

LETTER • **OPEN ACCESS**

Examining the changes in the spatial manifestation and the rate of arrival of large tornado outbreaks

To cite this article: Niloufar Nouri and Naresh Devineni 2022 *Environ. Res. Commun.* **4** 021001

View the [article online](#) for updates and enhancements.

You may also like

- [Advances in applications of the physics of fluids to severe weather systems](#)
Howard B Bluestein
- [Daily tornado frequency distributions in the United States](#)
J B Elsner, T H Jagger, H M Widen et al.
- [ARE GIANT TORNADOES THE LEGS OF SOLAR PROMINENCES?](#)
Sven Wedemeyer, Eamon Scullion, Luc Rouppe van der Voort et al.

Environmental Research Communications



LETTER

Examining the changes in the spatial manifestation and the rate of arrival of large tornado outbreaks

OPEN ACCESS

RECEIVED

11 November 2021

REVISED

24 January 2022

ACCEPTED FOR PUBLICATION

1 February 2022

PUBLISHED

10 February 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Niloufar Nouri^{1,2} and Naresh Devineni^{1,3} ¹ Department of Civil Engineering, The City University of New York (City College), New York, NY 10031, United States of America² Department of Data Science, The City University of New York (City College), New York, NY 10031, United States of America³ NOAA/Center for Earth System Sciences and Remote Sensing Technologies (CESSRST), City University of New York, New York, NY 10031, United States of AmericaE-mail: nnouri@ccny.cuny.edu and ndevineni@ccny.cuny.edu**Keywords:** tornado outbreak, extreme events, spatiotemporal analysis, clustering, return periodSupplementary material for this article is available [online](#)**Abstract**

This study presents an assessment of the spatial and temporal characteristics of large tornado outbreak (LTOs) days, in which several counties were impacted by tornadoes rated F2 (EF2) or greater on the Fujita (Enhanced Fujita) scale in one day. A statistical evaluation of changes in the LTO clusters for two periods, 1950–1980 and 1989–2019, has been performed. There is a geographical shift of the nucleus (central impact location) towards the southeast United States. This spatial shift is also accompanied by reduced spatial variance, suggesting LTOs have become less dispersed (or more localized) in the recent period. The overall inter-arrival rate of LTOs, and how it changed during successive 31-year climatological blocks between 1950–2019 was investigated using an exponential probability model. The arrival rate has changed from 124 days during 1950–1980 to 164 days during 1977–2007 and remained relatively constant during later periods, indicating that LTOs are becoming less frequent.

1. Introduction

Tornadoes are one of the most devastating severe weather events in the United States (US) that have posed a risk to human life and resulted in extensive property damage. The US has experienced an annual average loss of 5.4 billion dollars due to the severe thunderstorms accompanied by tornadoes [1]. Tornado outbreaks, defined as sequences of several tornadoes closely spaced in time, routinely rank among the deadliest severe weather events and cause billion-dollar losses [2, 3].

Several studies illustrate that the annual number of reliably reported US tornadoes have remained relatively constant; however, there is a decrease in the number of days per year with tornados [1, 4–6]. This suggests that tornadoes are beginning to cluster more in time [7]. According to Brooks *et al* [4], while there is a lower chance of a day having one tornado, there is a much higher probability of having many tornadoes if a day experiences a tornado. Elsner *et al* [8] study on tornado outbreaks (rated F1/EF1 or higher) also confirmed an increasing trend in the risk of having multiple tornado days with densely concentrated clusters. Tippett *et al* [3] also analyzed tornado outbreaks (rated F2/EF2 or higher) during 1977–2014 and showed that the mean number of tornadoes per outbreak and its variance had increasing trends.

Recent studies also indicate geographical shifts of tornadoes over time [5, 7, 9, 10]. For instance, Agee *et al* [9] compared tornado activity (rated F1/EF1 or higher) between the two periods 1954–1983 and 1984–2013 over a gridded domain covering the central region of tornado activity (80°–105°W, 30°–50°N). They showed that the maximum gridded tornado counts have shifted from the Great Plains to the Southeast. They also observed that the new center of violent tornado days (rated F3/EF3 or higher) is in northern Alabama and Tennessee.

In this study, we present a distinct analysis of the changing spatial manifestation and the inter-arrival rate of large tornado outbreaks, which impacted several US counties on any given day. Tornado outbreaks have been

previously defined as six or more tornadoes (F1 or greater) that are closely spaced in time [11]. We propose an alternate identification scheme to determine only the strong to violent outbreaks that impacted several counties simultaneously in a day. We term these ‘large tornado outbreaks (LTO).’ A systematic analysis of such simultaneous large tornado outbreaks in the principal region of tornado activity covering 28 US states is conducted.

