado parameter is a widely used meteorological composite index that combines several variables known to favor tornadic supercell thunderstorms. This research examines the spatial relationship between U.S. tornado frequency and the significant tornado parameter (the predictor covariate) across four seasons in order to establish a spatial–statistical model that explains significant amounts of variance in tornado occurrence (the predictand). U.S. tornadoes are highly dependent on the significant tornado parameter in a climatological sense. The strength of this dependence is seasonal, with greatest dependence found during December–February and least dependence during June–August. Additionally, the strength of this dependence has not changed significantly through the 39-yr study period (1979–2017). Results herein represent an important step forward for the creation of a predictive spatial–statistical model to aid in tornado prediction at seasonal time scales.

1. Introduction

It is well known that atmospheric conditions supportive of supercell thunderstorms are generally characterized by low static stability, high surface water vapor mixing ratios, a lifting mechanism (e.g., a boundary or orographic feature) to get air parcels to the level of free convection, and adequate deep-layer vertical wind shear (McNulty 1985; Johns and Doswell 1992; Rasmussen and Blanchard 1998; Rasmussen 2003). In addition to these, tornadic supercell thunderstorms also require minimal convective inhibition, relatively low lifting condensation level heights, and sufficient low-level storm-relative helicity (Brooks et al. 2003; Thompson et al. 2003). To concentrate these ingredients into a simple diagnostic metric, a composite index known as the significant tornado parameter (STP) was developed to statistically discriminate significant tornado [(E)F2+] from nontornadic environments (Thompson et al. 2003, 2004). STP was designed as a diagnostic environmental discriminator for significantly tornadic versus nontornadic supercells, but it is now becoming more widely used in various tornado research and forecasting applications (Potvin et al. 2010;

Grams et al. 2012; Thompson et al. 2012; Gensini and Marinaro 2016; Allen et al. 2018; Gensini and Brooks 2018; Molina et al. 2018).

The main goal of this work is to statistically model tornado frequency as a function of climatological aspects of STP. We examine the dependence of tornado frequency on the STP and quantify the tornado rate of occurrence as a function of this covariate. Initial tornado report locations for different seasons are considered as a realization of a spatial point pattern with density $\lambda(s)$, defined as the number of touchdowns per unit area for location s . If the process is homogeneous, $\lambda(s)$ would be constant across the whole study domain and no preference would be found for reports at any given location. If the process is independent, information about tornado report frequency in one region would not offer information about tornado report frequency in another region. Both homogeneity and independence would characterize a completely spatial random process (CSR). Initial exploratory data analysis indicates the lack of complete spatial randomness of tornado frequency across the United States with a nonhomogeneous density, suggesting a dependence of the density process $\lambda(s)$ on covariate predictor STP, with varying strengths of dependence. This dependence strength is quantified seasonally and

Corresponding author: Vittorio A. Gensini, vgensi@niu.edu

interannually during the period of 1979–2017. Finally, the tornado density process is modeled as a function of the STP for different seasons to quantify the density factor modifying baseline tornado reporting rate without considering the covariate effect.

The paper is organized as follows. [Section 2](#) provides a background description of the problem and previous attempts to apply statistical models of the convective environment to tornado reports. [Section 3](#) describes data used in this analysis and the selected methodologies/statistical tests to quantify tornado report dependence on climatological STP metrics. [Section 4](#) provides statistical evidence of the dependence of tornado reports on the STP covariate for two extreme, but opposite, report frequency years and quantification of the strength of this dependence as a function of season and year. In addition, we include estimations and parameter interpretations for the proposed density process models. Finally, a summary of the findings is provided in [section 5](#).

2. Background

A few recent studies have examined and modeled various aspects of the U.S. tornado report climatology. These studies investigate various temporal and spatial scales with the overarching goal of modeling tornado frequency as a function of some statistically significant covariate. At the climate scale, tornado report counts have been modeled using a negative binomial distribution with parameters depending on population density and elevation roughness to account for spatial variability of tornado activity ([Elsner et al. 2016](#)). This roughly county-level spatial statistical model focused on the fixed effects of population density and elevation roughness for a limited spatial area over the Great Plains of the United States. Results were most significant over the state of Kansas using this approach, and the model was found to best serve as a potential first guess for understanding the local climatological distributions of tornadoes, helping to confirm earlier work showing the importance of topography on U.S. tornado frequency ([Karpman et al. 2013](#)). However, these works did not attempt to address potential variability in tornado frequency as a function of atmospheric environmental parameters.

At the monthly scale, Poisson regression has also demonstrated skill in explaining the variance associated with monthly tornado frequency by using an observation-based covariate of monthly averaged atmospheric parameters ([Tippett et al. 2012, 2014](#)). Specifically, these studies used environmental measures of storm-relative helicity and convective precipitation to explain variance in U.S. tornado frequency by month. Results were

statistically significant for most months and proved skillful in representing the interannual variability of monthly tornado reports at the spatial scale of NOAA climate regions. Additional studies have demonstrated promising results using hierarchical Bayesian modeling using population ([Anderson et al. 2007](#)) and ENSO ([Wikle and Anderson 2003](#)), but the work herein is more closely related to studies such as [Tippett et al. \(2012\)](#), [Tippett et al. \(2014\)](#), and [Cheng et al. \(2016\)](#) that all utilize environmental variables pertinent to tornado formation to explain spatiotemporal variability. We differ from these works by using an inhomogeneous Poisson process to create a nonparametric modeled estimation of gridpoint seasonal tornado frequency using a well-documented meteorological composite index.

With a myriad of potential environmental metrics to choose from, this study focuses on a composite index parameter (STP) that is skillful in discriminating significantly tornadic and nontornadic supercells in a diagnostic setting ([Thompson et al. 2003, 2004](#)). At the prognostic time and space scales associated with nowcasting, STP has also proved useful in the development of conditional probabilities for significant tornadoes using logistic regression given some a priori knowledge of the convective mode ([Togstad et al. 2011](#)). In addition, recent research has focused on the use of STP as a climatological indicator of tornado frequency ([Gensini and Brooks 2018](#)), and the use of STP to explain spatial patterns of tornado frequency associated with various atmospheric teleconnections ([Gensini and Marinaro 2016](#); [Allen et al. 2018](#); [Molina et al. 2018](#)). The former study noted that the annual sum of the daily max STP value explained 45% of the interannual variance in annual U.S. tornado report counts, but was dependent on month. Notably, the month of August was found to have the lowest explanatory capability of any month, likely associated with the challenge of forecasting tornadoes during the boreal summer months ([Hart and Cohen 2016](#)). This portion of the seasonal cycle represents an ongoing challenge in the climatological modeling, analysis, and prediction of U.S. tornadoes.

This research has some caveats because of the use of an atmospheric composite parameter ([Doswell and Schultz 2006](#)). For example, tornadoes can occur with STP values of 0, and the STP alone does not uniquely define which aspects of the forecast parameter space (e.g., instability, vertical wind shear) are most favorable for the potential generation of tornadic supercell thunderstorms. Environments characteristic of low instability, but adequate vertical wind shear (and thus low STP values), pose a particular forecast challenge ([Sherburn and Parker 2014](#); [Sherburn et al. 2016](#)). It is also well known that environmental conditions favorable for

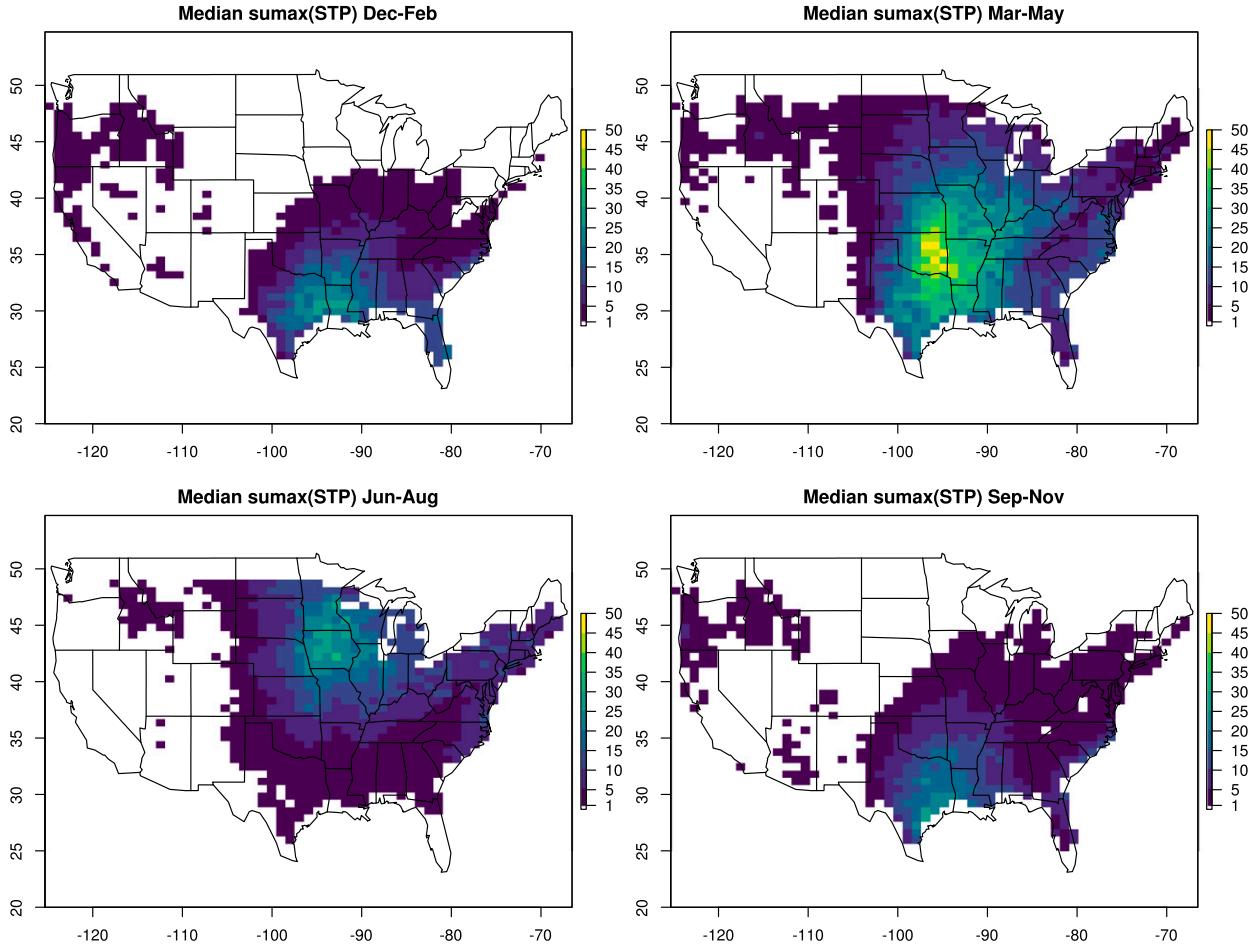


FIG. 1. Quarterly median values for the period of 1979–2017 of sumax(STP).

