An empirical relation between U.S. tornado activity and monthly environmental parameters

MICHAEL K. TIPPETT*

International Research Institute for Climate and Society, Columbia University, Palisades, New York

Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

ADAM H. SOBEL

Department of Applied Physics and Applied Mathematics, Department of Earth and Environmental Sciences, Columbia University, New York, New York

SUZANA J. CAMARGO

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

JOHN T. ALLEN

International Research Institute for Climate and Society, Columbia University, Palisades, New York

Corresponding author address: M. K. Tippett, International Research Institute for Climate and Society, The Earth Institute of Columbia University, Lamont Campus 61 Route 9W, Palisades New York 10964, USA.

E-mail:tippett@iri.columbia.edu

ABSTRACT

Recently an empirical relation between U.S. monthly tornado activity and monthly-averaged environmental parameters was established. The relation is expressed in the form of an index, and here a detailed comparison is made between the index and reported tornado activity. The index is a function of two environmental parameters taken from the North American Regional Reanalysis: convective precipitation (cPrcp) and storm relative helicity (SRH). Additional environmental parameters are considered for inclusion in the index, among them convective available potential energy, but their inclusion does not significantly improve the overall climatological performance of the index. The aggregate climatological dependence of reported monthly U.S. tornado numbers on cPrcp and SRH is well described by the index, though the index fails to capture nonsupercell and cool season tornadoes. The contributions of the two environmental parameters to the index annual cycle and spatial distribution are examined with the seasonality of cPrcp (maximum during summer) relative to SRH (maximum in winter) accounting for the index having its peak value in May. The cPrcp spatial distribution enhances the index in the south and southeast and surpresses it west of the Rockies and over elevation. At the scale of the NOAA climate regions, the largest deficiency of the index climatology occurs over the Central region where the index peak in spring is too low and where in late summer the drop-off in the reported number of tornadoes is poorly captured. This index defficiency is related to its sensitivity to SRH, and increasing the index sensitivity to SRH improves the representation of the annual cycle in this region. The ability of the index to represent the interannual variability of the monthly number of U.S. tornadoes can be ascribed during most times of the year to interannual variations of cPrcp rather than of SRH. However, both factors are important during the peak spring period. Additionally, the index shows some skill in representing the interannual variability of monthly tornado numbers at the scale of NOAA climate regions.

1. Introduction

Identifying environmental "ingredients" for severe weather and tornadic storms provides useful guidance for forecasters in interpreting observed soundings and short-range forecasts, and numerous studies have examined the question of which local environments are associated with the formation of tornadoes (Maddox 1976; Brooks et al. 1994; Rasmussen and Blanchard 1998; Brooks et al. 2003; Grams et al. 2011). Various measures of vertical wind shear and potential updraft strength have been found to be conducive to tornado activity. However, tornadogenesis is a highly complex phenomenon which depends on multiple small-scale processes as well as the ambient environment (Wurman et al. 2012). Therefore, even when the environment is favorable and a thunderstorm has formed, the occurrence or nonoccurrence of tornadoes remains highly uncertain.

The ingredients approach is also useful in addressing the question of how climate signals such as ENSO and changes in radiative forcing may influence tornado activity. Direct treatment of these questions is challenging because a high-quality homogeneous observation record is unavailable, and climate models currently do not resolve thunderstorms or tornadoes. Mixtures of dynamical and ingredients-based statistical downscaling have been used to investigate projected changes in severe convection environments (Trapp et al. 2009, 2007) and to relate other climate signals to tornadic storms (Lee et al. 2012; Weaver et al. 2012; Thompson and Roundy 2013).

Recently an empirical relationship, expressed in the form of an index, has been shown between monthly-averaged environmental quantities and tornado activity over the contiguous US (CONUS; Tippett et al. 2012). The use of monthly averages differs substantially from previous work which has used environmental quantities on shorter (typically 6-hourly) time-scales. The degree to which the monthly index co-varies with reported tornado activity is somewhat surprising since the lifetime of a tornadic event is no more than a few hours and often only a few minutes. At least conceptually,

changes in the frequency of extreme environments associated with tornadoes can be thought of as being caused by both changes in the mean or the variance of the distribution of environments during a month. Success of the monthly index suggests that changes in monthly-mean environment do reflect to some extent changes in the frequency of extreme environments. Some motivations for the use of monthly averages are that this time-scale is better simulated and projected in relatively low-resolution numerical models and that the impacts of climate signals are better understood on longer time-scales. A obvious disadvantage of using monthly-averaged environmental parameters is that the joint high-frequency behavior of the ingredients is lost.

Tippett et al. (2012) investigated some general properties of the monthly tornado index including the climatological number of CONUS tornadoes per month predicted by the index, the annually averaged spatial pattern of the index and the interannual variability of the number of CONUS tornadoes predicted by the index. However, more detailed analysis of the index is necessary if the index is be used as a tool to diagnose the impact of climate signals tornado activity. Here we examine the properties of the index in more depth, including aspects of the environmental parameter selection, systematic deficiencies and regional behavior. The paper is organized as follows: Tornado report and environmental data are described in Section 2. Index construction and parameter selection is discussed in Section 3. The annual cycle of the index is described in Section 4, and its interannual variability is described in Section 5. A summary and future prospects are given in Section 6.

2. Data

a. Tornado Data.

U.S. tornado data covering the period 1950-2010 is provided by the NCEP Storm Prediction Center (SPC) Tornado, Hail and Wind Database (Schaefer and Edwards 1999). As has been discussed extensively by other authors, substancial variability in the tornado report record is unrelated to climate variations and is due to variations in reporting practices, introduction of Doppler radar, etc. (Verbout et al. 2006; Doswell et al. 2009). The annual number of reported weak (F0) tornadoes has increased dramatically, roughly doubling over the last 60 years (Fig. 1a) consistent with the findings of Brooks and Doswell (2001). The annual numbers of reported F1 (Fig. 1b) and F2-F5 (Fig. 1c) tornadoes do not show such strong trends, but there are some indications of changes occurring in the late 1970s and 1980s, especially in the F2-F5 reports. Reported annual totals from the last 20 years seem relatively homogeneous across each of the intensity levels. Ideally, the non-physical variations in the observational record could be removed from the tornado record, and the SPC does compute an "inflation adjusted"[1] annual number of U.S. tornadoes using a trend line for the period 1954-2007 and taking 2007 as a baseline. However, there is no rigorous justification for the use of a linear correction to tornado frequency which may in itself may introduce artifacts. Spatially varying features of the observational records are even more difficult to quantify. Limiting our attention to reports of more intense tornadoes or to more recent periods (last two decades) has the disadvantage of substantially reducing the sample size which may be a problem, especially for the spatial dependence.

The temporal inhomogeneities associated with the tornado record are a primary reason for tak-

[1]The "Inflation-adjusted" tornado count was developed by Harold Brooks, NSSL and Greg Carbin, of the SPC and is described at http://www.spc.noaa.gov/wcm/adj.html

ing the monthly-averaged tornado report *climatology* as the target of the index fitting procedure (Tippett et al. 2012). Doing so avoids the possibility of the statistical analysis spuriously associating nonphysical changes (trends and shifts) in the tornado record with coincidental physical variations. This choice also leaves the interannually varying record as an independent data set for further verification. Disadvantages of this approach are that the range of co-variability of tornado occurrence with environmental parameters in the climatology data is greatly reduced, the joint distribution of parameters is only climatological, and there is no association of particular tornadic events with the physical environment. The monthly tornado data (F0 and greater) is put onto a $1^\circ \times 1^\circ$ latitude-longitude grid (25N - 50N, 130W - 60W) for the 32-year period 1979-2010. The upward trend in the number of reported tornadoes presumably results in the gridded monthly tornado climatology being negatively biased with respect to the most recent period.

b. North American Regional Reanalysis.

Monthly averaged environmental parameters are taken from the North American Regional Reanalysis (NARR; Mesinger and Coauthors 2006). NARR data are provided on a 32-km Lambert conformal grid which we interpolate to a $1^\circ \times 1^\circ$ latitude-longitude grid over the CONUS (25N - 50N, 130W - 60W). Only data over land points are used. We consider monthly averages of the following NARR variables: surface convective available potential energy (CAPE), surface convective inhibition (CIN), best (4-layer) lifted index (4LFTX), the difference in temperature at the 700 hPa and 500 hPa levels divided by the corresponding difference in geopotential height (lapse rate), the average specific humidity between 1000 hPa and 900 hPa (mixing ratio), 3000-0m storm relative helicity (SRH), the magnitude of the vector difference of the 500 hPa and 1000 hPa winds (vertical shear), precipitation, convective precipitation (cPrcp) and elevation. Lapse rate and vertical shear are computed using monthly averages of the constituent variables. We take the natural loga-

rithm of CAPE, SRH, vertical shear, precipitation and cPrcp, consistant with previous analysis of environmental factors impacting severe weather on synoptic time scales (e.g., Brooks et al. 2003).