We examine large tornado outbreaks’ temporal properties, including their inter-arrival rates, seasonality, and quasi-periodic behavior. We illustrate their spatial characteristics and compare the statistics on the number of LTO days and the spatial variation between two (early and late) 31-year periods. Finally, we investigate how their return time has changed over 70 years using an exponential probability model. Understanding spatial and temporal shifts in large tornado outbreaks are of great importance, knowing that the recent Southeastward change could involve regions with a high density of mobile homes and forested terrain. The findings of this study have important implications for tornado risk management in the United States.

2. Methods

2.1. Large tornado outbreaks

Researchers have provided several definitions for the tornado outbreak depending on the requirement of their research. Some of them specified a threshold and provided a measurable characteristic, and some did not [12]. In a number of previous studies [13, 14] the number of tornadoes in an outbreak are typically between 6 to ten, no matter they impacted one county or more. In our study, we considered a new definition which is close to the typical definition of an outbreak, while provides additional information on the affected municipalities which can be later helpful for policy making and insurance agencies.

Several studies also reported that population change and better observational tools contributes to an increasing report of ‘weaker’ tornadoes over the last decades [4, 15, 16], since weaker tornadoes may not leave an observable damage, especially in areas with lower population. Evidence also suggests that (E)F2–(E)F5 tornadoes are more likely to be reported due to their greater intensity and longer duration [17, 18] and there is no secular trend in their frequency [19]. Therefore, we excluded (E)F0–(E)F1 tornadoes and used F2(EF2) or greater as the benchmark for LTOs.

In this study, a large tornado outbreak day (LTO) is characterized by a day when at least eight US counties (among the major tornado impacted States (figure S1 (available online at stacks.iop.org/ERC/4/021001/mmedia))) experienced a category F2(EF2) or greater tornado (a day begins at midnight, 00:00 CST, and the last minute of the day begins at 23:59 CST). We arrived at this LTO measure as follows. Consider $x_{i,t}$ as the indicator for the Fujita or enhanced Fujita scale of the tornado occurrence for a county i and day t . Using $x_{i,t}$, we construct a large tornado outbreak binary matrix $X_{i,t}$ in which the elements are either zero or 1; The matrix elements indicates the occurrence of a tornado in county i , and day t . On a given day/county, if there was an (E)F2 or greater tornado, the corresponding element is one, otherwise it is zero.

$$X_{i,t} = \begin{cases} 1 & \text{if } x_{i,t} \geq F2 (EF2); \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

where $i = 1: 2495$ counties, and $t = 1: 25566$ days

Using $X_{i,t}$, we compute the total affected counties for each day as the sum of the counties that have

$$x_{i,t} \geq F2 (EF2):$$

$$TO_t = \sum_{i=1}^n X_{i,t} \quad (2)$$

Where n is the number of counties among the selected States.

We identify large tornado outbreaks as the days when the number of affected counties exceed the 95th percentile of non-zero TO . Out of the 25,566 days between Jan 1950 to Dec 2019, there are 14548 days where TO is zero (no tornados). Based on our data, we find that the 95th percentile of non-zero TO is eight. There are 186 LTOs, i.e., days when eight or more counties simultaneously experienced an F2/EF2 or greater intensity tornado. The goal is to identify tornado outbreaks that are relatively extreme with regard to impacted counties. The 95th percentile threshold (which will screen approximately 5% of the non-zero TO days as LTOs) provided large enough sample size for the statistical analyses.

Ćwik *et al* [12] recently presented an overview of tornado outbreak definitions since 1950. Our definition of LTO uses intensity and spatial extent as thresholds and provides a planning context for the affected states/counties. We also included the table of affected land area in all LTO days in the supplemental document (table S3).

Given that some tornadoes might happen during 11:00 PM and 1:00 AM, we examined the time of occurrence of all the tornadoes in 186 LTO days and confirmed that out of 2934 tornado hits, there were only 37 tornadoes that happened between 11:00 PM and 1:00 AM (approximately 1%). Therefore, the analysis in this study is not significantly affected by midnight tornadoes.

2.2. Changes in the spatial manifestation of LTOs

Using the filtering algorithm, we found 186 LTO days during 1950–2019. For each LTO, we employed the k-medoids clustering method [20, 21] to identify geographic clusters. The algorithm is more robust in the presence of outliers compared to k-means [21, 22]. We used Duda-Hart hypothesis test to check if there are at least two clusters in each LTO [23]. The test is based on calculating the within-cluster sum of squared errors for two clusters and the overall sum of squared errors when only one cluster exists.

