Regional Characterization of Tornado Activity

ROBERT J. TRAPP

Department of Earth, Atmospheric, and Planetary Sciences, and Purdue Climate Change Research Center, Purdue University, West Lafayette, Indiana

HAROLD E. BROOKS

National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 4 July 2012, in final form 27 November 2012)

ABSTRACT

In the United States, tornado activity of a given year is usually assessed in terms of the total number of human-reported tornadoes. Such assessments fail to account for the seldom-acknowledged fact that an active (or inactive) tornado year for the United States does not necessarily equate with activity (or inactivity) everywhere in the country. The authors illustrate this by comparing the geospatial tornado distributions from 1987, 2004, and 2011. Quantified in terms of the frequency of daily tornado occurrence (or “tornado days”), the high activity in the South Atlantic and upper Midwest regions was a major contributor to the record-setting number of tornadoes in 2004. The high activity in 2011 arose from significant tornado occurrences in the Southeast and lower Midwest. The authors also show that the uniqueness of the activity during these years can be determined by modeling the local statistical behavior of tornado days by a gamma distribution.

1. Introduction

In the United States, tornado activity for a given year is usually assessed in terms of the total number of human-reported tornadoes and by maps of the geographical locations and tracks of each event (Fig. 1). Other information, such as the number of tornado-induced fatalities and the estimated costs resulting from damage to property, can be used to characterize the socioeconomic impact.

Comparison with a reference level is somewhat problematic because of the acknowledged biases and clear trend in the historical record of tornado reports (see Fig. 1) (Brooks et al. 2003, hereinafter BDK03; see also Diffenbaugh et al. 2008). Nonetheless, one popular gauge of interannual tornado activity is the annual number of reported tornadoes relative to some (arbitrary) mean. So, for example, 1987 and 2004 were years of extremely low (662) and high (1841) numbers of reports, respectively, relative to a 10-yr (2001–10) mean of 1321.

Such assessments fail to account for the well-known regional dependence of tornado occurrence and thus for the seldom-acknowledged fact that an active (or inactive) tornado year for the United States does not necessarily equate with activity (or inactivity) everywhere in the country. In this regard, a geographically aggregated assessment does little to help in understanding the physical basis for the activity in the context of internal climate variability and even in terms of anthropogenic climate change.

One purpose of this paper is to introduce a simple technique to characterize the tornado activity of a given year, season, or month relative to a climatological mean (see section 2). Key to this technique is the use of binary tornado occurrence counts on day of year and within spatial grid cells. Although still based on reports, such tornado days provide a measure that is less susceptible to the possibility of nonphysical biases and trends (see Fig. 1) and is uninfluenced by single-day tornado outbreaks (e.g., BDK03; Changnon 1982). Thus, high activity here relates to large numbers of tornado days per time interval rather than simply to large numbers of tornadoes. We show, not surprisingly, that even during a perceived highly active year there is significant regional variation.
Another purpose of this paper is to introduce an approach to quantify the uniqueness of the tornado activity of a given year, month, or season. As described in section 3, this involves expression of tornado days relative to the cumulative probability associated with the local gamma distribution. Using the events of April and May of 2011 for illustration, we reach a quantitatively justified conclusion that the tornado activity in the southeastern United States during April of 2011 was highly anomalous.

2. Geospatial representation

Geospatial representation of tornado occurrence is based on the method of BDK03. In accord with this approach, we begin with a calculation of the mean tornado occurrence on a given day within predefined grid cells:

$$d_{x,y,n} = D_{x,y,n}/N, \quad (1)$$

where \((x, y)\) refers to the gridpoint location, \(D_{x,y,n}\) is the number of years of tornado occurrence on day of year \(n\), and \(N\) is the number of years of consideration, herein equal to 30. Note that \(d_{x,y,n}\) gives the mean tornado days at location \((x, y)\); this is equivalent to the probability of tornado occurrence on day \(n\) at \((x, y)\). Following the method in BDK03, we make no attempt to remove non-physical biases or trends from \(D_{x,y,n}\), because these are convolved with the real data in an unknown way. While we cannot claim that our tornado-day data are stationary, the slope of the linear fit to the 30-yr time series of consideration (Fig. 1) is a mere 0.34 days per year, which fails tests of statistical significance at the 95% confidence interval.