tornadoes do not necessarily indicate whether or not tornadoes will occur. For example, the lack of a dynamic lifting mechanism is associated with a known bias when examining convective environments and not accounting for a process to commence thunderstorm development (i.e., convective initiation). This conservative bias is shown at the monthly scale for STP and tornado report standardized anomalies during the period of 1979–2017 (Gensini and Brooks 2018). Further discussion on the topic of environmental bias is provided in Gensini and Ashley (2011). Despite these caveats, environments form the basis for severe convective storm forecasting through use of an ingredients-based methodology (Johns and Doswell 1992).

3. Methodology

a. Tornado reports

Tornado observations were obtained from the National Centers for Environmental Information (NCEI) Storm Data (Schaefer and Edwards 1999) archive for the period

of 1979–2017. Reports of tornadoes are sensitive to spatial and nonmeteorological biases relating to variations in population, subjectivity in (E)F rating, and other discontinuities in the record (Verbout et al. 2006). Climatologically speaking, such errors are much less likely influence the stratified results herein given the large sample sizes ($N > 40\,000$).

b. NARR

STP was calculated from the North American Regional Reanalysis (NARR; Mesinger et al. 2006) 3-hourly 3D isobaric (29 pressure levels spanning 1000–100 hPa) data for all CONUS land points on the native 32-km Lambert conformal grid for the period of 1979–2017. STP terms include surface-based convective available potential energy (sbCAPE), surface-based lifted condensation level (sbLCL), 0–1-km storm-relative helicity (SRH), and 0–6-km bulk vertical wind shear (BWS). All surface-based parcels are calculated using the virtual temperature correction (Doswell and Rasmussen 1994). The 0–1-km SRH was calculated using the Bunkers storm motion

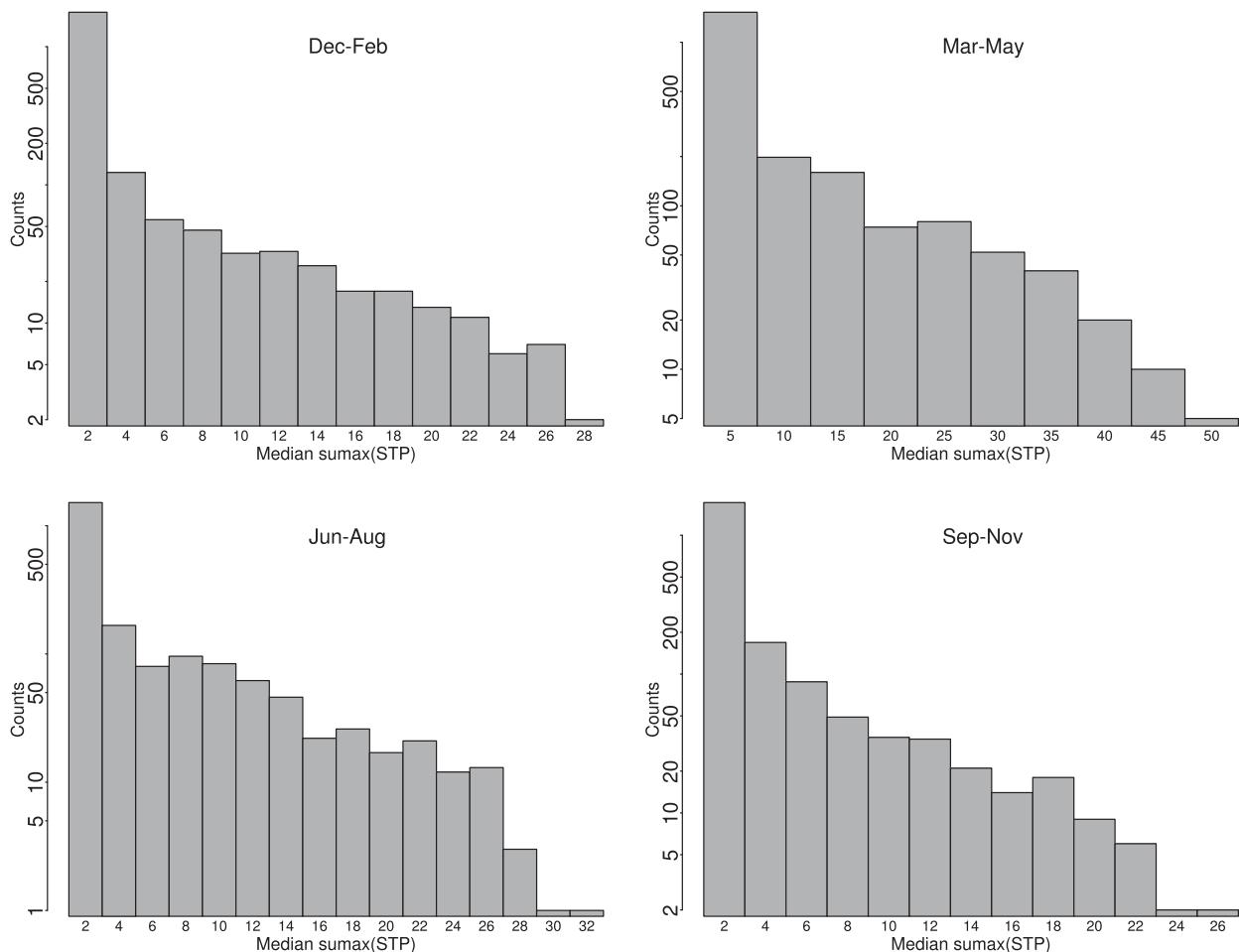


FIG. 2. Quarterly histograms for the median values of sumax(STP) for the period of 1979–2017.

vector (Bunkers et al. 2000), and 0–6-km BWS was calculated by vertically interpolating NARR isobaric data to AGL height coordinates and finding the magnitude of the shear vector between 10 m and 6 km. STP is a unitless composite index, and is calculated here following the Thompson et al. (2003) fixed-layer calculation:

$$\text{STP} = \frac{\text{sbCAPE}}{1500 \text{ J kg}^{-1}} \times \frac{2000 \text{ m} - \text{sbLCL}}{1000 \text{ m}} \\ \times \frac{0\text{--}1\text{-km SRH}}{150 \text{ m}^2 \text{ s}^{-2}} \times \frac{0\text{--}6\text{-km BWS}}{20 \text{ m s}^{-1}}. \quad (1)$$

The sbLCL term was set to 0 if $>2000 \text{ m}$ and 1 if $<1000 \text{ m}$; 0–6-km BWS was set to 0 if $<12.5 \text{ m s}^{-1}$ and 1.5 if $>30 \text{ m s}^{-1}$. STP variable cutoff values are related to thresholds identified by previous studies (e.g., Brooks et al. 1994; Rasmussen and Blanchard 1998; Thompson et al. 2003), which aid in discriminating between significantly tornadic and nontornadic supercells. In addition, STP was set to 0 in the presence of

surface-based convective inhibition $\leq -50 \text{ J kg}^{-1}$ in an attempt to only analyze environments with a limited capping inversion. NARR has been shown to underestimate the strength of the capping inversion in portions of the southern Great Plains, so caution should be used when interpreting relationships between environments and reports in these regions (Gensini and Ashley 2011; Gensini et al. 2014).

c. Methods

1) STP SUMMARY STATISTICS

Following results from Gensini and Brooks (2018), the daily (1200–1200 UTC) maximum STP value was used to examine aspects of climatological tornado frequency in this study. The quarterly (December–February, March–May, June–August, September–November) sum of the daily maximum STP value, herein defined as sumax(STP), was considered as a representative STP summary statistic. To represent the spatial and seasonal