We found that only four LTOs out of 186 LTOs have more than one geographic cluster. These happened on March 12, 1976, Nov 22, 1992, May 4, 2003, and April 27, 2011. For the days with more than one cluster, the optimum number of medoids was determined using the average silhouette width criterion [24]. Subsequently, each LTO is summarized using the number of tornado landfalls (i.e., the number of geo-locations that the tornado struck), centroid(s) of the cluster, latitudinal and longitudinal standard deviations. To explore the climatological changes of LTOs, we separated them into the early 31-years (1950–1980) and the recent 31-years (1989–2019). We investigated the changes in the spatial manifestation using the spatial kernel density estimation method. We also verified the shift between the two periods in the number of LTO days and the spatial variance of the LTO clusters (as summarized by latitudinal and longitudinal standard deviations) using a bootstrap hypothesis test.

2.3. Large tornado outbreak inter-arrival rate

The analysis of LTOs indicated that the time between two consecutive LTOs could vary from one day to 999 days with an average of 137 days. Our objective is to investigate the temporal changes in the LTO inter-arrival rate during the period of study. Assuming that two successive LTOs occur independently and continuously at a constant rate, an exponential probability function can be assumed for the arrival time data. Mathematically, the probability density function is represented as:

$$f(t|\lambda) = \lambda e^{-\lambda t} \quad (3)$$

Where $f(t|\lambda)$ is the probability density function of arrival time, and t is the time between two successive LTOs, and λ is the arrival rate. It should be noted that here we fit ‘return time’ data that is the time between two LTOs, so exponential distribution is an appropriate choice.

We also determined the inter-arrival rate of LTOs in consecutive 31-year blocks during 1950–2019, such that each 31-year block is one year ahead of the next one, i.e., 1950–1980, 1951–1981, 1952–1982, ..., 1989–2019. We calculated the return time for each of these 31-year periods (i.e., the number of days between two successive LTOs) and approximated the inter-arrival rate using an exponential distribution function. λ_T from this model is the estimate of average inter-arrival rate over a 31-year window T . We interpret the trends in λ_T .