3. Poisson regression and parameter selection

Tippett et al. (2012) related the climatological monthly number of U.S. tornadoes to climatological monthly averages of collocated NARR atmospheric parameters using Poisson regression. The same method was used to develop a tropical cyclone genesis index (Tippett et al. 2011). The monthly number of tornadoes summed over T years in a grid box is assumed to be a Poisson distributed random variable with expected value μ. The expected value μ is the *monthly tornado activity index* and is assumed to have a log-linear dependence on the environmental parameters modeled by

$$\mu(\mathbf{x}) = \exp\left[\mathbf{b}^T\mathbf{x} + c + \log\left(\Delta x\ \Delta y\ T\cos\phi\right)\right]\ , \tag{1}$$

where $\mathbf{x}$ is a vector of environmental parameters, $\mathbf{b}$ is a vector of regression coefficients, c is an intercept term, ϕ is the latitude, Δx and Δy are the longitude and latitude spacings in degrees, respectively, and T is the number of years. The last term accounts for the differing area of each grid box and the number of years used in the climatology and removes the dependence of the coefficients on grid resolution and climatology length. The regression model (1) with the same coefficients is used at all locations and all times of the year. In addition to relating tornado activity with environmental parameters, the regression can correct spatially and seasonally uniform systematic errors in the NARR environmental parameters. The regression coefficients are estimated by maximum likelihood, and a commonly used goodness of fit measure, deviance, is also determined (McCullagh and Nelder 1989).

A key issue is the choice of environmental parameters included in the index. Including too few environmental parameters gives a model which poorly fits the data, while including too many leads

to overfitting and poor performance on independent data. Tippett et al. (2012) took the previously listed set of 10 monthly-averaged parameters associated with tornado occurrence and used a forward selection procedure to find the best set of parameters for a given number of parameters. This approach reduces the parameter selection problem to one of selecting the number of parameters. Increasing the number of parameter always improves the (in-sample) fit of the index to the data. However, evaluation of the fit on out-of-sample data using cross-validation showed that including more than 2 parameters did not produce a significant increase in the overall fit. This finding does not rule out that additional parameters might result in significant improvements in fit for particular regions or months nor does it say anything about the utility of additional parameters outside the climatological setting.

In the simplest sense, potential updraft strength and vertical wind shear are the two basic environmental factors considered favorable for tornado activity. However, there are many related parameters which measure these conditions. The deviance-based R-squared (Cameron and Windmeijer 1996) values of the six best (in the sense of minimizing mean cross-validated deviance) 2-parameter models are shown in Fig. 2 and range from 0.53 to 0.67. The best one-parameter model uses cPrcp and has a deviance-based R-squared value of 0.46, giving an indication of the benefit of including an additional predictor. The uncertainty of the estimates, shown as plus and minus one standard deviation error bars, is computed from 10 repetitions of 10-fold cross-validation.[2] These six statistical models include one parameter associated with convective instability (CAPE or cPrcp) and one associated with vertical shear (SRH, mixing ratio, or vertical shear). The model with smallest deviance uses cPrcp and SRH, and replacing SRH with vertical shear does not give a significantly different fit.

[2] 10-fold cross-validation consists of splitting the data into 10 randomly selected set, fitting the coefficient on 9 of those sets and validating on the tenth.

The energy-helicity index (EHI), the product of SRH and CAPE, is often used on synoptic time scales as a forecast parameter (Rasmussen and Blanchard 1998). A related quantity is the significant severe parameter, the product of CAPE and vertical shear (Davies and Johns 1993). Both of these quantities are included in the Poisson regression model framework and correspond to choosing the entries of $\mathbf{b}$ to be unity for the appropriate choice of parameters. Interestingly, the fit of the Poisson regression model with CAPE and SRH as parameters is significantly worse than that of the best 2-parameter model, and the model with CAPE and shear is close to being significantly worse as well. Given the widespread use of the EHI and other CAPE-based measures on synoptic time scales, the poor performance of its constituent parameters here is unexpected.

Of the many possible explanations for the relatively poor performance of the CAPE-based Poisson regressions, foremost is that monthly-averaged CAPE simply fails to capture the relation with tornado activity that is observed in higher frequency data. Another possibility is that the relation between monthly-averaged CAPE and tornado activity is not well-fit by the functional form of the Poisson regression model. The coefficients Poisson regression model can be interpreted as the sensitivity of the expected monthly number of tornadoes to changes in the environmental parameters. Specifically, for a small change $\delta\mathbf{x}$ in the environmental variables, the change $\delta\mu$ in the expected number of tornadoes is given by

$$\frac{\delta\mu}{\mu} \approx \mathbf{b}^T\delta\mathbf{x} = b_1\delta x_1 + b_2\delta x_2\,.$$

That is, for a 0.01 unit change in one of the environmental parameters, the value of its coefficient is the corresponding percent change in the expected value μ. Equivalently, the Poisson regression coefficients are the partial logarithmic derivatives of the expected monthly number of tornadoes with respect to the environmental variables since

$$\frac{\partial}{\partial x_i}\log\mu = b_i\,.$$

Since the coefficients are constant, the Poisson regression model assumes that the sensitivity of the number of tornadoes to the environmental parameters is constant, and, in particular, does not depend on the values of the environmental parameters.

The extent to which the tornado data and environmental parameters satisfy the Poisson regression functional form was investigated using the same approach as in Tippett et al. (2011). For each parameter, we compute its Poisson regression coefficient for different ranges of that parameter while allowing the other parameters to vary. Essentially we are computing the partial logarithmic derivative for different values of the parameters and determining if that derivative is constant. Note that the above procedure is different from computing the average number of tornadoes as a function of one of the variables and checking for a linear relation which would be equivalent to taking the ordinary derivative and would give rather different results for correlated quantities. Specifically, here we compute the Poisson regression coefficient of each parameter over four ranges defined by the 10-30, 30-50, 50-70 and 70-90 percentiles of the parameter. Error bars for the coefficient estimate are defined as twice the standard deviation of 100 bootstrap estimates of the coefficients. Figure 3a shows clearly that the coefficient of $\log(\text{CAPE})$ is not constant. There is enhanced sensitivity of climatological tornado occurrence to CAPE in the 10-30 percentile range which decreases until $\log(\text{CAPE}) \approx 4$ at which point the coefficient is roughly the same as that obtained when the complete data is used. We hypothesize that this mismatch between the observed sensitivity to CAPE and that imposed by the Poisson regression functional form is the reason for the relatively poor performance of the CAPE-based models. More sophisticated models may be better able to accommodate the variable sensitivity of climatological monthly tornado activity to monthly averaged CAPE (Mestre and Hallegatte 2009; Villarini et al. 2010), or the behavior might be ameliorated with the inclusion of additional parameters. The choice of which strategy to pursue, non-log-linear dependence or additional parameters, would essentially depend on whether the behavior in 3a re-

flects a physical property or is an artifact of the analysis. On the other hand, Fig. 4 shows that the coefficient of cPrcp is approximately constant over the range of values and consistant with the value estimated from the complete data.

The sensitivity of the expected number of monthly tornadoes to SRH is similar whether SRH is used in conjunction with either cPrcp or CAPE. In both cases, the SRH coefficient confidence intervals over the 30-50 percentile range fail to include the value estimated from the complete data, and there is some indication of greater sensitivity to SRH, especially in combination with CAPE (Figure 3b). We return to this finding in later sections.

4. Climatological features

a. Dependence on cPrcp and SRH

We first compare the dependence of the index on cPrcp and SRH with that of the observations. The index $\mu(\mathrm{cPrcp}, \mathrm{SRH})$ expresses the expected number of tornadoes for given values of cPrcp and SRH. The corresponding observed quantity is the average number of tornadoes at all locations and months of the year with the specified values of cPrcp and SRH. The observed climatological numbers of tornadoes are binned according to the corresponding values of cPrcp and SRH. Bin boundaries of cPrcp and SRH individually correspond to percentiles and range from the 5th to the 95th percentile with a width of 5 percent. Figure 5 shows the average number of observed tornadoes and the index as functions of cPrcp and SRH. The log-linear form of the index means that its isolines as a function of $\log(\mathrm{cPrcp})$ and $\log(\mathrm{SRH})$ are straight lines. The index isolines are overlaid on the observed distribution to aid in comparison, and, for the most part, there is a reasonable match between the functional dependence of the observations and the index, especially for the parameter ranges associated with the largest number of tornadoes. The isolines of the

observed distribution are not precisely straight and indicate a tendency of a greater sensitivity to large values of SRH, consistent with the results of the previous section (Figs. 3b and 4b). The difference of the observations and index show little indication of systematic bias over the parameter ranges associated with the majority of tornadoes The largest discrepancies between the observations and the index are seen for simultaneously low values of cPrcp and SRH (the gray box marked B1 in Fig. 5c), fore which parameter values there are more observed tornadoes than predicted by the index. Conversely, for intermediate values of SRH and low values of cPrcp, (the gray box marked B2 in Fig. 5c) there are no observed tornadoes while the index predicts small numbers.