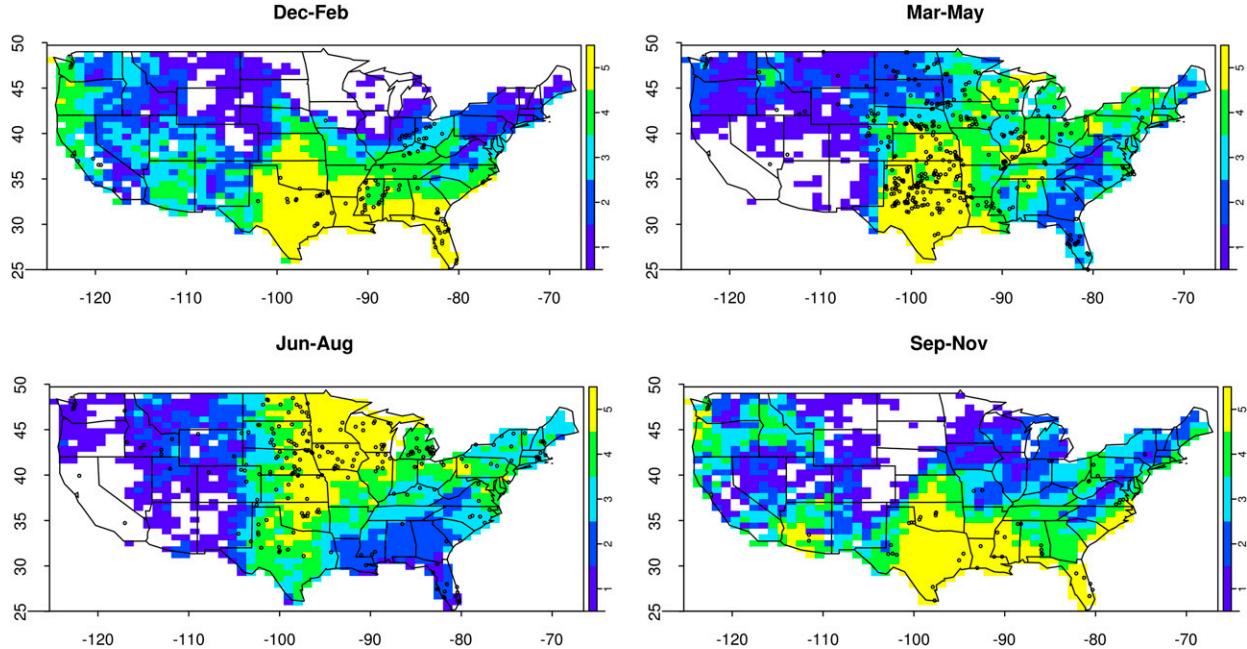


FIG. 3. For year 1987 quarterly tornado reports (circles) with sumax(STP) percentile groups defined by probability intervals [0.5, 0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 0.9), and [0.9, 1.0], colored and labeled as 1, 2, 3, 4, and 5, respectively.

patterns of the STP summary statistic, the median of all quarterly values (39-yr record) was calculated for each NARR CONUS grid cell. The well documented seasonal cycle of CONUS tornado frequency is visually apparent in the analysis of quarterly sumax(STP) median values (Fig. 1). Histograms of the quarterly median

sumax(STP) exhibit a logarithmic distribution, positive skew, and largest values during March–May (Fig. 2).

2) TORNADO REPORTS

Latitude and longitude of tornado reports (initial starting location) were aggregated by quarter and year.

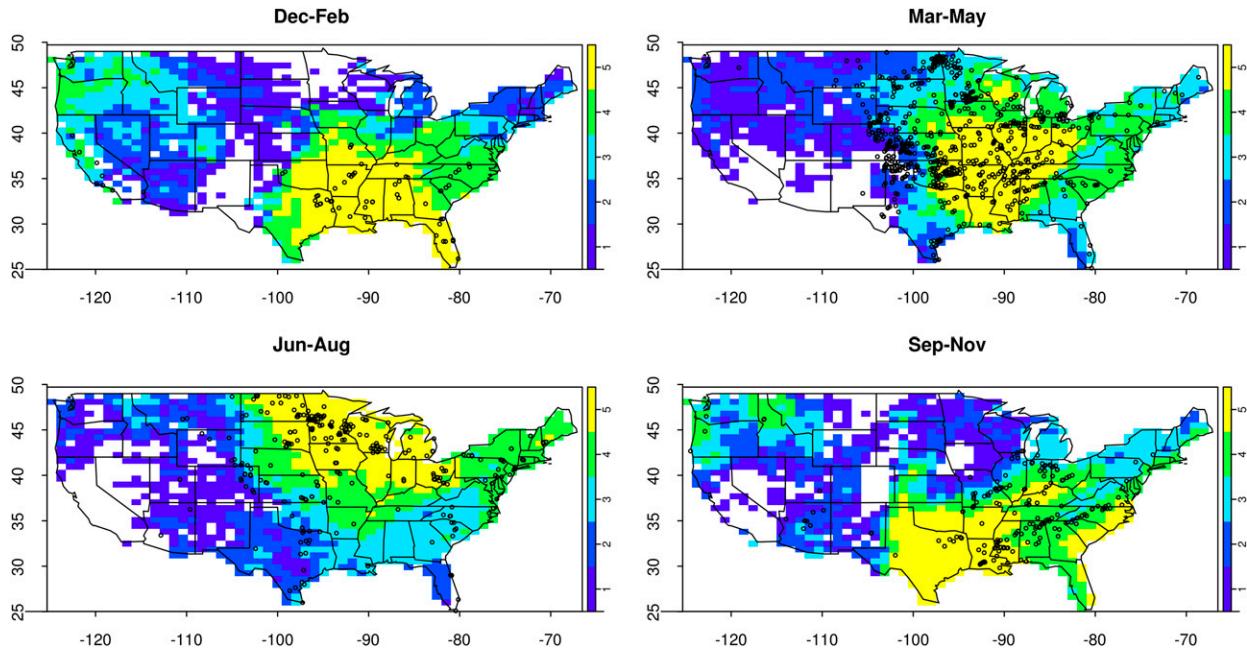


FIG. 4. As in Fig. 3, but for year 2011.

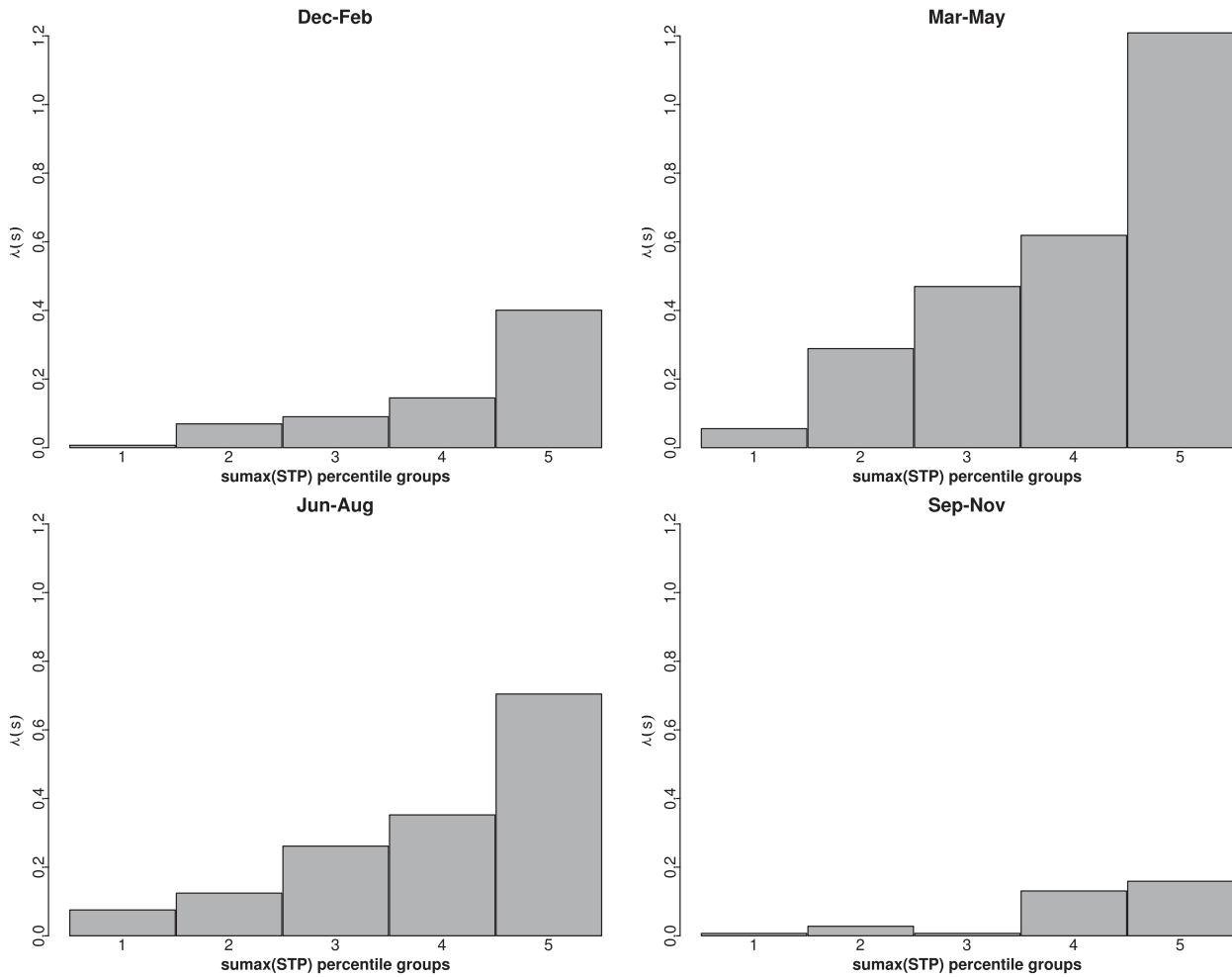


FIG. 5. For year 1987 quarterly tornado report density vs sumax(STP) percentile groups defined by probability intervals [0.5, 0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 0.9), and [0.9, 1.0] labeled as 1, 2, 3, 4, and 5 respectively.