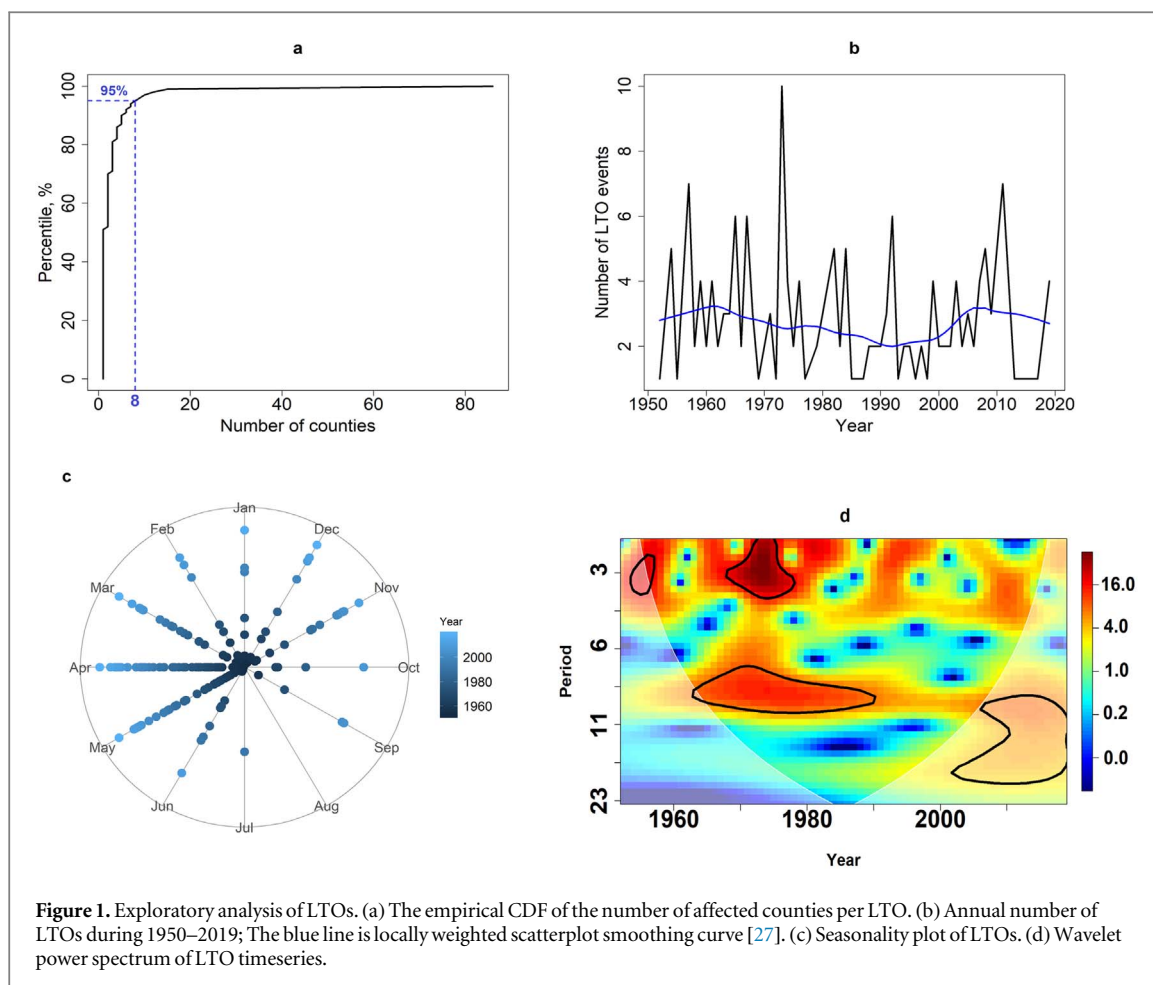
3. Data

Historical tornado records were obtained from the US Storm Prediction Center. This dataset is compiled from the National Weather Service Storm Data publications and reviewed by the US National Climate Data Center [16, 25, 26]. It includes information about the date and time, location, tornado path and length, intensity [Fujita or enhanced Fujita (EF) scale], property losses, crop damages, fatalities, and records of injuries for all tornado incidents in the US during 1950–2019. We only consider data from the major tornado-affected states located in the South, Southeast, Ohio Valley, Upper Midwest, and Northern Rockies. The study region is presented in supplementary figure S1. The final dataset consists of daily tornado data from 1950 to 2019 for 2232 counties across 28 states.

4. Results and discussion

4.1. Exploratory analysis of LTOs

The empirical cumulative frequency distribution of Tornado Outbreaks (TO) is presented in figure 1(a). As mentioned before, TO_t is the basis for identifying LTOs; we identified large tornado outbreaks as days that exceed the 95th percentile of TO . Based on our data, we found that the 95th percentile of TO is eight, and there are 186 LTOs, i.e., days when eight or more counties experienced an F2/EF2⁺ tornado. Figure 1(b) represents the number of LTO days per year. We used locally weighted scatterplot smoothing (LOESS) method to better



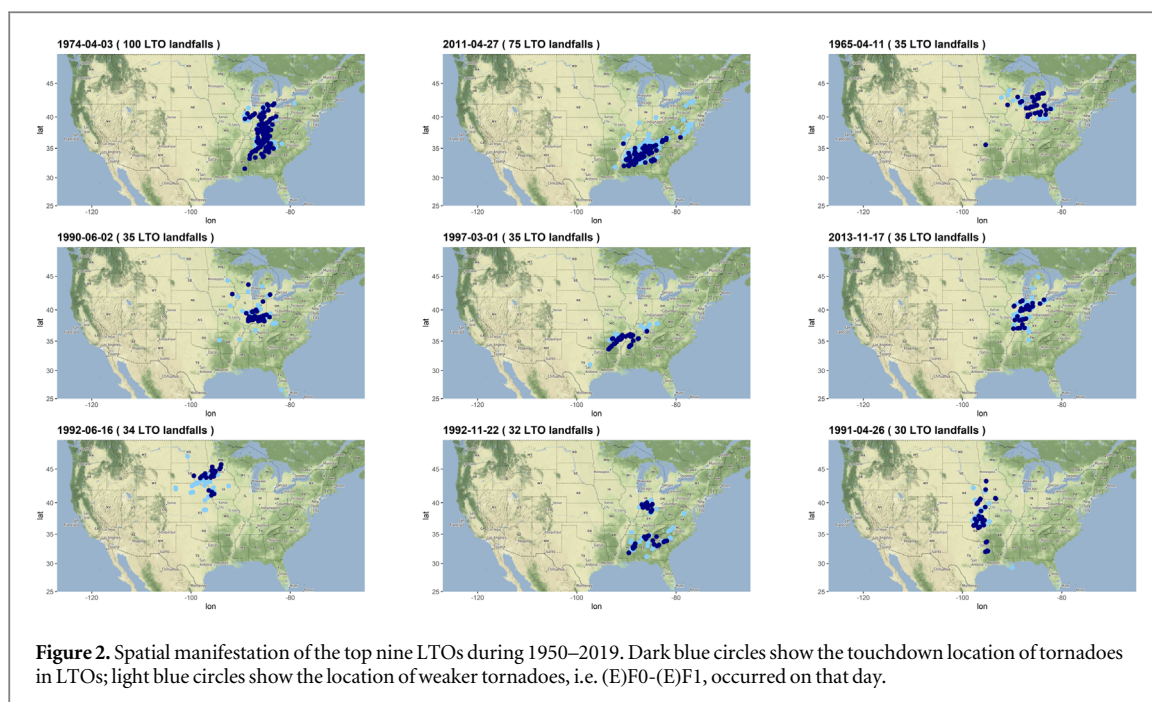
visualize changes in the number of LTOs over time. It is a non-parametric approach that fits multiple polynomial regressions in local neighborhood; The size of neighborhood, i.e., smoothing span, is 0.3 [27]. The average annual LTOs is 3 with a large inter-annual variation (coefficient of variation = 60%). In 1973 and 2011, there were 10 and 7 LTOs, with an average number of affected counties of 13 and 18, respectively. Seasonality plot (figure 1(c)) indicates a strong seasonality during March–April–May (approximately 69% of LTOs), in line with the typical tornado season in these states.