The index biases associated with the parameter ranges in B1 and B2 correspond to fairly well-defined geographical regions and calendar months. Figure 6 shows the spatial distributions and annual cycles of the data with parameters in boxes B1 and B2. The negative bias in box B1 is seen to be due to the failure of the index to produce observed April-November tornadoes occurring west of the Rockies, concentrated in Southern California and corresponding to about 2.4 tornadoes per year. These tornadoes are likely associated with different environmental conditions than the index is designed to detect and mainly comprised of low CAPE and high shear environments (Hanstrum et al. 2002; Monteverdi et al. 2003; Kounkou et al. 2009). The positive bias in box B2 is is due to the index indicating tornado activity mainly west of 100°W during October through April and corresponds to about 7.4 tornadoes per year. Both observations and index (by construction) have 999 tornadoes per year.

b. Contribution of cPrcp and SRH to annual cycle and spatial pattern

The annual cycle of the total number of reported tornadoes and the annual cycle of the index are shown in Fig. 7a. The index captures the general phasing with maximum values in May and

minimum values in winter. Overall the index shows less variability through the seasonal cycle than do the observations. The index shows a positive bias in August and September, a feature that we will examine in more detail later. The simplicity of the tornado index makes it possible to diagnose the contribution of the two environmental factors to the annual cycle. We compute the index with the annual cycle of SRH suppressed and with the annual cycle of cPrcp suppressed. In the first case only cPrcp contributes to the annual cycle and in the second case only SRH. The annual cycles of these single-factor indices are shown in Fig. 7(b). The contribution of SRH to the annual cycle has maximum values in winter and minimum values in late summer. Nearly out of phase, the contribution of cPrcp to the annual cycle has maximum values in June and July and minimum values in winter. Although the most unstable atmospheric conditions occur in June and July, the contribution from SRH is nearly minimum at that time of the year.

The index can be written as the normalized product of the two single-factor indices

$$\mu(x_1, x_2) = \frac{\mu(x_1, \overline{x}_2)\mu(\overline{x}_1, x_2)}{\mu(\overline{x}_1, \overline{x}_2)}\,, \tag{2}$$

where $\overline{()}$ denotes annual average. At each location, the annual cycle is exactly the product of the two single-factor indices normalized by $\mu(\overline{x}_1, \overline{x}_2)$ which has no annual cycle. The normalized product of the spatially summed single-factor indices

$$\frac{\langle\mu(x_1, \overline{x}_2)\rangle\langle\mu(\overline{x}_1, x_2)\rangle}{\langle\mu(\overline{x}_1, \overline{x}_2)\rangle}\,, \tag{3}$$

is not the same as the index annual cycle $\langle\mu(x_1, x_2)\rangle$; the notation $\langle\cdot\rangle$ denotes the spatial sum. However, Fig. 7(b) shows that this product does have its maximum in May like the complete index and the observations. The two factors are nearly but not quite out of phase. The minimum of the SRH factor is in August, while the maximum of the cPrcp factor is in June. This difference in phasing is the reason that the product of the annual cycles of the two factors has its maximum in late spring when the contribution from cPrcp is already large and that from SRH is still fairly large.

This result indicates that the May maximum of the index can be explained by the phasing of the annual cycles of the cPrcp and SRH contributions.

A similar approach can be used to determine how the two factors contribute to the annually-summed spatial distribution of tornado occurrence. The annual distribution of reported (3x3 box-averaged smoothing) tornadoes and index values shown in Figs. 8(a) and 8(b), respectively, have similar overall patterns. The index is missing the observed maximum in the northeast corner of Colorado where non-supercell tornadoes are common and local effects contribute to the low-layer shear in this area (Wakimoto and Wilson 1989). The index values do not extend far enough into the northern high plains and extend too far south into Texas. To quantify the impact of the two environmental parameters on the spatial distribution, we compute the index with the spatial variability of SRH suppressed and the index with the spatial variability of cPrcp suppressed. In the first case only cPrcp contributes to the spatial variability and in the second case only SRH. cPrcp enhances tornado index activity in the south and southeast, and limits it elsewhere (Fig. 8c). The SRH factor enhances the index in the "tornado alley" region and suppresses activity in the southeast (Fig. 8d).

c. Regional features of the annual cycle

We compute the annual cycle of the index and the tornado reports in the 9 NOAA climate regions (Karl and Koss 1984) (Fig. 9); overlapping gridpoints are weighted according to the fraction of area within the region. The Pearson (rank) correlation between the observation and index regional annual cycles exceeds 0.85 (0.83) in all regions except for the Northwest and West where the correlation is 0.38 and 0.68 (0.24 and 0.58), respectively. Positive biases are seen for the months of August through October in the South, Central, Upper Midwest and Plains regions, a feature we will examine in some detail. The index shows a substantial negative bias in the Southeast during September that may be related to tornadoes associated with tropical cyclones which are observed

to have a different relation with environmental parameters on synoptic time scales (Schultz and Cecil 2009; Edwards et al. 2012). The index has substantially fewer tornadoes than reported in the Southwest during the period May-July and indicates too many tornadoes in the Northwest especially during the months November-June. The index has roughly the correct phasing in the West but with positive biases in winter and early spring.

An overall measure of the similarity between the observed and index climatological spatial patterns is given by their monthly pattern correlation shown in Fig. 10. The lowest pattern correlation values occur in late summer and early fall, with the minimum occurring in September irrespective of whether the pattern correlation is centered (map average is removed) or uncentered (map average is not removed). The reason for the low pattern correlation values is seen in the spatial distributions of the July-September monthly index and tornado report climatologies (Fig. 11). Both the index and report climatologies show the northward shift of values in July. In August and September, the index weakens somewhat and shifts slightly southward. The behavior of the report climatology is rather different showing substantially less tornado activity than does the index over the central US. This discrepancy is especially striking in September when the index has maximum values in the upper-midwest while the maximum report values are in the eastern and southern seaboard states. This behavior of the index is precisely the reason for the positive bias in the annual cycle noted earlier.

The erroneous spatial structure of the index in August and September concentrated in the northern central US reflects that of the SRH, suggesting that the index response to SRH may be to blame. To understand better how the environmental factors are responsible for the positive bias of the index during the late summer and early fall, we fit the index using data restricted to the box 33N to 42N and 100E to 90E. Figure 12a shows the annual cycle of tornado reports and the annual cycles of two indices: the index using coefficients estimated from all the data ("US coef.") and an

index using coefficients estimated from the box data ("box coef."). The report annual cycle shows a much sharper decline in tornado activity in August than does the US index. On the other hand, the behavior of the box index is more similar to that of the report data. The box index coefficients of cPrcp and SRH are 1.41 and 4.36, respectively, indicating that while the regional sensitivity to cPrcp is similar to its all-US value, the regional sensitivity to SRH is more than double its all-US value. Figure 12b shows seasonal cycle of box-averaged cPrcp and SRH. Solid lines show the isolines of the all-US index and dashed lines those of the box index. The isolines show that the index value in July using the all-US coefficients is between that of April and May. Increasing the sensitivity of the index has the effect of increasing the slope of the isolines. The isolines of the index with box coefficients show that the value of the index in July is close to that of March, which is a more realistic result. Roughly speaking, the increased sensitivity to SRH results in a more vigorous annual cycle with enhanced maximum spring values and a more abrupt decline in late summer. This differing sensitivity to SRH may be due to either time-averaging or neglected factors. We do not believe that the sensitivity to SRH varies by location, all other factors being the same.

5. Interannual variability

The CONUS-summed index values computed with interannually varying NARR data were shown to correlate well with total numbers of CONUS reported tornadoes on a monthly as well as on annual basis (Tippett et al. 2012). We assess the relative importance of the two environmental parameters for characterizing interannual variability by computing the index using climatological values of one of the parameters and interannually varying values of the other parameter, and then computing the correlation of the resulting index with reported tornado numbers. Table 1 shows the Pearson and rank correlations between CONUS sums of index values and reported numbers

of tornadoes by calendar month. During most months of the years, the index computed with climatological SRH and interannually varying cPrcp has nearly the same correlation level with CONUS totals as does the full index. On the other hand, when only annually varying SRH is included in the index, the resulting correlation is insignificant in the majority of months. Just as in an overall sense cPrcp explained more of the annual cycle variability, it also explains more of the interannual variability. Only in May and June does the inclusion of interannually varying SRH lead to a marked increase in the correlation. This finding suggests that during the peak activity period both factors contribute to interannual variability, a result with important implications for prediction. First, Tippett et al. (2012) showed that, on average, monthly predictions of cPrcp had lower skill than those of SRH in initialized coupled GCM forecasts. Second, accurate prediction of peak season variability requires accurate forecasts of both cPrcp and SRH. Table 1 also shows the corresponding correlation when the index is constructed using CAPE rather than cPrcp; the values are somewhat lower, especially in April.