Data were converted to a point pattern data format using the R library spatstat ppp function (Baddeley et al. 2016). Several tools are available within the spatstat library to investigate spatial properties of a point pattern, and in particular, to examine whether the hypotheses of homogeneity and complete spatial randomness are plausible assumptions. In addition, to investigate these assumptions, there are a number of methods to quantify the dependence of a point pattern on a covariate given by a spatial raster field. These methods, and their relevance to this study, are described in the following subsection.

3) TESTING DEPENDENCE ON A COVARIATE

Different test statistics were calculated to demonstrate the dependence of tornado reports on the values of sumax(STP). In point process theory, process density refers to “process intensity.” We use the term process density to avoid confusion with strength of the tornado. For this demonstration, and because of the need for

brevity, we selected 2 years of data: one with the highest observed quarterly reports (highest process density); and another year with the lowest observed frequency of quarterly reports (lowest process density). The selected years were 2011 and 1987, respectively. Highest report frequencies occurred in March–May 2011. Year 1987 had its lowest observed quarterly report during December–February. The following test statistics were applied to all quarters for these two years to examine dependence in two vastly different annual cycles of tornado frequency.

(i) χ^2 test on quadrats defined by a covariate

A formal dependence test of the density process on a covariate was performed by using a χ^2 test for quadrat counts. The quadrats are regions defined according to the values of the covariate sumax(STP). These regions can be of the same or different sizes. The null hypothesis is that the point density process is a homogeneous point pattern process with a Poisson distribution (completely

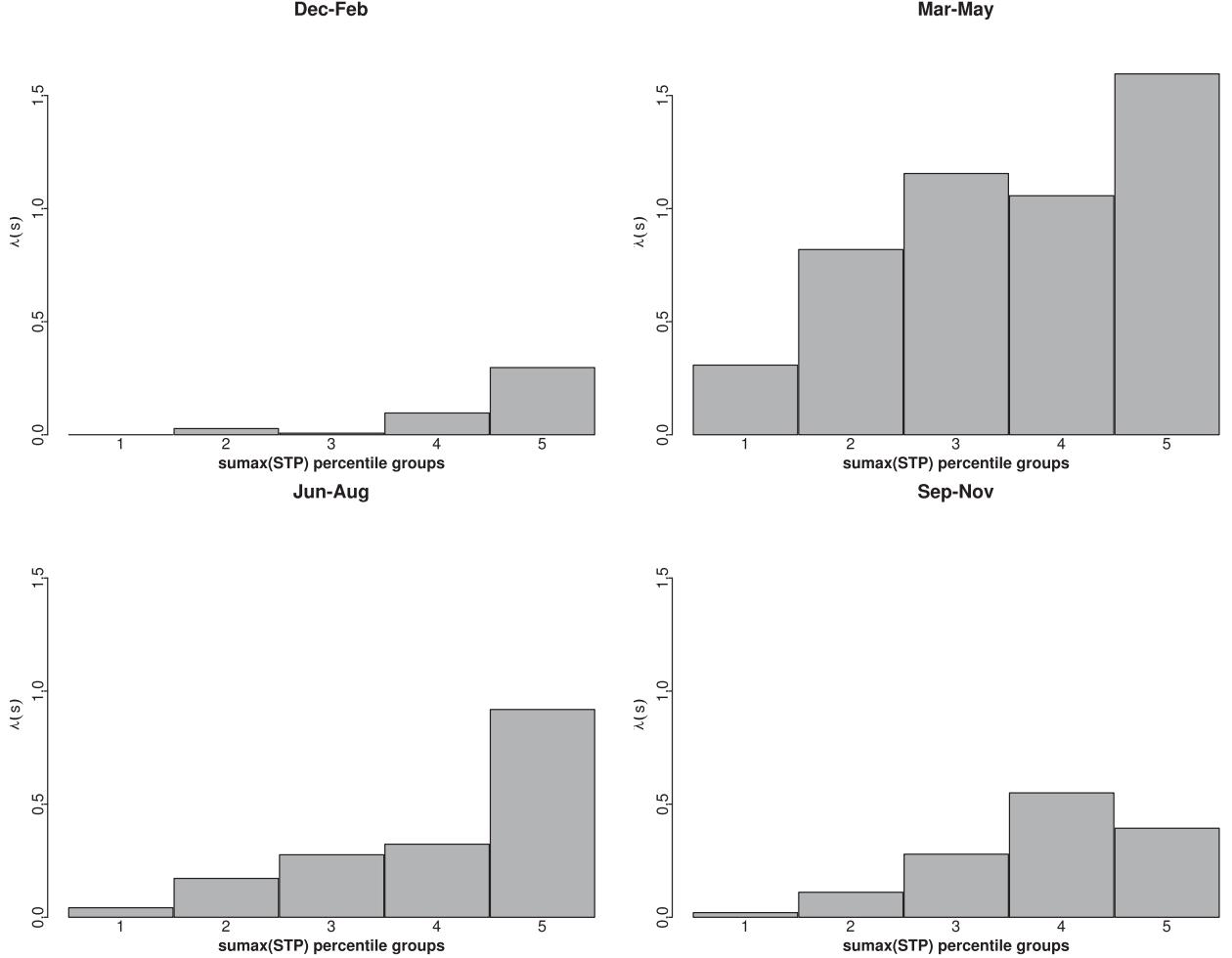


FIG. 6. As in Fig. 5, but for year 2011. Note the different y-axis scale here vs Fig. 5.

spatially random assumption), and the theoretical counts are compared with the observed counts in the same fashion as using a χ^2 test for independence. This test is often used when comparing observed and expected frequencies in climatological studies. Because of the high proportion of observed zeroes in the covariate, quadrats were created according to the STP quantiles defined by the following limiting probability values: (0.5, 0.6, 0.7, 0.8, 0.9, 1.0). Quadrat counts were compared with theoretical counts under the assumption of a Poisson process with quadrat counts proportional to quadrat area. The chi-squared test statistic is calculated as

$$\chi^2 = \sum_j \frac{(\text{observed} - \text{expected})^2}{\text{expected}}, \quad (2)$$

where the sum is over the total number of quadrats. If the homogeneity assumption is rejected, tornado reports are not homogeneous across the different quadrats.

(ii) Anderson–Darling test based on the exact values of the covariate

If the point pattern is completely random and independent of a covariate, the values of the covariate at the report points, $z_i = \mathbf{Z}(x_i)$, would be a random sample of the covariate field. The cumulative distribution functions (CDF) of the observed values of sumax(STP), z_i , at the report points $F(z)$ is compared to the CDF of the covariate at all spatial locations $F_0(Z)$. The Anderson–Darling test statistic is calculated as a weighted average in the form

$$A = n \int_{-\infty}^{\infty} \frac{[\hat{F}(z) - \hat{F}_0(z)]^2}{F_0(z)[1 - F_0(z)]} dF_0(z). \quad (3)$$

The empirical CDF for $F(z)$ is estimated as

$$\hat{F}(z) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}\{z_i \leq z\}, \quad (4)$$

where \mathbf{I} is the indicator function. The CDF $F_0(z)$ is estimated as

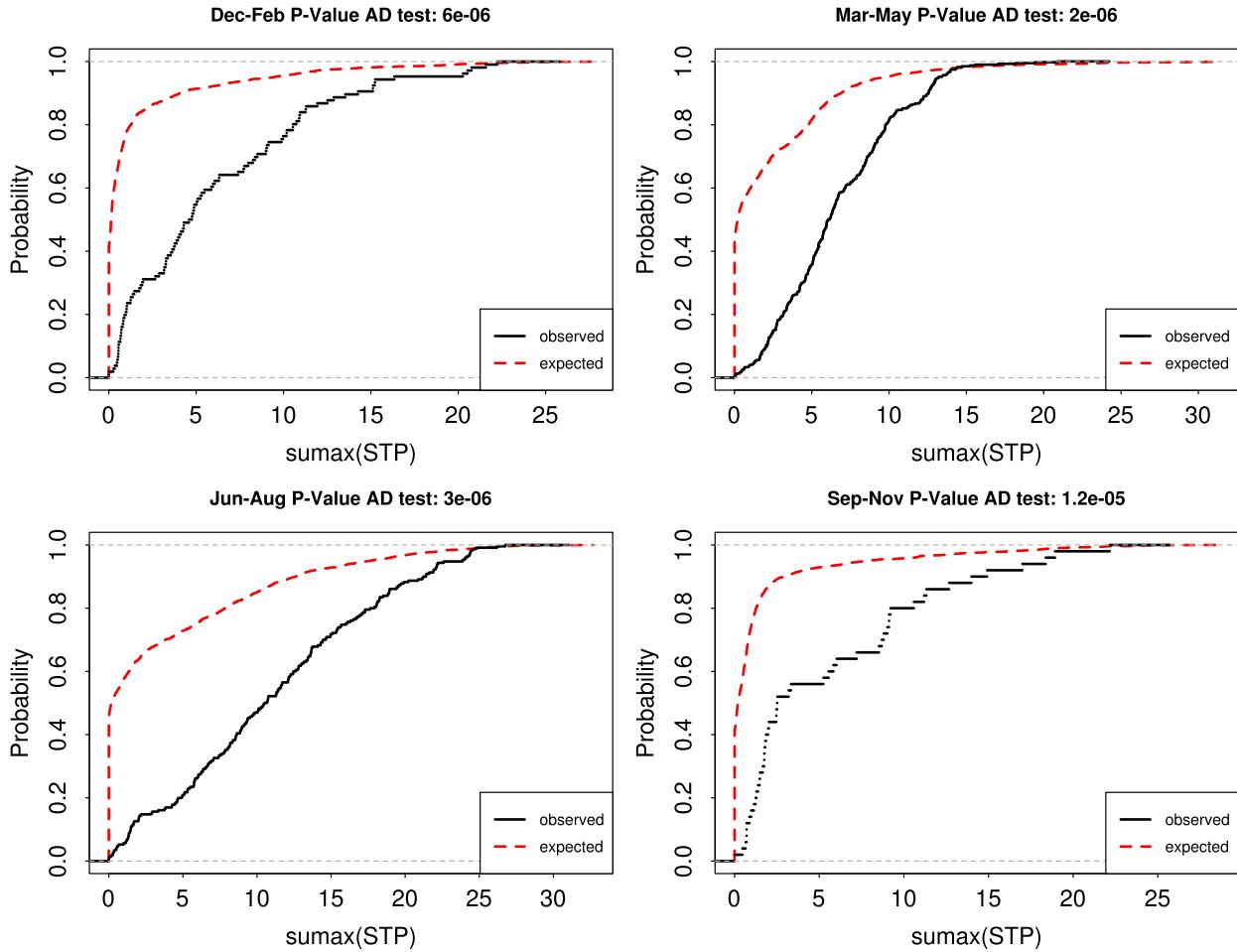


FIG. 7. Anderson–Darling test results for 1987. Observed CDF of sumax(STP) at all locations (black solid line). Expected CDF of sumax(STP) at tornado report locations (red dashed line).