Owing to the spatial and temporal intermittency of tornadoes, the geospatial representation of tornado days benefits from smoothing. Following the method of BDK03, Gaussian kernel smoothing in time and space is applied to \(d_{x,y,n}\):

$$c_{x,y,n} = \sum_{k=1}^{366} d_{x,y,n} \exp\left[-\frac{1}{2} \left(\frac{n-k}{\sigma_t}\right)^2\right] \quad \text{and} \quad (2)$$

$$C_{x,y,n} = \sum_{i=1}^{J} \sum_{j=1}^{I} c_{x,y,n} \exp\left[-\frac{1}{2} \left(\frac{d_{ij}}{\sigma_s}\right)^2\right], \quad (3)$$

where \(k\) is a day-of-year index and \(\delta_{ij}\) is the Euclidean distance between the analysis \((x, y)\) and surrounding

![Fig. 1. Example of locations of tornadoes that occurred in the United States during 2004 (figure provided through the courtesy of the National Oceanic and Atmospheric Administration Storm Prediction Center). The inset shows the time series of total reported tornadoes and total tornado days over the period of 1980–2009.](image-url)
FIG. 2. Smoothed mean number of tornado days per (a) year, (b) April, and (c) May over the period of 1980–2009.

FIG. 3. Tornado-day departures from the annual mean in (a) 1987, (b) 2004, and (c) 2011.
grid points within the domain defined by $I$ and $J$ grid points. When used to generate analyses of annual tornado days, the temporal and spatial smoothing parameters are assigned values of $s_t = 15$ days and $s_s = 120$ km, respectively, following the testing and justification provided in BDK03; for analyses of monthly tornado days, we use $s_t = 2$ days on the basis of our testing with analytically defined data.

The procedure in Eqs. (1)–(3) is applied to the National Centers for Environmental Prediction 221 grid, which has a horizontal gridpoint spacing of approximately 32 km (exactly at 40°N latitude) over a North American domain. The use of this particular grid accommodates our parallel work that seeks to relate tornado activity to regional atmospheric forcing, as derived from North American Regional Reanalysis (NARR) data. The period of analysis is 1980–2009 ($N = 30$), which also coincides with NARR availability.

Calculation of the 30-yr mean monthly or annual tornado days $C_{x,y}$, hereinafter referred to as the tornado “climatology,” follows from

$$C_{x,y} = \frac{1}{(m_2 - m_1 + 1)} \sum_{n=m_1}^{m_2} C_{x,y,n},$$

using the relevant intervals of $m_1 \leq n \leq m_2$, that is, for mean annual occurrence, $m_1 = 1$ and $m_2 = 366$; for mean monthly occurrence, $m_1 (m_2) =$ day of year of the beginning (ending) of month. The annual and monthly (April and May) climatological distributions shown in Fig. 2 highlight the known tornado frequency within the central United States, with especially high May occurrence in the Great Plains states of Texas, Oklahoma, and Kansas (e.g., BDK03).

As mentioned, the current work is motivated by the desire to compare the geospatial distribution of tornado days over a single year with that of the long-term mean. Accordingly, the procedure in Eqs. (1)–(4) is repeated for a specific year (and hence with $N = 1$), with the result, denoted here as $S_{x,y}$, used to compute the anomaly, or departure of this single year from the climatological mean:

$$A_{x,y} = S_{x,y} - C_{x,y}. \quad (5)$$

Figure 3 illustrates well the previous statement regarding the geospatial distribution of tornado activity from year to year. The year 1987 was characterized by anomalously low activity over most of the country, especially in the tornado-prone regions. Much of the tornado occurrences during the extremely active year of 2004 were contributed from regions outside of the Great Plains. In particular, the anomalously high activity along the East Coast was a result of numerous landfalling tropical cyclones and the tornadoes they induced. The strongly positive anomaly in the southeastern United States in 2011 was associated with the record numbers of tornadoes in April of 2011.