We also investigated the annual time series of LTOs for quasi-periodic behavior using wavelet transformation of the data. Wavelet transforms perform an orthogonal decomposition of the time series using base functions differing in time and frequency resolutions [28, 29]. Figure 1(d) shows wavelet decomposition of annual LTOs presented in figure 1(b). Significant power is exhibited at the decadal time scale. Locally, during 1970–1980, we also found a significant inter-annual mode with 2–5-year periodicity that could be coherent with El-Niño Southern Oscillation. Besides, there is also a significant decadal mode with an 8–10 year periodicity that could be coherence with decal climate oscillations.

The spatial distributions of the top nine LTOs are presented in figure 2. Dark blue circles show the touchdown location of tornadoes in LTOs; light blue circles show the location of weaker tornadoes, i.e. (E)F0–(E)F1, occurred on that day. The date of occurrence and the number of landfalls – geo-locations tornado had a touchdown in – are also shown for each of the top nine LTOs.

Looking into the spatial manifestation of the top nine LTOs, we make the following observations. Three of the LTOs occurred in the southern US, while the rest happened in the Central and North Central US. While most LTOs appear to be localized, a few of them occurred in large clusters. For instance, the LTO on November 22, 1992, had two distinct centers in the Ohio Valley and Southeast. The touchdown locations indicate that they are not always in traditional Tornado Alley or Dixie Alley, the two regions with higher historical records of tornadoes. For instance, in April 1965, 35 tornadoes (rated F2–F5) struck the upper Midwest, including Indiana, Ohio, Michigan, Wisconsin, Illinois, and Iowa. On June 2, 1990, and December 17, 2013, the outbreaks were in the Ohio Valley. The top nine LTOs happened during March, April, June, and December. It is also interesting to note that seven out of the top nine LTOs occurred post-1990.

The most extensive tornado outbreak happened on April 4, 1974, in which 100 tornadoes (rated F2–F5) struck 86 counties across the US. The outbreak extended from the upper-Midwest to the Southeast and involved



ten states, including Illinois, Indiana, Michigan, Ohio, Kentucky, Tennessee, Alabama, Mississippi, Georgia, and North Carolina. It was also the most violent tornado outbreak ever recorded, with around 30 confirmed F4/F5 tornadoes. The tornado outbreaks on April 3 and April 4 with a combined path length of 2,600 mi caused more than \$600 million loss (2020 equivalent is \$3.3 billion) in the US [30, 31]. The LTO that occurred on April 27, 2011, was notable for the most tornado touchdowns in 24 h, with more than 200 tornadoes (including 75 tornadoes greater than EF2). More than half of these touchdowns were recorded in Alabama and Tennessee. It also included four confirmed cases of EF5 tornadoes, which is the highest-ranking possible on the Enhanced Fujita scale and resulted in 316 deaths and approximately 3000 injuries according to the National Climate Data Center [32]. We note that most of the damaging tornadoes have been detected using our LTO measure. We summarized the top nine LTOs' details, including the affected states, number of landfalls, and total number of affected counties in supplementary table S1.