$$F_0(z) \approx \frac{\text{No. of } \{ \text{pixels } u : Z(u) \leq z \}}{\text{No. of pixels}}. \quad (5)$$

In short, the A statistic measures the discrepancy between the two CDFs, where the null hypothesis is $H_0: F_0 = F$, and the alternative is $H_1: F_0 \neq F$.

(ii) Berman test for CSR

The Berman test represents another method to test the dependence of a point pattern on a covariate Z . The $Z2$ Berman test was used in this context to test tornado report dependence on the covariate sumax(STP). Let the observed values of the covariate at the report locations be represented as z_i . These values are transformed to $F_0(z_i) = u_i$, which is the spatial CDF of the covariate Z . Under the null hypothesis of independence, the values of u_i have a uniform distribution from $[0, 1]$, with mean 0.5 and standard deviation $1/\sqrt{12}$. The test statistic $Z2$ is the standardized sum of the transformed values u_i , where

$$Z2 = \sqrt{\frac{12}{n}} \left(\sum_i u_i - 0.5 \right); \quad (6)$$

$Z2$ has an asymptotic normal distribution under the null hypothesis of independence of the point pattern on the covariate.

4) STRENGTH OF THE DEPENDENCE ON A COVARIATE

Besides testing the dependence of a point pattern on a covariate, it is also important to measure the strength of this dependence. Even if one rejects the null hypothesis of independence for a given level of significance, it is still plausible that the dependence of the density process on the covariate might be rather weak. To provide a dependence quantification, it is common to measure the area under the curve (AUC) of a receiving operating characteristic (ROC) curve. A ROC curve enables the comparison of covariate CDF values $\hat{F}(z)$ evaluated at the report locations $Z(x_i)$,

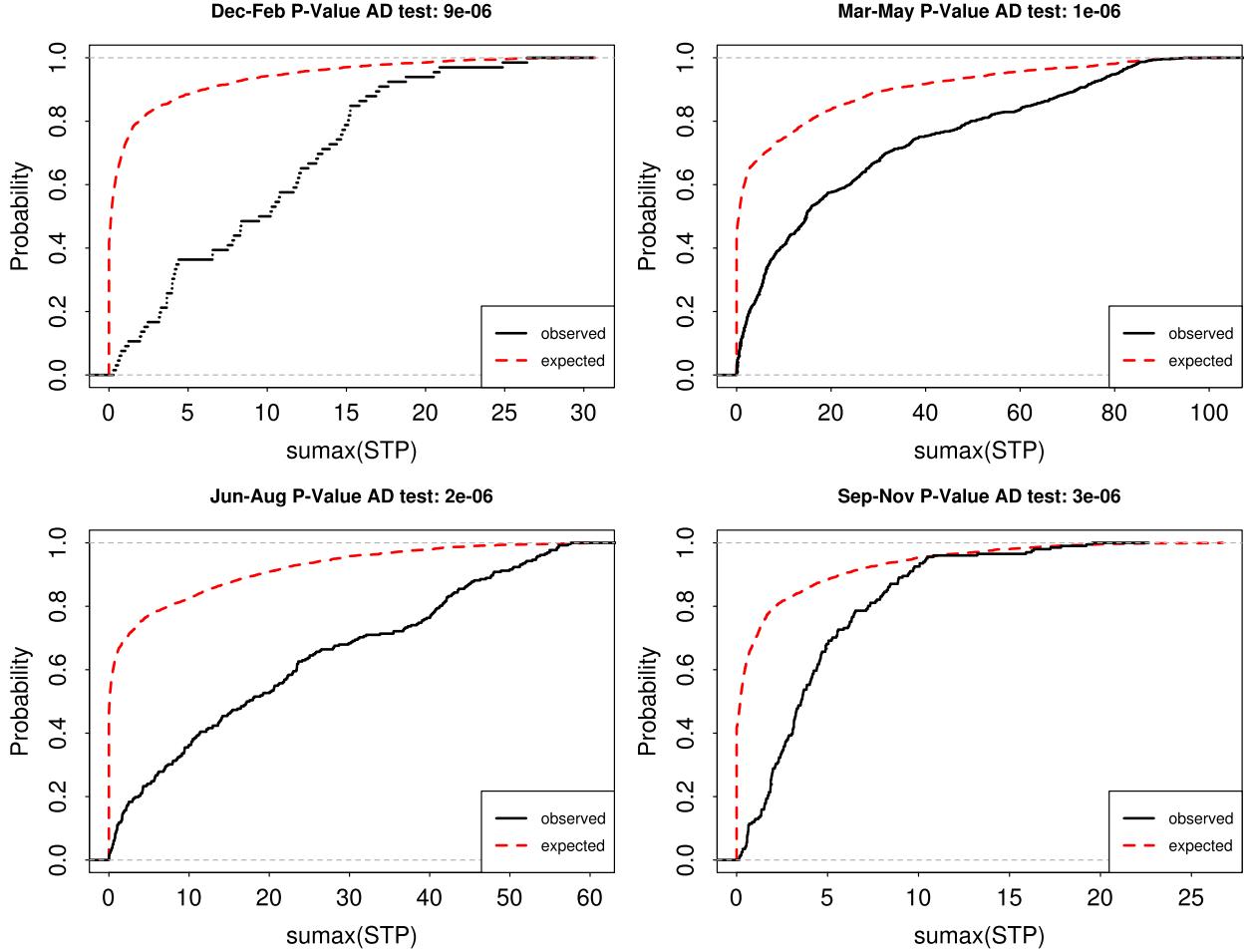


FIG. 8. As in Fig. 7, but for year 2011. Note the different x -axis scale here vs Fig. 7.

against the CDF values of the covariate evaluated at all locations u [$\mathbf{Z}(u)$], $F_0(z)$, which is an estimate of the fractional area less than z . The graphical representation is similar to a probability–probability plot with $\hat{F}(z)$ in the y axis and $F_0(z)$ in the x axis. If there is a preference of the point pattern for high values of the covariate, high values of $\hat{F}(z)$ would correspond to high values of the probability $F_0(z)$. AUC values range from 0 to 1 where values close to 0.5 indicate poor dependence of the point pattern process on the covariate $\mathbf{Z}(u)$ (i.e., low discrimination), while AUC values close to zero indicate high positive dependence (i.e., higher discrimination) of the point pattern process for spatial domains with higher values of the covariate.

4. Results

a. Dependence of tornado report density $\lambda(s)$ on different interval values of the covariate

Quarterly values of $\text{sumax}(\text{STP})$ were split into percentile intervals with limiting value probabilities given

by [0.5, 0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 0.9), and [0.9, 1.0], and the resulting values were plotted with associated tornado reports corresponding to each subinterval, with different colors assigned to each category labeled from 1 to 5 (not to be confused with tornado strength). Results are presented here for year 1987 (lowest quarterly report frequency on record; Fig. 3) and year 2011 (highest quarterly report frequency on record; Fig. 4). Formal dependence of the density process on the covariate was performed by using a chi-squared test for quadrat counts. In this case, the quadrats are irregular areas for the observed $\text{sumax}(\text{STP})$, defined by the percentile intervals described above. The null hypothesis is that this process is completely spatially random (CSR) and independent of $\text{sumax}(\text{STP})$ value. The null hypothesis was rejected for all quarters with p values ≤ 0.0001 , and thus it is concluded that $\lambda(s)$ is dependent on $\text{sumax}(\text{STP})$.