To assess the ability of the index to represent regional tornado activity, we computed the monthly and annual number of tornadoes for each of the nine NOAA climate regions and compared the resulting time series with the corresponding index values. Regional Pearson and rank correlations on a monthly and annual basis are given in Tables 2 and Table 3, respectively. Regions and months averaging less than 1 tornado per year are omitted. The South, Southeast and Central regions average more than one tornado per month throughout the year, and significant skill is seen in most months with August - October tending to have poor skill depending on region and skill measure. Deficiencies in explaining the annual cycle seem to impact the representation of interannual variability. Regional correlations are generally lower than CONUS ones reflecting increased noise due to reduced averaging. Correlation of annual values are generally less than those of monthly values since the correlation of the annual total is negatively impacted by temporally

varying biases in mean and amplitude. Even in the Central and Upper Midwest where there is a mean bias, the correlation is still fairly good. The correlation values for the index computed with observed parameters is presumably an upper bound for the skill of forecasts based on this index since forecast skill is limited by the imperfect relation between index and tornado reports as well as the ability to predict the parameters.

6. Summary and conclusions

We have examined the properties of a recently developed index (Tippett et al. 2012) designed to represent the expected monthly number of US tornadoes as a function of monthly-averaged convective precipitation (cPrcp) and storm relative helicity (SRH) taken from the North American Regional Reanalysis. Here we have examined index and its construction in more depth, including aspects of the environmental parameter selection, systematic deficiencies and regional behavior. While the convective available potential energy (CAPE) appears as a factor in many tornado indices, we find here that CAPE does not fit the log-linear functional form of the Poisson regression, and cPrcp takes its place as an indicator of potential updraft strength. This use of cPrcp adresses the initiation issue, at least in in a climatological sense (Trapp et al. 2009), but introduces the complication that the detailed features of cPrcp are sensitive to model convective parameterization schemes.

Pooling all locations and months of the calendar year, we find that the index favorably represents the climatological dependence of monthly tornado numbers on cPrcp and SRH. The index does fail to account for significant number of presumably non-supercell tornadoes in Colorado and Florida, as well as modest numbers of cool-season tornadoes in southern California that occur when according to the index the values of cPrcp, and implicitly instability, are too low. The index also indicates that SRH values are adequate for small numbers of tornadoes to occur west of the

Rockies when none are reported.

The contributions of the two environmental parameters to the index are mostly independent, both with respect to annual cycle and spatial distribution. The annual cycle of the index and of the reported tornado numbers show similar phasing though the index fails to capture the peak magnitude in May. The May peak of the index can be inferred from the relative phases of the annual cycles of SRH and cPrcp considered separately. In May, cPrcp is increasing and already fairly large, and SRH though declining from its winter peak is still large. In terms of the climatological spatial distribution, cPrcp serves to favor the southern part of the US and suppresses the index west of the Rockies and over elevation. SRH strongly enhances the Central US and counteracts the role of cPrcp in the southeast. These findings only apply to the monthly climatology and may be less relevant for day to day variability.

The largest deficiency in the annual cycle of the index occurs in late summer over the Central US where it indicates a greater number of tornadoes than are reported. We found that this behavior can be explained in terms of the sensitivity of the index to SRH and that when the index was fit using only data from this region, the sensitivity to SRH was more than doubled. Increasing the sensitivity of the index to SRH resulted in the index having a more vigorous annual cycle with a larger spring peak value and a more rapid decline in late summer.

The index demonstrates some ability to represent the interannual variability of the number of US tornadoes per month. During most months, cPrcp explains more of this variability than does SRH. However, both factors are important during the peak spring period. The regional variability of the index at the scale of the NOAA climate regions captures aspects of both annual cycle and interannual variability.

Acknowledgments.

MKT and JTA are supported by grants from the National Oceanic and Atmospheric Administration (NA05OAR4311004 and NA08OAR4320912), the Office of Navel Research (N00014-12-1-0911) and a Columbia University Research Initiatives for Science and Engineering (RISE) award. AHS and SJC acknowledge support from NOAA Grant NA08OAR4320912. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

REFERENCES

Brooks, H. E., , 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.

Brooks, H. E. and C. A. Doswell, III, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56**, 191 – 201.

Brooks, H. E., C. A. Doswell, III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618.

Cameron, A. C. and F. A. G. Windmeijer, 1996: R-Squared measures for count data regression models with applications to health-care utilization. *J. Bus. Econ. Stat.*, **14**, pp. 209–220.

Davies, J. M. and R. H. Johns, 1993: *The Tornado: Its Structure, Dynamics, Prediction, and Hazards*, chap. Some wind and instability parameters associated with strong and violent tornadoes. 1: Wind shear and helicity, 573–582. No. 79 in Geophys. Monogr., Amer. Geophys. Union.

Doswell, C. A., III, H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563.

Edwards, R., A. R. Dean, R. L. Thompson, and B. T. Smith, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part III: Tropical cyclone tornadoes. *Wea. Forecasting*, **27**, 1507–1519, doi:10.1175/WAF-D-11-00117.1.

Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2011: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, doi:10.1175/WAF-D-11-00008.1.

Hanstrum, B. N., G. A. Mills, A. Watson, J. P. Monteverdi, and C. A. Doswell, 2002: The cool-season tornadoes of California and Southern Australia. *Wea. Forecasting*, **17**, 705–722, doi: 10.1175/1520-0434(2002)017⟨0705:TCSTOC⟩2.0.CO;2.

Karl, T. R. and W. J. Koss, 1984: Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983. Historical climatology series 4-3, National Climatic Data Center, Asheville, NC. 38 pp.

Kounkou, R., G. Mills, and B. Timbal, 2009: A reanalysis climatology of cool-season tornado environments over southern Australia. *Int. J. Climatol.*, **29**, 2079–2090, doi:10.1002/joc.1856.

Lee, S.-K., R. Atlas, D. Enfield, C. Wang, and H. Liu, 2012: Is there an optimal ENSO pattern that enhances large-scale atmospheric processes conducive to tornado outbreaks in the United States? *J. Climate*, **26**, 1626–1642, doi:10.1175/JCLI-D-12-00128.1.

Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133–142, doi:10.1175/1520-0493(1976)104⟨0133:AEOTPW⟩2.0.CO;2.

McCullagh, P. and J. A. Nelder, 1989: *Generalized Linear Models*. Chapman and Hall, London.

Mesinger, F. and Coauthors, 2006: North American Regional Reanalysis. *Bull. Am. Meteor. Soc.*, **87**, 343–360.

Mestre, O. and S. Hallegatte, 2009: Predictors of tropical cyclone numbers and extreme hurricane intenisties over the North Atlantic using generalized additive and linear models. *J. Climate*, **22**, 633–648.

Monteverdi, J. P., C. A. Doswell, and G. S. Lipari, 2003: Shear parameter thresholds for forecasting tornadic thunderstorms in northern and central California. *Wea. Forecasting*, **18**, 357–370.

Rasmussen, E. N. and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164.

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.

Schultz, L. A. and D. J. Cecil, 2009: Tropical cyclone tornadoes, 1950–2007. *Mon. Wea. Rev.*, **137**, 3471–3484, doi:10.1175/2009MWR2896.1.

Thompson, D. B. and P. E. Roundy, 2013: The relationship between the Madden-Julian oscillation and U.S. violent tornado outbreaks in the spring. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-12-00173.1.

Tippett, M. K., S. J. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357.

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**, doi:10.1029/2011GL050368.

Trapp, R. J., N. S. Diffenbaugh, H. E. Brooks, M. E. Baldwin, E. D. Robinson, and J. S. Pal, 2007: Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing. *Proc. Natl. Acad. Sci. (USA)*, **104**, 19 719–19 723.

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

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.

Villarini, G., G. A. Vecchi, and J. A. Smith, 2010: Modeling of the dependence of tropical storm counts in the North Atlantic basin on climate indices. *Mon. Wea. Rev.*, **138**, 2681-2705. doi: 10.1175/2010MWR3315.1.

Wakimoto, R. M. and J. W. Wilson, 1989: Non-supercell tornadoes. *Mon. Wea. Rev.*, **117**, 1113–1140, doi:10.1175/1520-0493(1989)117⟨1113:NST⟩2.0.CO;2.

Weaver, S. J., S. Baxter, and A. Kumar, 2012: Climatic role of North American low-level jets on U.S. regional tornado activity. *J. Climate*, **25**, 6666–6683, doi:10.1175/JCLI-D-11-00568.1.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second verification of the origins of rotation in tornadoes experiment: VORTEX2. *Bull. Am. Meteor. Soc.*, **93**, 1147–1170, doi:10.1175/BAMS-D-11-00010.1.