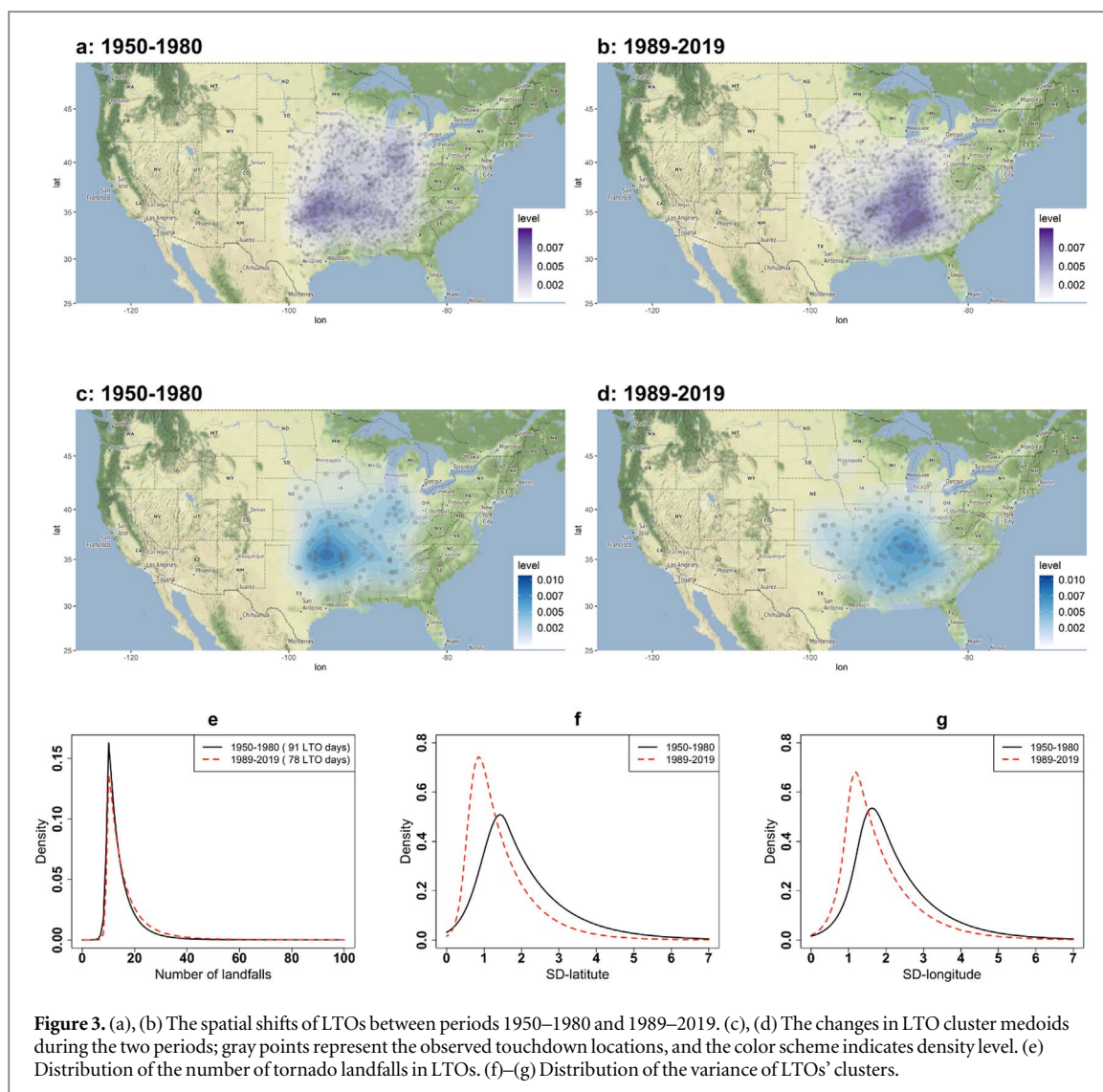
4.2. Spatial characteristics of LTOs

Analyzing LTO spatial density maps revealed an eastward shift of tornado touchdowns. We performed a two-dimensional kernel density estimation with an axis-aligned bivariate normal kernel on LTO's geographical coordinates and compared the density level during the two periods – 1950–1980 and 1989–2019 – in figures 3(a) and (b). We used 100 grid points in each direction for density estimation and selected the bandwidth using the normal reference bandwidth rule [33, 34].

In figures 3(a), (b), we used the kernel density estimator on all the touchdown data that correspond to LTOs. In figures 3(c), (d), we used the kernel density estimator only on the cluster centroids to get a more specified view.

During the early 31-year record (1950–1980), LTOs are primarily concentrated in the south of Great Plain (Oklahoma and Arkansas) and Ohio valley. In the recent 31-years, they are denser in Alabama, Tennessee, and Kentucky. We also observed that the LTO days during the recent 31 years (1989–2019) had a slightly higher density level than the initial years. Multiple studies have also shown evidence of an eastward shift of tornado activity [5, 7, 9, 10]. Our study complements these efforts by analyzing the clusters of LTOs between two 31-year periods and presenting hypothesis testing to verify changes in spatial clusters' characteristics. It is believed that the eastward shift in tornado activity is due to the rising tornado-favorable environments in the Southeast [35, 36]