For both a record low report quarter year (1987; Fig. 3) and a record-high report quarter year (2011; Fig. 4), it is evident that there is a preference for tornado reports to

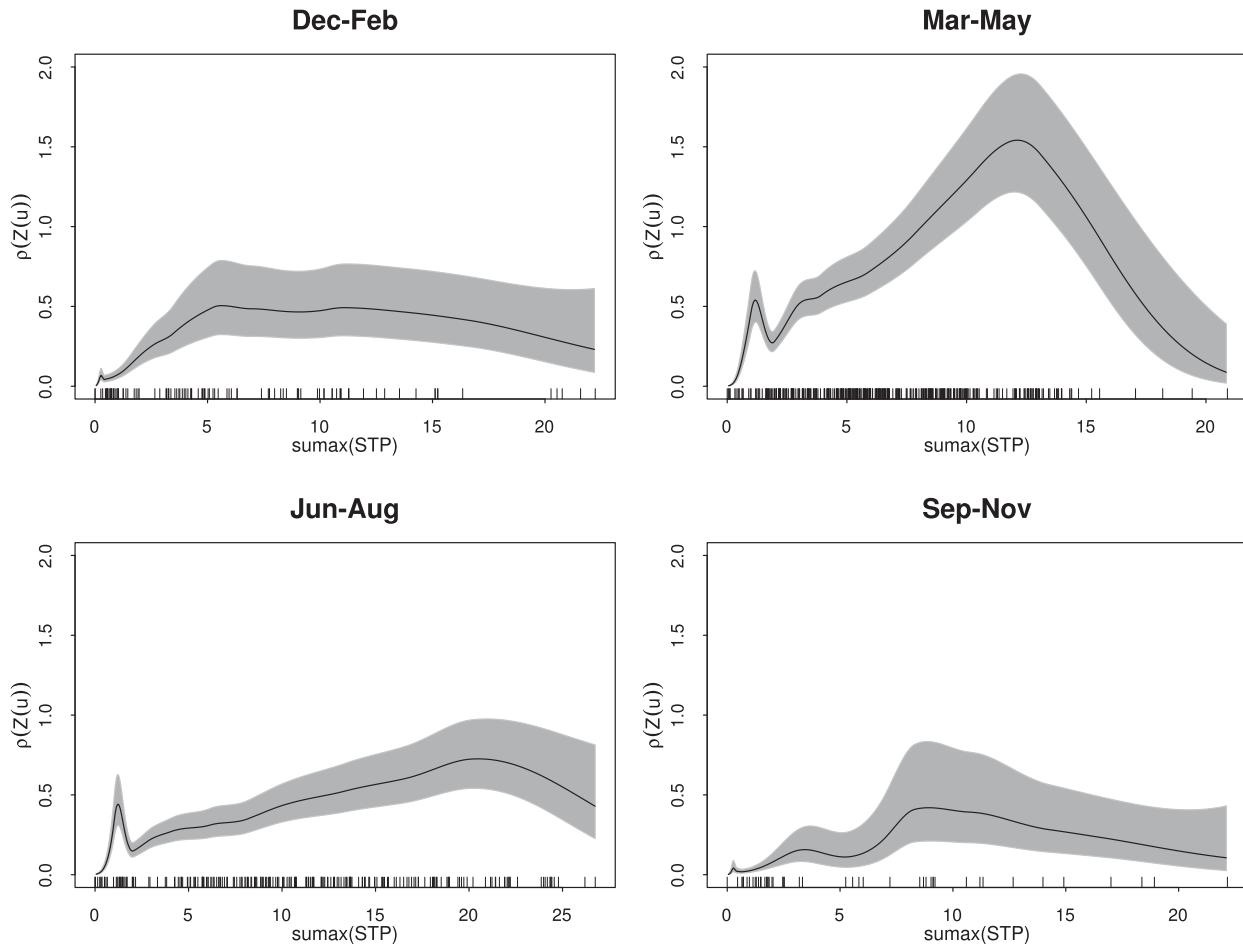


FIG. 9. For year 1987 nonparametric estimation of the tornado density process on the covariate sumax(STP). Solid black lines are estimated values of $\hat{p}[Z(u)]$ and gray bands indicate the 95% confidence intervals. The x -axis tick marks represent observed values of sumax(STP) from tornado reports for each quarter.

occur in the highest percentiles of sumax(STP) values. Another, perhaps more simple, representation of the density process as a function of the covariate values (as defined by percentile ranges) is to use a bar chart to quantify the density process against percentile classes for the different sumax(STP) percentile ranges. These plots are again compared for 1987 (Fig. 5) and 2011 (Fig. 6). The y axis represents tornado report rate of occurrence per unit area, which is referred to as the point process density. This analysis indicates that quarters 2 and 3 (March–May and June–August, respectively) represent the highest tornado point process density and are associated with the highest values of sumax(STP) (which are percentiles corresponding to a [0.9, 1.0] probability interval).

b. Anderson–Darling test for CSR

The Anderson–Darling test was performed to measure potential discrepancies between the CDF of sumax(STP)

at the report locations with the CDF at all locations over the CONUS. Another approach for this kind of test is to use the Kolmogorov–Smirnov statistic. However, the Anderson–Darling test has proven more meaningful since it is more sensitive to the tails of the distribution than the Kolmogorov–Smirnov (Razali and Wah 2011). If the covariate values at report locations are a random sample from sumax(STP), the two CDFs would be identical, indicating complete independence of the density process on the covariate. The Anderson–Darling test was applied by quarters for the two extreme years considered in the previous analyses (Figs. 7 and 8). In all quarters of both years, the p values were ≤ 0.0001 and the null hypothesis that the point pattern density is independent of the sumax(STP) covariate was rejected. There is a big gap between the continuous line representing the cumulative probability distribution of the covariate sumax(STP) at the report locations, from the dashed line representing the cumulative probability

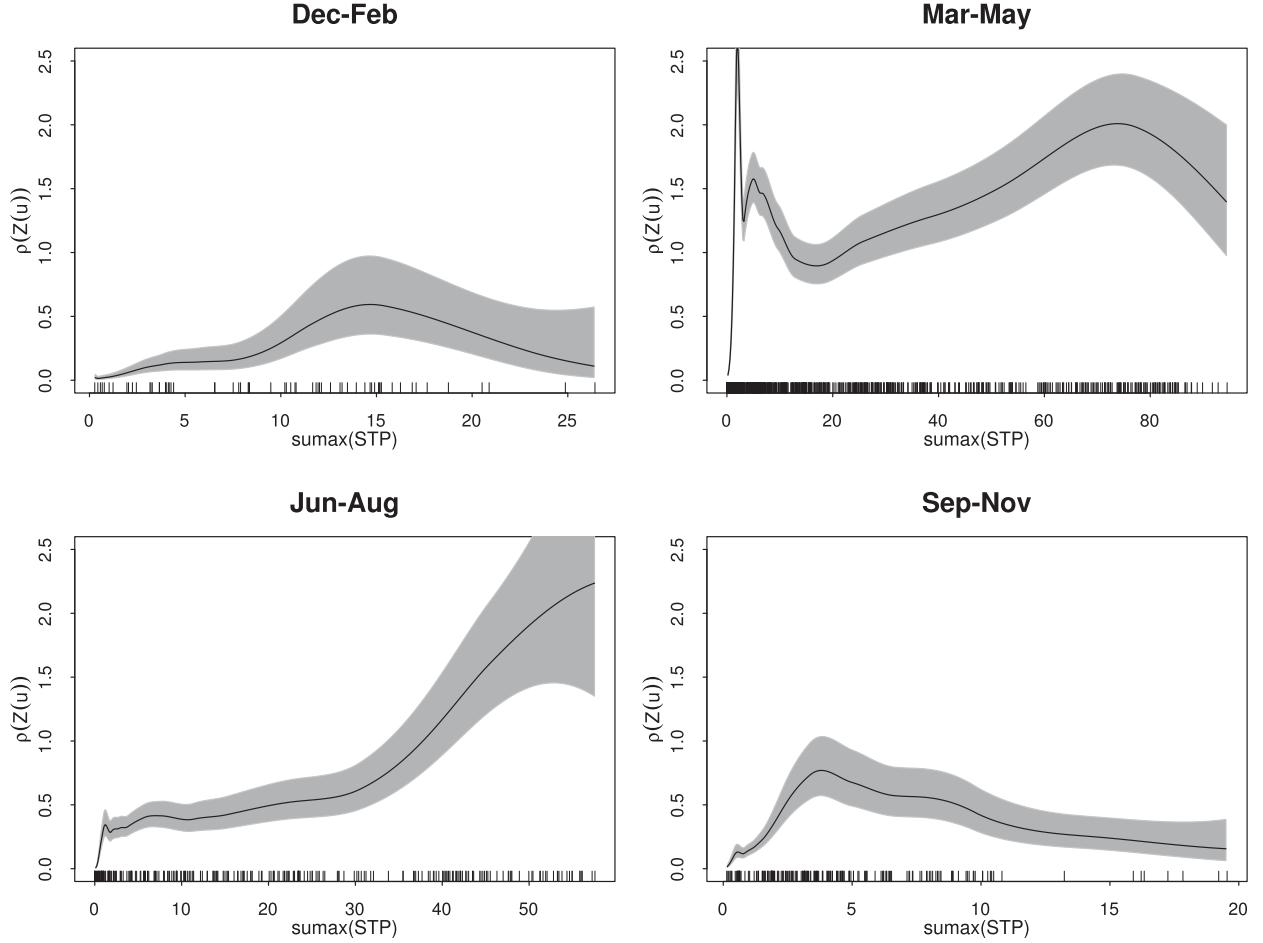


FIG. 10. As in Fig. 9, except for year 2011. Note the different x - and y -axis scales here vs Fig. 9.

distribution of $\text{sumax}(\text{STP})$ at all locations. The proportion of observations for low to moderate values of $\text{sumax}(\text{STP})$ at the tornado report locations is consistently lower than the proportion of these values in the study domain.

c. Nonparametric estimation of $\lambda(s)$ as a function of covariate $\text{sumax}(\text{STP})$

A formal test of dependence on a covariate, the Berman Z^2 test, was applied to each quarter for years 1987 and 2011 to formally test the dependence of the tornado reports on the covariate $\text{sumax}(\text{STP})$. The test yields very small p values ($p \approx 0$) for all quarters for both years, which rejects the hypothesis of independence in all cases.

All previous test statistics give enough evidence regarding the dependence of the density process $\lambda(s)$ on a covariate $Z(s) = \text{sumax}(\text{STP})$. Thus, it is reasonable to assume there exists a function $\rho(\cdot)$ that represents the dependence of the process density on the covariate $Z(s)$.