List of Tables

	Pearson correlation											
Parameters	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
cPrcp:SRH	**0.75**	**0.64**	**0.54**	**0.50**	**0.60**	**0.67**	**0.75**	**0.40**	0.15	0.25	**0.48**	**0.74**
$\overline{\text{cPrcp}}$:SRH	0.24	0.12	0.14	0.32	**0.41**	**0.39**	**0.52**	0.32	-0.16	0.12	0.21	**0.37**
cPrcp:$\overline{\text{SRH}}$	**0.76**	**0.59**	**0.62**	**0.47**	0.30	**0.48**	**0.64**	0.34	0.15	0.25	**0.53**	**0.73**
CAPE:SRH	**0.66**	**0.42**	**0.44**	0.27	**0.50**	**0.50**	**0.62**	0.11	-0.15	**0.38**	0.32	**0.50**
	Rank correlation											
Parameters	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
cPrcp:SRH	**0.73**	**0.55**	**0.56**	**0.55**	**0.69**	**0.72**	**0.63**	**0.50**	0.25	**0.44**	**0.57**	**0.58**
$\overline{\text{cPrcp}}$:SRH	**0.42**	0.04	0.23	0.33	**0.36**	0.35	**0.55**	0.33	-0.13	0.18	**0.37**	**0.40**
cPrcp:$\overline{\text{SRH}}$	**0.74**	**0.61**	**0.59**	**0.47**	**0.40**	**0.52**	**0.50**	**0.36**	0.34	**0.36**	**0.58**	**0.56**
CAPE:SRH	**0.69**	**0.40**	**0.43**	0.31	**0.61**	**0.49**	**0.59**	0.20	-0.08	**0.48**	**0.47**	**0.43**

TABLE 1. Pearson and rank correlation (Spearman's rho) between reported number of tornadoes and North American Regional Reanalysis (NARR) Poisson regression estimates for the period 1979-2010. Correlations significant at the 95% level are indicated by bold font. The rows labeled $\overline{\text{cPrcp}}$:SRH and cPrcp:$\overline{\text{SRH}}$ indicate the use of the climatological cPrcp and SRH, respectively, in the index. The row CAPE:SRH indicates the results for the index based on CAPE and SRH.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
South	**0.66**	**0.51**	**0.52**	**0.69**	**0.50**	**0.47**	**0.57**	0.31	0.12	**0.46**	**0.60**	**0.71**	**0.53**
Southeast	**0.53**	**0.54**	**0.36**	**0.47**	**0.68**	**0.46**	**0.54**	**0.42**	**0.67**	**0.41**	**0.57**	**0.69**	0.30
Central	**0.68**	**0.69**	**0.65**	**0.53**	**0.56**	**0.73**	**0.65**	0.35	**0.42**	0.26	0.28	**0.73**	**0.51**
Upper Midwest	-	-	**0.60**	**0.55**	**0.71**	**0.57**	**0.56**	0.14	**0.54**	**0.56**	-	-	**0.45**
Plains	-	-	**0.63**	**0.58**	**0.80**	**0.53**	**0.81**	**0.49**	**0.55**	0.23	-	-	**0.51**
Northeast	-	-	-	**0.38**	0.13	**0.61**	**0.50**	**0.41**	**0.37**	**0.71**	0.29	-	**0.36**
Southwest	-	-	-	0.21	0.13	**0.37**	0.32	**0.40**	0.02	0.31	-	-	0.22
Northwest	-	-	-	0.03	**0.44**	**0.36**	-	0.07	-	-	-	-	0.15
West	-	**0.49**	**0.60**	-	-	-	-	-	-	-	-	-	0.34

TABLE 2. Correlation between the index and reported number of tornadoes by U.S. climate region and month for the period 1979-2010. Significant correlations are in bold font. Regions and months with less than 32 reported tornadoes during the period are omitted.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
South	**0.55**	**0.50**	**0.50**	**0.63**	**0.59**	**0.46**	**0.48**	0.35	**0.36**	**0.63**	**0.62**	**0.69**	**0.43**
Southeast	**0.68**	**0.39**	**0.46**	**0.61**	**0.64**	**0.45**	**0.38**	0.21	**0.44**	**0.42**	**0.57**	**0.72**	**0.35**
Central	**0.72**	**0.59**	**0.61**	**0.48**	**0.78**	**0.56**	**0.41**	**0.47**	**0.57**	**0.50**	**0.55**	**0.61**	**0.54**
Upper Midwest	-	-	**0.70**	**0.51**	**0.56**	**0.64**	**0.68**	**0.37**	**0.55**	**0.51**	-	-	**0.47**
Plains	-	-	**0.37**	**0.50**	**0.64**	**0.67**	**0.63**	**0.49**	**0.56**	0.31	-	-	**0.55**
Northeast	-	-	-	**0.50**	**0.42**	**0.55**	**0.38**	**0.57**	0.32	0.34	**0.57**	-	0.34
Southwest	-	-	-	**0.51**	0.18	0.32	0.28	0.24	0.17	**0.41**	-	-	0.23
Northwest	-	-	-	-0.10	**0.53**	0.35	-	0.13	-	-	-	-	0.22
West	-	**0.42**	**0.64**	-	-	-	-	-	-	-	-	-	**0.36**

TABLE 3. As in Table 2 but for rank correlation.

List of Figures

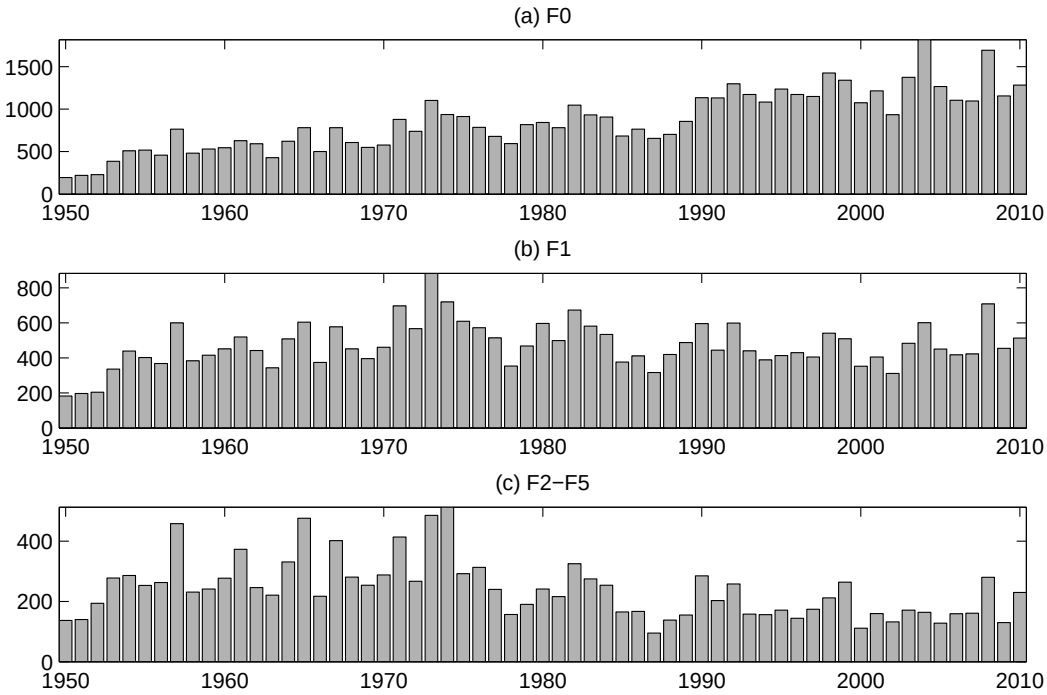

FIG. 1. Annual number of reported (a) F0, (b) F1 and (c) F2-F5 tornadoes for the period 1950-2010.

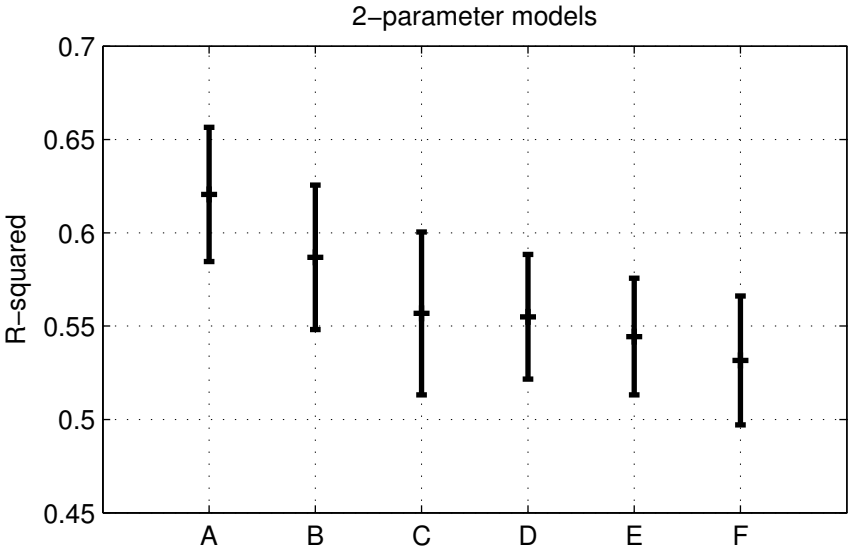

FIG. 2. Deviance-based R-squared values of the 2-parameter models based on (A) cPrcp and SRH, (B) cPrcp and vertical shear (C) cPrcp and mixing ratio (D) CAPE and vertical shear (E) CAPE and mixing ratio (F) CAPE and SRH.