3. Statistical modeling

The information in Fig. 3 would, by itself, appear to be very useful in assessing the tornado activity during 2011 and other years. An open question, though, is how best
to characterize the uniqueness of this activity relative to the climatological mean. One could readily compute (as we have) the sample variance with an assumption of normally distributed data and then form conclusions that are based on this moment. However, as demonstrated by the distribution of April tornado days at grid points near Oklahoma City, Oklahoma; St. Louis, Missouri; and Tuscaloosa, Alabama, during the 30-yr period (Fig. 4), the daily occurrence data are nonnegative, tend to be highly skewed (toward low occurrences), and therefore are not represented well by the normal distribution.

We attempt here to model the statistical behavior of tornado days by a gamma distribution (e.g., see Wilks 2006). The gamma scale parameter $\beta$ and shape parameter $\alpha$ are fit to the smoothed 30-yr data ($C_{x,y}$ for individual years, seasons, and months) at each grid point, using the Thom (1958) approximation to the maximum-likelihood estimator. Goodness-of-fit tests (Lilliefors test with the Kolmogorov–Smirnov test statistic; see Wilks 2006) of the distributions shown in Fig. 4 pass at least at the 10% level and thereby support the use of the gamma distribution. Following the approach of Husak et al. (2006), instances of $C_{x,y} = 0$ are necessarily excluded from the estimator calculations; at least 24 data values are required herein for the local fits (at least 80% of the years during the 30-yr period). Inspection of the resultant shape parameters (not shown) from the local fits reveals larger values and thus relatively more symmetrical distributions throughout the Great Plains.

Fig. 5. Anomalies in tornado days in (a) April and (c) May 2011, and the (b) April and (d) May 2011 tornado days relative to the cumulative probability associated with the local gamma distribution. In (b) and (d), the analyses are masked at all points at which the smoothed daily occurrence was zero on 7 or more years during the 30-yr period and hence at which the gamma-distribution fits are considered to be compromised by less-than-adequate data (see text).
further confirming the above conclusions about the locations of highest tornado occurrence.

Similar to what is done routinely for monthly summaries of precipitation (e.g., Blunden et al. 2011), we express the monthly tornado days relative to the cumulative probability associated with the local gamma distribution. The events of April and May of 2011 provide a good illustration of this approach. The tornado-day anomalies are particularly strong and positive in the southeastern United States in April of 2011; these are reflected in gamma probabilities as high as 0.9 throughout Alabama, as well as in North Carolina, Virginia, and West Virginia (Figs. 5a,b). In May of 2011, the strongest positive anomalies are shown along the Illinois–Indiana border, Wisconsin, Pennsylvania, and southwestern Missouri. A negative anomaly spans the Texas Panhandle. Referring to the climatological distributions in Fig. 2, it should not be too surprising that the Wisconsin anomaly was associated with the unique (most extreme) activity in May of 2011, relative to the local gamma distribution.

4. Closing comments

We have focused herein on the frequency of daily occurrence of all tornadoes, regardless of (damage based) intensity. We are applying our method to tornadoes that are stratified by Fujita/enhanced Fujita (F/EF) scale, which will allow us to draw conclusions about spatiotemporal characterizations as a function of intensity. The method could, in principle, also be applied to other severe-thunderstorm phenomena such as hail.

We foresee numerous applications to our general approach. For example, these characterizations of activity allow for an alternative assessment of impact and quantification of future risk. They also provide a means to constrain statistical correlations for diagnoses, such as between tornado occurrence and modes of climate variability. Along related lines, this approach is facilitating our work that seeks to relate tornado activity to regional atmospheric forcing and, in turn, attribution of climate extremes. Our approach also contributes to our ongoing efforts to develop long-lead (monthly and seasonal) predictions of high-impact convective weather.

Acknowledgments. This research was supported in part by NSF ATM-0756624 (RJT) and contributes to the Clouds, Climate, and Extreme Weather initiative at Purdue University. This is Purdue Climate Change Research Center paper 1246.

REFERENCES