We can then infer $\lambda(s) = \rho[Z(s)]$ and use a nonparametric estimation of $\rho[Z(s)]$. This is performed using the function `rhohat` of the R library `spatstat` (Baddeley et al. 2016). A ratio estimator of the form

$$\hat{\rho}(z) = \frac{1}{|W|G'(z)} \sum_i \kappa[Z(s_i) - z] \quad (7)$$

is used for $\rho[Z(s)]$, where s_1, s_2, \dots, s_n are the report locations; $|W|$ is the total number of pixels in the study window; $G(z) = [\text{No. of } \{ \text{pixels } s : Z(s) \leq z \}] / |W|$; and $\kappa(z)$ is a Gaussian kernel. $G'(z)$ is a smooth estimate of $G(z)$. This estimator provides values of $\lambda(s)$ (y axis) as a direct function of the covariate $\text{sumax}(\text{STP})$ (x axis). The lower and upper 95% confidence intervals and estimated values of $\rho[Z(s)]$ are provided for each quarter of years 1987 (Fig. 9) and 2011 (Fig. 10). For 1987, quarter 1 indicates a general increase in density dependence as $\text{sumax}(\text{STP})$ increases, whereas quarter 2 shows a peak in dependence around a $\text{sumax}(\text{STP})$ value of 15. Quarters 3 and 4 during 1987 show a broad nonparametric

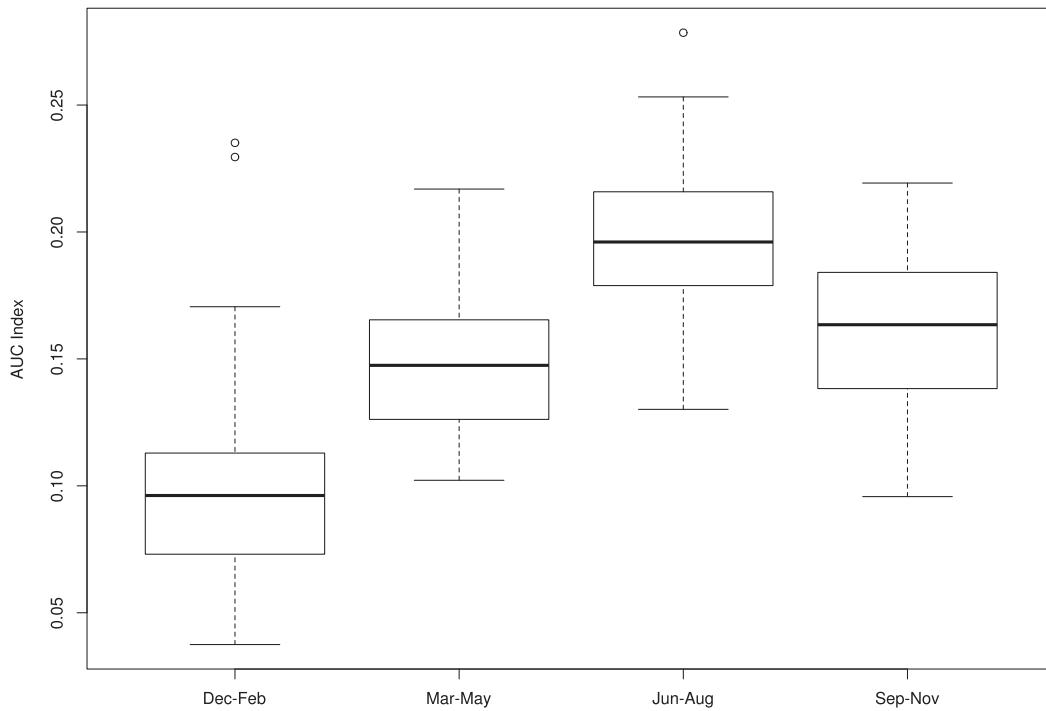


FIG. 11. Quarterly box-and-whisker plots of the AUC values for each ROC curve comparing $\hat{F}(z)$ with $F_0(z)$.

estimation of dependence across all sumax(STP) values indicating no strong preference for dependence based on sumax(STP) value. Similar results are noted for a much higher tornado frequency year (2011) for all quarters except quarter 3. In 2011, quarter 3 (June–August) showed more of an increase in tornado density dependence on sumax(STP) values. For both years, the tornado density process increases with increasing sumax(STP) during the annual climatological peak of U.S. tornado frequency (March–May).

d. Yearly and seasonal variation of dependence strength

ROC curves comparing the empirical cumulative probability distribution $\hat{F}(z)$ of the covariate sumax(STP) at the report locations with the cumulative probability distribution of the covariate sumax(STP) at all locations [$F_0(z)$] were calculated for each quarter and year of the study period (1979–2017). Figure 11 shows box-and-whisker plots indicating quarterly AUC values for all years. Thick horizontal lines in the boxes represent the median AUC values for each quarter. Recalling that lower AUC values mean high process density dependence on the covariate, we can confirm that the third quarter (June–August) shows the highest AUC values, resulting in a weaker discrimination of the tornado reports on sumax(STP), in comparison with the first (December–February), second (March–May), and fourth

(December–February) quarters. This is consistent with previous research showing composite parameters (such as STP) typically exhibit a reduction in explanatory capability in the boreal summer (Hart and Cohen 2016; Gensini and Brooks 2018).

A quarterly AUC time series was constructed for the study period to examine any potential trends associated with the covariate dependence on tornado reports (Fig. 12). The seasonal dependence of sumax(STP) AUC becomes quite apparent from this analysis, with the greatest dependence during quarter 1 (December–February), and the least dependence during quarter 3 (June–August). This is again consistent with the monthly standardized anomaly correlations shown in previous research (Gensini and Brooks 2018). In addition, there is a slight downward trend in the AUC time series over the study period, indicating a greater dependence on sumax(STP) through time (though this trend does not appear to be significant). To quantify this further, a multiple linear regression model was created with a seasonal and trend component:

$$\text{AUC}_t = \beta_1 t + \alpha_j + \varepsilon_t. \quad (8)$$

This model was fitted to the AUC time series, where β_1 represents the linear trend coefficient for quarterly data, α_j represents the quarterly seasonal effects with $j = 1, 2, 3, 4$ and ε_t represents a normally distributed random

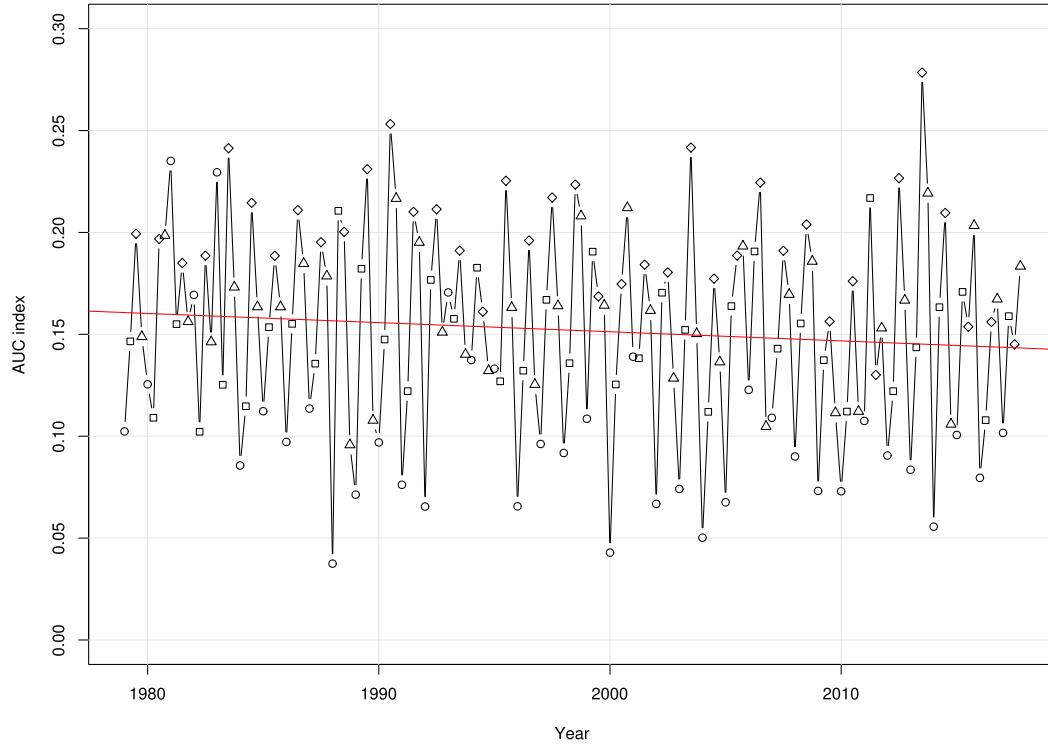


FIG. 12. Quarterly time series of the AUC values for each ROC curve comparing the empirical CDF of sumax(STP) at report locations $[\hat{F}(z)]$ with the CDF of sumax(STP) at all locations $[F_0(z)]$ for the period of 1979–2017.

error. The fitted model has an $R^2 = 0.957$ and the estimated coefficient β_1 is negative ($-0.000\,486\,1$) with corresponding p value = 0.04 for the slope significance t test.

e. Models fitted to the tornado density process as a function of the covariate

From these statistical tests, it is reasonable to conclude that there is a significant effect of the covariate sumax(STP) on the tornado report density function, and that a point process model can be fitted to quantify impact of the covariate on the report process density. Thus, an inhomogeneous Poisson process was fitted to each quarter of the years 1987 and 2011, with process density expressed as a parametric model of the covariance in the form

$$\lambda(u) = \exp[\alpha + \beta Z(u)], \quad (9)$$

where the density $\lambda(u)$ is a log-linear model of the covariate $Z(u)$, and α and β are model parameters to be estimated.