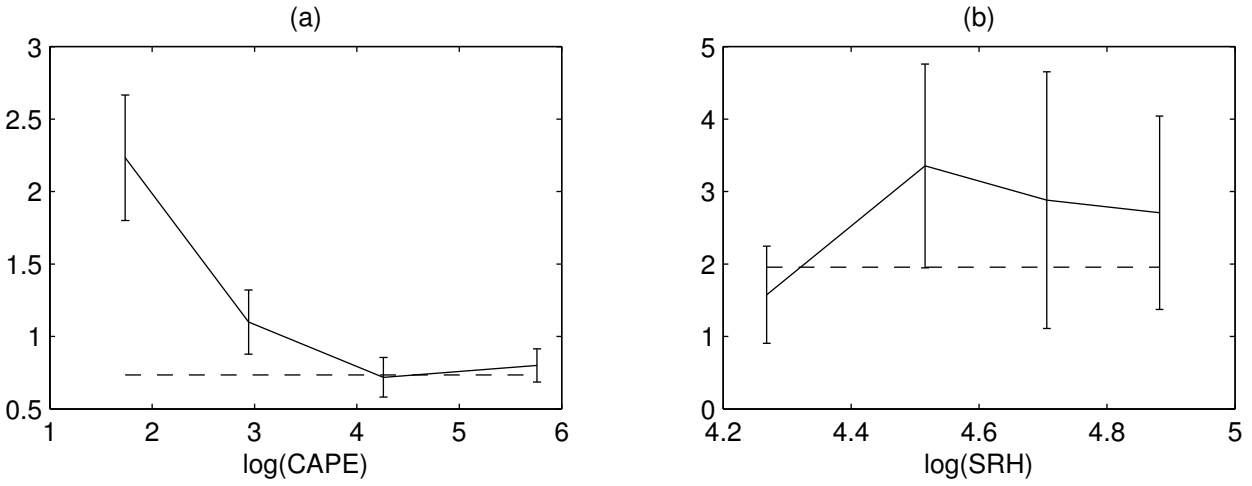

FIG. 3. The Poisson regression coefficients of (a) CAPE and (b) $\log(\mathrm{SRH})$. Error bars show 95% bootstrap estimated intervals and the dashed lines show the Poisson regression coefficient estimated from the complete data. See text for additional details.

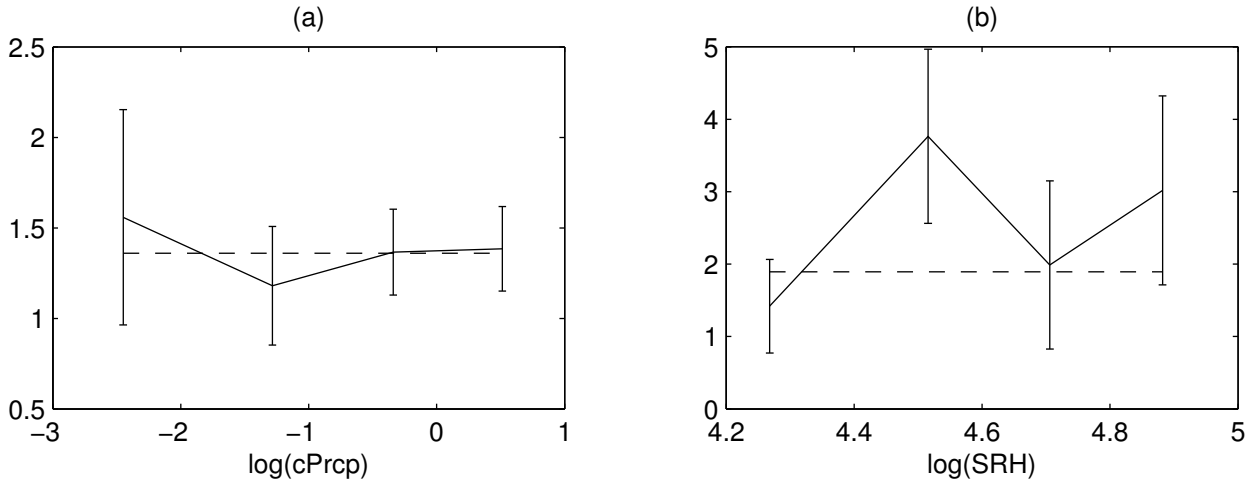

FIG. 4. As in 3 but for (a) cPrcp and (b) $\log(\mathrm{SRH})$.

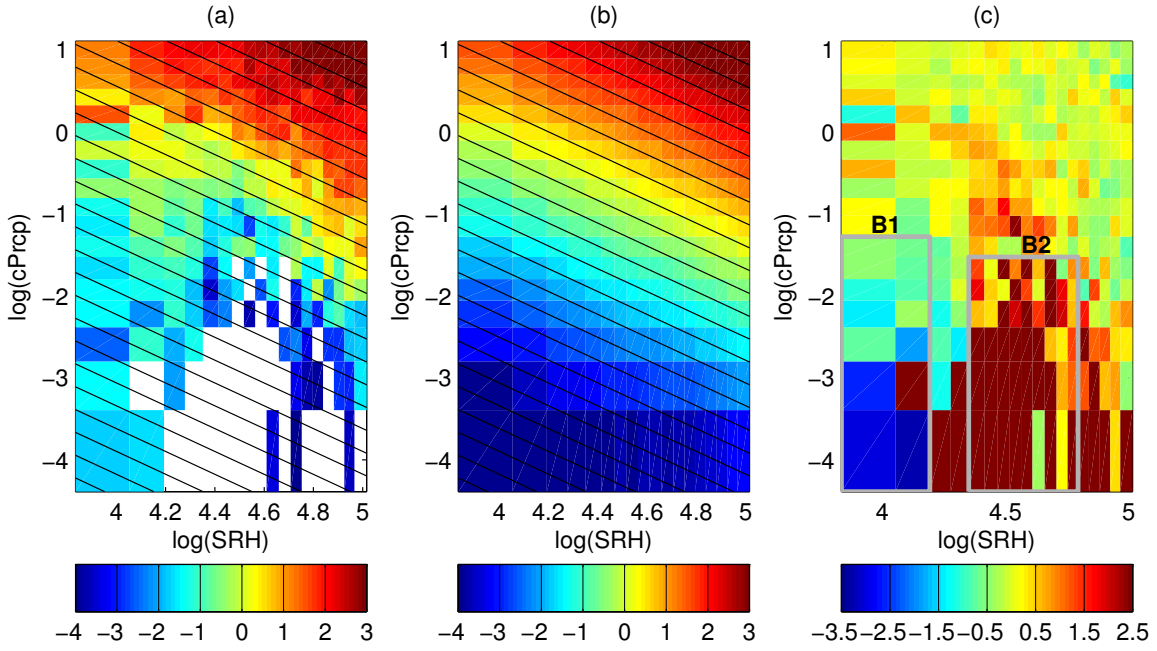

FIG. 5. Logarithm of the average number of tornadoes as a function of $\log(\text{cPrcp})$ and $\log(\text{SRH})$ in (a) observations and (b) the index and (c) the difference observations minus index. Black lines indicate isolines of the index in (a) and (b). White regions in (a) indicate no reported tornadoes. The gray boxes in (c) marked B1 and B2 are described in text.

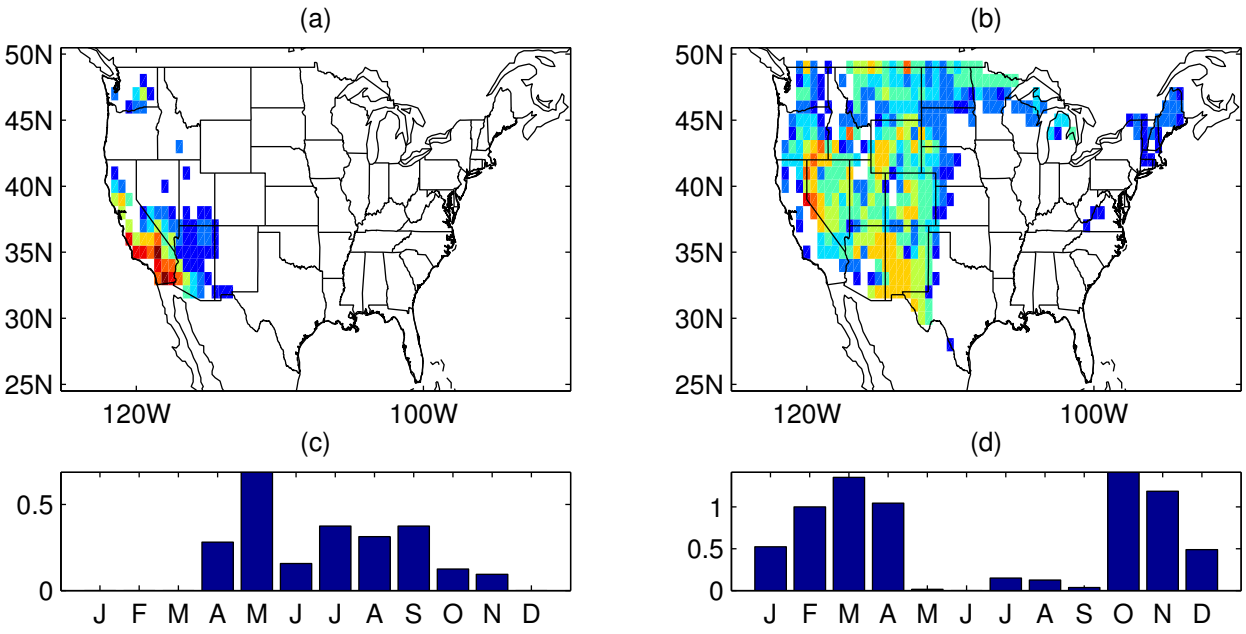

FIG. 6. The spatial and temporal distribution of the data in boxes B1 and B2 of Fig. 5. Panels (a) and (b) indicate the number of months per year that the parameter values of each grid box fall in boxes B1 and B2, respectively. The color red corresponds to 9 months and blue to one month. The number of reported tornadoes per year in box B1 and the corresponding index quantity for box B2 are shown in panels (c) and (d), respectively, by calendar month.