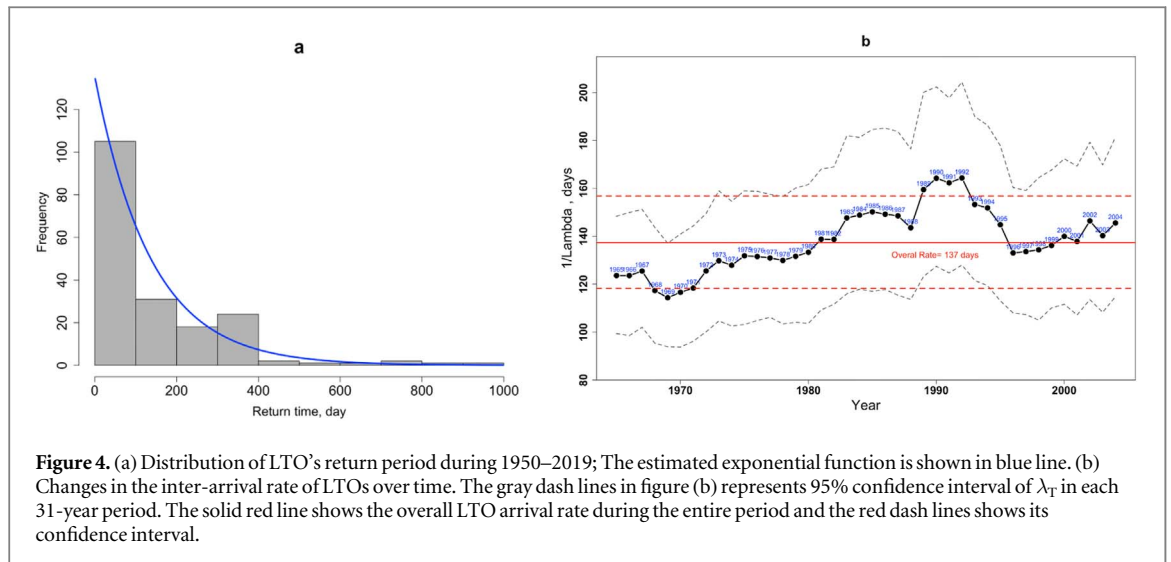
To identify changes in the spatial characteristics of LTO clusters, we implemented the k-medoid clustering algorithm on geographic coordinates of LTO touchdowns and estimated two-dimensional kernel density on the cluster medoids. We see a similar and clearer eastward shifting pattern in LTO's cluster centers (figures 3(c) and (d)). During the early 31-year record, the region with the maximum density of cluster medoids happens in Arkansas, while in the recent 31-year record, it shifted to Tennessee. In addition to examining the shift in the overall number of tornado landfalls and the cluster medoids, we also examined the clusters' latitudinal and



longitudinal standard deviations between the two periods to understand LTO's spatial dispersion changes over time.

In figures 3(e)–(g), we present the probability density functions of tornado landfalls, longitudinal and latitudinal standard deviations of the early 31-year and recent 31-year LTOs, respectively. A logsplines density estimator is used for this purpose [37]. We can see that the distributions of tornado landfalls (i.e., the number of geo-locations that the tornado had a touchdown in) during the early 31-year and recent 31-year are similar (see figure 3(e)). However, the LTO clusters' variance during the two periods is different in both longitude and latitude directions. The changes in the peak and tail of the distribution of clusters' standard deviation suggest that LTOs are getting less dispersed (more localized) between the two periods 1950–1980 and 1989–2019. We run a significance test (using the bootstrap method with 10,000 resamples) to compare the mean and tail of the distribution of clusters' standard deviation (latitude and longitude directions) between the two periods. The *p-values* of the null hypothesis for comparing mean and 90th percentile of clusters' standard deviation in latitude direction were $1e-04$ and 0.030, respectively. They were 0.002 and 0.03 respectively in the longitudinal direction. The null hypothesis that there is no change in the distributions is rejected.

The hypothesis test results confirm that the spatial dispersion of the clusters is significantly different between the two periods 1950–1980 and 1989–2019. The cluster size has shrunk and got more concentrated lately. A recent study [7] used a standard deviation ellipse to measure tornadoes' geographical dispersion. By calculating ellipsoid characteristics, the method captured 50%–85% of tornadoes that are normally distributed around the mean center. Their analysis showed that tornadoes were becoming less dispersed over time, particularly in seasons with more tornado outbreaks. In our study, we only focused on LTOs and used an unsupervised learning approach that finds clusters of tornadoes based on how close the touchdown geo-coordinates are, and then analyzed spatial shift using clusters' variance. Our results support Moore's 2019 [7] study and reveal that changes



in spatial dispersion of tornadoes can also be extended to the tornado outbreaks when strong tornadoes simultaneously impact several counties.