Table 1 presents the estimated α and β values for each quarter of 1987 and 2011. A value of $\beta = 0$ implies a constant density and independence of the covariate. In all cases, we were able to reject the null hypothesis that the coefficient β is equal to zero (p value ≈ 0). Interpretation of the estimated model parameters is as follows: For 1987, the estimated value of λ for the second

quarter (March–May) is $\exp(-1.938) = 0.14$ tornadoes per unit area for a value of sumax(STP) = 0. This amount would increase by to $\exp(-1.938 + 0.102) = 0.16$ tornadoes per unit area, when the sumax(STP) value increases 1 unit. This is equivalent to 1.79 tornadoes per $10\,000\text{ km}^2$ per unit increase of sumax(STP).

5. Summary

This research explored using the significant tornado parameter (STP) as a covariate to U.S. tornado frequency

TABLE 1. Inhomogeneous Poisson process parameters fitted to the tornado report density as a function of covariate sumax(STP). The model was fitted to each quarter of years 1987 and 2011. Values in parenthesis are the standard errors of the estimates.

Quarter	1987		2011	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
1	-4.469 (0.238)	0.222 (0.021)	-3.845 (0.175)	0.248 (0.018)
2	-1.938 (0.070)	0.102 (0.005)	-1.649 (0.062)	0.038 (0.002)
3	-2.281 (0.085)	0.104 (0.006)	-1.870 (0.068)	0.053 (0.002)
4	-2.880 (0.113)	0.109 (0.0123)	-2.521 (0.095)	0.156 (0.009)

in a climatological context. More specifically, the seasonal (i.e., quarterly) sums of the daily maximum STP value were tested for spatial dependence on tornado reports. Through several statistical tests, we conclude that STP is a statistically significant covariate to U.S. tornado frequency at the aggregated time and space scales examined herein. Spatial dependence of a tornado report on STP was found to vary by season, with greatest dependence noted in the winter months (December–February). The boreal summer season (June–August) exhibited the lowest dependence which is consistent with previous research. In addition, an interannual analysis of all quarters from 1979 to 2017 shows that the dependence of tornado reports on STP has not changed significantly throughout time.

The most novel aspect of this research developed nonparametric estimations of the spatial tornado density process (i.e., tornado reports per unit area) as a function of sumax(STP) for two vastly different tornado report frequency years. In general, for both years and all quarters, the tornado density point process increases as the quarterly sumax(STP) value increases. The difference in shape and slope of the nonparametric curves highlights the need for inclusion of seasonal cycle information into any climatological tornado model.

Finally, an inhomogeneous Poisson process was fitted for both 1987 and 2011 by using tornado point density expressed as a parametric model of the covariance. These results develop a spatial–statistical model that aids in the understanding of tornado density (dependent variable) as a function of quarterly sums of daily maximum STP (independent variable) by season. Future work may add a seasonal and interannual component to this model and/or explore the incorporation of tornado strength into the point process model. This analysis will assist in the future development of predictive spatial–statistical models that may aid in seasonal tornado forecasting.

Acknowledgments. The authors would like to thank the efforts of three anonymous reviewers in enhancing the quality of this manuscript.

REFERENCES

- Allen, J. T., M. J. Molina, and V. A. Gensini, 2018: Modulation of annual cycle of tornadoes by El Niño–Southern Oscillation. *Geophys. Res. Lett.*, **45**, 5708–5717, <https://doi.org/10.1029/2018GL077482>.
- Anderson, C. J., C. K. Wikle, Q. Zhou, and J. A. Royle, 2007: Population influences on tornado reports in the United States. *Wea. Forecasting*, **22**, 571–579, <https://doi.org/10.1175/WAF997.1>.
- Baddeley, A., E. Rubak, and R. Turner, 2016: *Spatial Point Patterns: Methodology and Applications with R*. Taylor and Francis/CRC Press, 810 pp.
- Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618, [https://doi.org/10.1175/1520-0434\(1994\)009<0606:OTEOTA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0606:OTEOTA>2.0.CO;2).
- , J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67**, 73–94, [https://doi.org/10.1016/S0169-8095\(03\)00045-0](https://doi.org/10.1016/S0169-8095(03)00045-0).
- Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79, [https://doi.org/10.1175/1520-0434\(2000\)015<0061:PSMUAN>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0061:PSMUAN>2.0.CO;2).
- Cheng, V. Y., G. B. Arhonditsis, D. M. Sills, W. A. Gough, and H. Auld, 2016: Predicting the climatology of tornado occurrences in North America with a Bayesian hierarchical modeling framework. *J. Climate*, **29**, 1899–1917, <https://doi.org/10.1175/JCLI-D-15-0404.1>.
- Doswell, C. A., III, and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting*, **9**, 625–629, [https://doi.org/10.1175/1520-0434\(1994\)009<0625:TEONTV>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2).
- , and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electron. J. Severe Storms Meteor.*, **1** (3), <http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/11/12>.
- Elsner, J. B., and Coauthors, 2016: The relationship between elevation roughness and tornado activity: A spatial statistical model fit to data from the central Great Plains. *J. Appl. Meteor. Climatol.*, **55**, 849–859, <https://doi.org/10.1175/JAMC-D-15-0225.1>.
- Gensini, V. A., and W. S. Ashley, 2011: Climatology of potentially severe convective environments from the North American regional reanalysis. *Electron. J. Severe Storms Meteor.*, **6** (8), <http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/85>.
- , and A. Marinaro, 2016: Tornado frequency in the United States related to global relative angular momentum. *Mon. Wea. Rev.*, **144**, 801–810, <https://doi.org/10.1175/MWR-D-15-0289.1>.
- , 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>.
- , T. L. Mote, and H. E. Brooks, 2014: Severe-thunderstorm reanalysis environments and collocated radiosonde observations. *J. Appl. Meteor. Climatol.*, **53**, 742–751, <https://doi.org/10.1175/JAMC-D-13-0263.1>.
- Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, <https://doi.org/10.1175/WAF-D-11-0008.1>.
- Hart, J. A., and A. E. Cohen, 2016: The challenge of forecasting significant tornadoes from June to October using convective parameters. *Wea. Forecasting*, **31**, 2075–2084, <https://doi.org/10.1175/WAF-D-16-0005.1>.
- Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612, [https://doi.org/10.1175/1520-0434\(1992\)007<0588:SLSF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2).
- Karpman, D., M. A. Ferreira, and C. K. Wikle, 2013: A point process model for tornado report climatology. *Stat.*, **2**, 1–8, <https://doi.org/10.1002/sta4.14>.
- McNulty, R. P., 1985: A conceptual approach to thunderstorm forecasting. *Natl. Wea. Dig.*, **10** (2), 26–30.
- 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>.

- Molina, M. J., J. T. Allen, and V. A. Gensini, 2018: The Gulf of Mexico and ENSO influence on subseasonal and seasonal conus winter tornado variability. *J. Appl. Meteor. Climatol.*, **57**, 2439–2463, <https://doi.org/10.1175/JAMC-D-18-0046.1>.
- Potvin, C. K., K. L. Elmore, and S. J. Weiss, 2010: Assessing the impacts of proximity sounding criteria on the climatology of significant tornado environments. *Wea. Forecasting*, **25**, 921–930, <https://doi.org/10.1175/2010WAF2222368.1>.
- Rasmussen, E. N., 2003: Refined supercell and tornado forecast parameters. *Wea. Forecasting*, **18**, 530–535, [https://doi.org/10.1175/1520-0434\(2003\)18<530:RSATFP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)18<530:RSATFP>2.0.CO;2).
- , —, 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](https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2).
- Razali, N. M., and Y. B. Wah, 2011: Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Anal.*, **2** (1), 21–33.
- Schaefer, J. T., and R. Edwards, 1999: The SPC Tornado/Severe Thunderstorm Database. Preprints, *11th Conf. Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 215–220.
- Sherburn, K. D., and M. D. Parker, 2014: Climatology and ingredients of significant severe convection in high-shear, low-CAPE environments. *Wea. Forecasting*, **29**, 854–877, <https://doi.org/10.1175/WAF-D-13-00041.1>.
- , —, J. R. King, and G. M. Lackmann, 2016: Composite environments of severe and non-severe high-shear, low-CAPE convective events. *Wea. Forecasting*, **31**, 1899–1927, <https://doi.org/10.1175/WAF-D-16-0086.1>.
- 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](https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2).
- , —, and C. M. Mead, 2004: An update to the supercell composite and significant tornado parameters. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P8.1, <https://ams.confex.com/ams/pdfpapers/82100.pdf>.
- , B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, <https://doi.org/10.1175/WAF-D-11-00116.1>.
- Tippett, M. K., A. H. Sobel, and S. J. Camargo, 2012: Association of U.S. tornado occurrence with monthly environmental parameters. *Geophys. Res. Lett.*, **39**, L02801, <https://doi.org/10.1029/2011GL050368>.
- , —, —, and J. T. Allen, 2014: An empirical relation between U.S. tornado activity and monthly environmental parameters. *J. Climate*, **27**, 2983–2999, <https://doi.org/10.1175/JCLI-D-13-00345.1>.
- Togstad, W. E., J. M. Davies, S. J. Corfidi, D. R. Bright, and A. R. Dean, 2011: Conditional probability estimation for significant tornadoes based on Rapid Update Cycle (RUC) profiles. *Wea. Forecasting*, **26**, 729–743, <https://doi.org/10.1175/2011WAF2222440.1>.
- 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>.
- Wikle, C. K., and C. J. Anderson, 2003: Climatological analysis of tornado report counts using a hierarchical Bayesian spatio-temporal model. *J. Geophys. Res.*, **108**, 9005, <https://doi.org/10.1029/2002JD002806>.