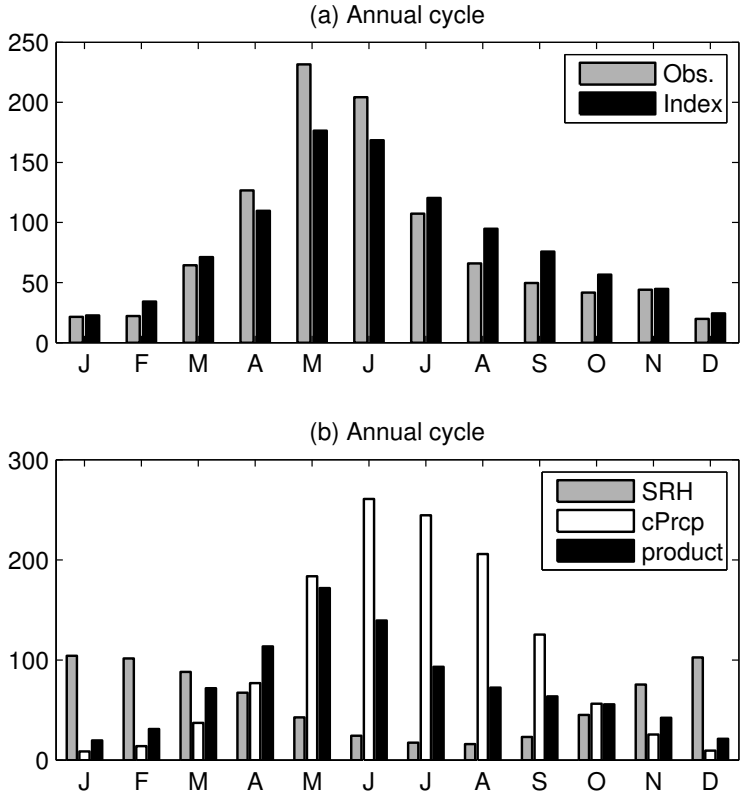

FIG. 7. (a) Annual cycle of the reported number of tornadoes and corresponding index values. (b) Annual cycle of the index factor with monthly varying SRH (gray), the factor with monthly varying cPrcp (white), and the product of the annual cycles of the three factors defined in (3) .

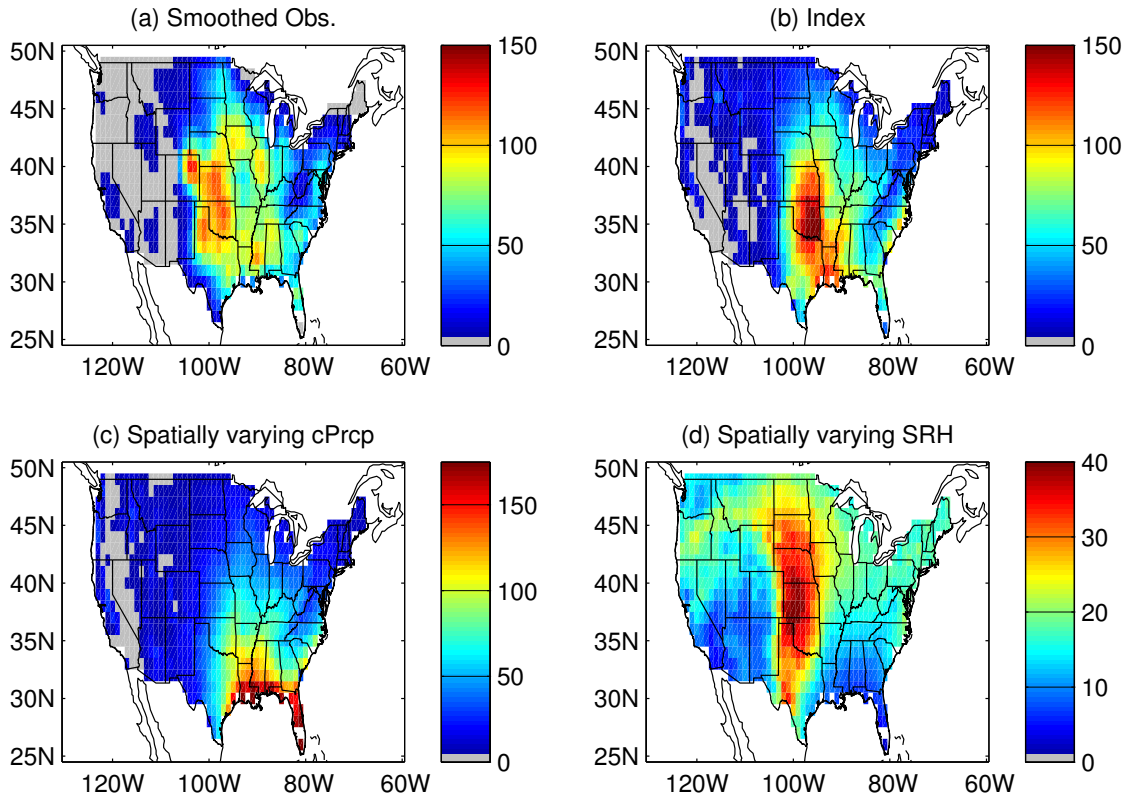

FIG. 8. Spatial distribution of the annual average number of (a) reported tornadoes, (b) the corresponding index values, and the index factors with spatially varying (c) cPrcp and (d) SRH, respectively.

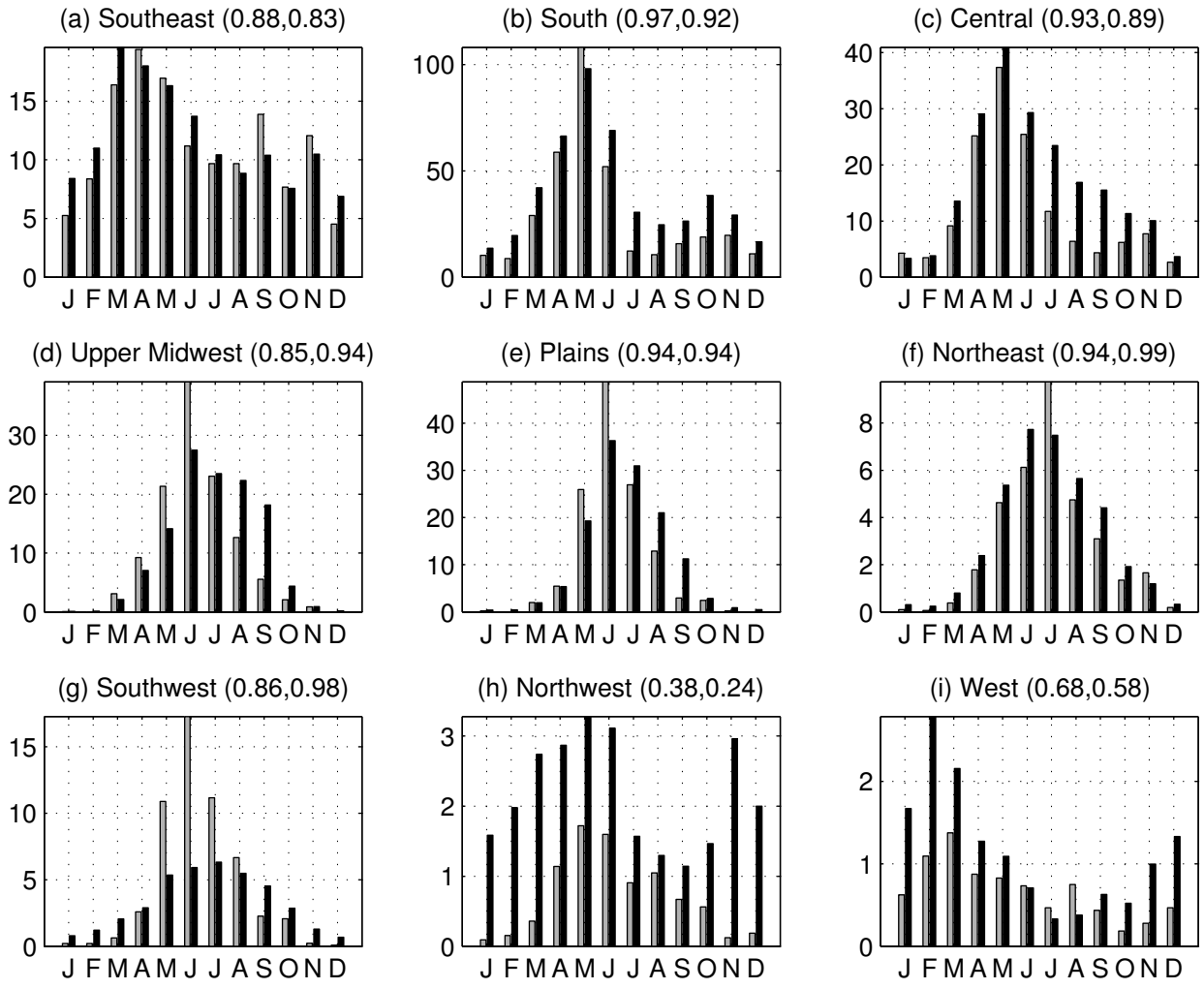

FIG. 9. Annual cycle of the observations (gray) and index (black) for the 9 NOAA climate regions. The values in parenthesis are the Pearson and rank correlations between the observations and index.

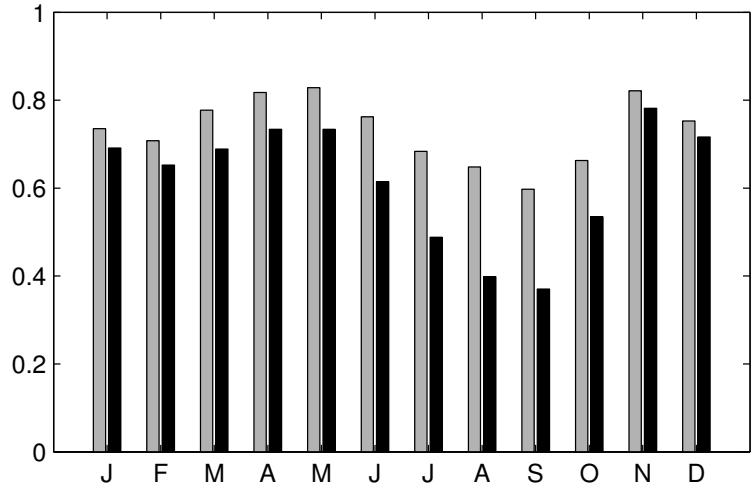

FIG. 10. Pattern correlation between index and observation climatology as a function of calendar month. Gray and black bars indicate uncentered and centered pattern correlations, respectively.

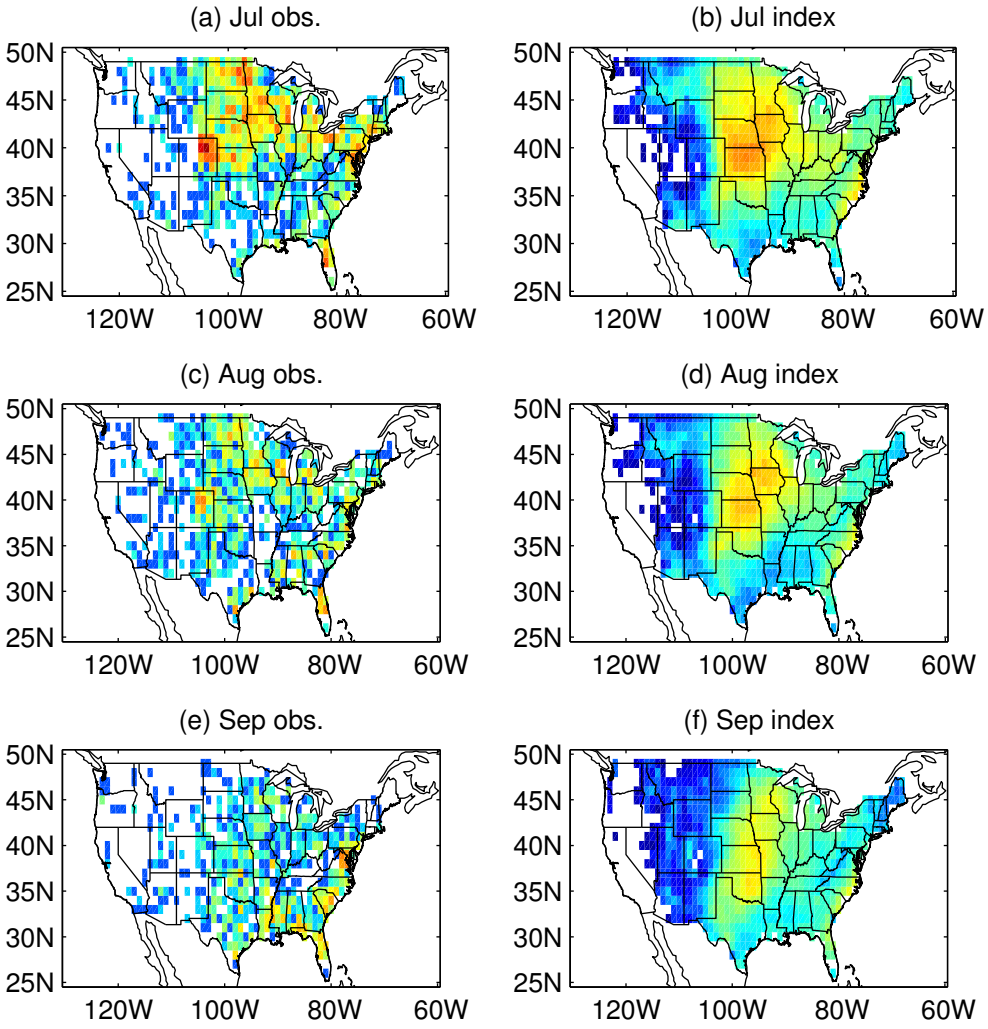

FIG. 11. Logarithm of monthly climatology of (a,c,e) tornado reports and the index (b,d,f) for the months July through September.

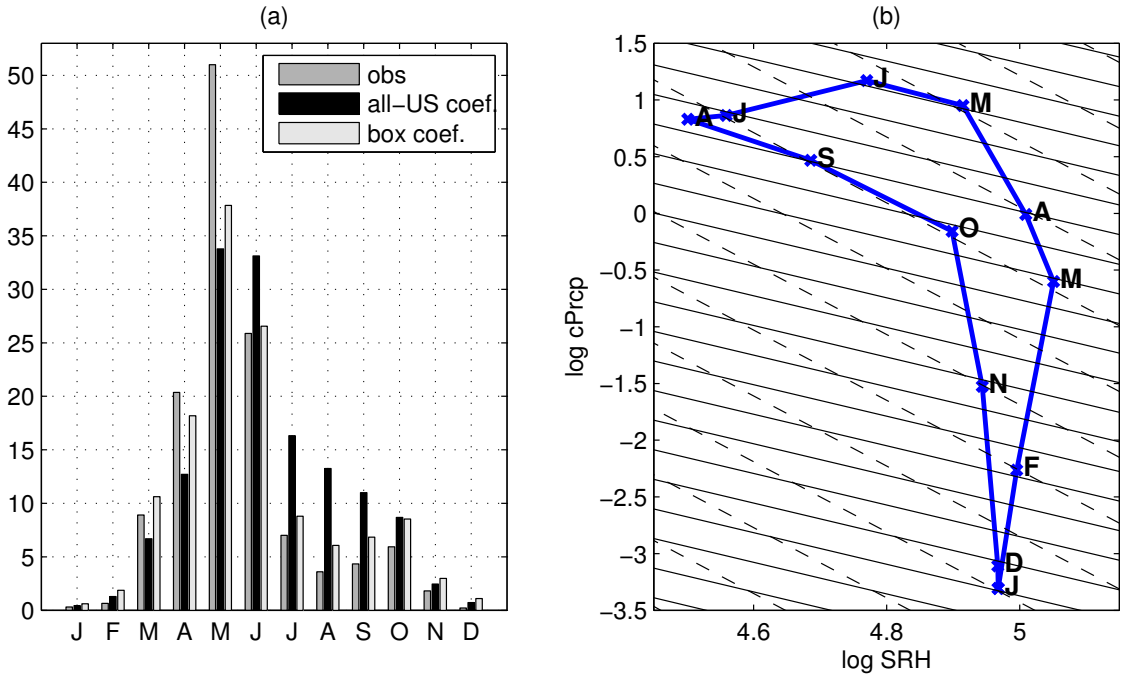

FIG. 12. (a) The annual cycles of reported tornadoes (dark gray), all-US index (black) and local index in the box 33N to 42N, 100E to 90E. (b) The box-averaged values of log cPrcp and log SRH for each calendar month with J, F, M, etc. indicating the month. The solid (dashed) lines are the isolines of the all-US (local) index.