To explore the impact of LTO definition (i.e., the 95th percentile threshold), we repeated the analysis using a different threshold (90th percentile). Using the 90th percentile threshold, the number of LTO days is about 400. However, the top nine events which are mapped in figure 3 would remain the same. We also included the map of spatial shift in the supplemental document (figure S2), which confirmed that there is not much of a distinction between the two thresholds.

4.3. LTO return time

LTO temporal analysis shows that their return time (i.e., the inter-arrival period) varies between 1 and 999 days, with a mean arrival rate of 137 days. To our best of knowledge there has not been a detailed analysis of tornado outbreak return time and arrival rate during the period of records. However, some studies might have briefly mentioned the return time of tornado for a particular region as part of an example. For instance, Farney and Dixon [5] focused on the years with high-frequency tornadoes and reported that during 1960–1989 the return time of high-frequency tornado-year is once every five years.

The histogram in figure 4(a) shows the distribution of the return time (the number of days between two successive LTOs) during the entire period (1950–2019). More information about the return time has been included in table S2 of the supplemental document. We found that 39% of the LTOs have a return time between 1–30 days, and 19% of LTOs happened within one week of their previous outbreak.

To identify the temporal changes in the LTO's inter-arrival rate, we estimated the arrival rate of LTOs for 40 consecutive 31-year blocks during 1950–2019. As discussed in the methodology, each 31-year block is one year ahead of the next one. We calculated the return times for each of these periods and approximated the inter-arrival rates, λ_T , using an exponential distribution function. Figure 4(b) presents the time-changing inter-arrival rate, λ_T , of LTOs, along with the overall inter-arrival rate (λ) and its confidence interval. The overall arrival rate of LTOs is 137 days, with 95% confidence interval of the rate parameter extending to ± 20 days.

The 31-year window T , is labeled using the middle year in figure 4(b). For instance, the first data point shows λ_T for 1950–1980, where the middle year is 1965. The LTO return period analysis shows that the arrival rate has changed from 124 days ($\lambda_T = 0.0081$) during 1950–1980 (labeled as 1965) to 164 days ($\lambda_T = 0.0061$) during 1977–2007 (labeled as 1992), indicating that the LTOs were becoming less frequent between the two periods. The LTOs were more frequent during the next four 31-year blocks, as evidenced by the decreasing arrival time. It was approximately 133 days during 1981–2011 (labeled as 1996) and remained relatively constant during the later periods. Considering the 95% confidence interval of the overall arrival rate (137 days), we can say that the changes in return time have mostly fallen into the uncertainty margins except for four successive 31-year periods between 1975 and 2007, including 1975–2005, 1976–2006, 1977–2007, in which the return time was maximum, indicating less frequent LTOs during these periods. It should be noted that the findings about LTO return time is not contradictory to some of the previous findings [3, 4], as the focus of current analysis is the time between two LTOs.

Knowing that during 1975–2007 warm phase of climate variability in the North Pacific was dominant [38], it is our conjecture that the higher inter-arrival rate of LTOs could relate to the warm phase of Pacific Decadal Oscillation (PDO). PDO is defined by the leading pattern of climate variability in the northeast and tropical

Pacific Ocean. Positive PDO corresponds to the anomalously cool sea surface temperature and the bellow average sea level pressures is in North Pacific [38]. Spencer [39] also suggests that the positive PDO phase is associated with fewer intense tornadoes (EF3-EF5) in the US during the mid-1970s to mid-2000s. Moreover, Elsner *et al* [40] showed the connection between a positive phase of the North Atlantic Oscillation (NAO) and a lower tornado activity across southeastern states. Another recent study [41] also showed that NAO negatively influences annual tornado frequency. Knowing that the winter index of NAO (December to March) during mid-1970s to early-2000s was primarily in its positive phase [42], we can also associate less frequent LTOs with the positive phase of NAO. However, additional study using a causality framework is needed to verify the connection between decadal climate oscillations and arrival rates of LTOs.

5. Discussion

In this study we provided a systematic analysis of tornado outbreaks that impacted several US counties in one day. Our approach is unique in representing a broad range of analysis including trend, seasonality, quasi-periodic behavior, clustering behaviors, geographical shifts and the inter-arrival rates for high-strength tornado outbreaks.

The difference between previous studies and our study can be explained by: (1) the changes in the period of study (we analyzed tornadoes during the entire available records) and the definition of tornado outbreak, (2) the comprehensive exploratory analysis on LTO days, and (3) the use of statistical learning methods for studying spatial shifts.

Previous studies have provided several definitions for the tornado outbreak depending on the intended purpose of the research; Some of them specified a threshold and provided a measurable characteristic, and some did not [12]. In this study, by incorporating the counts of impacted US counties into the LTO definition, we offered additional insight in terms of affected municipalities which can be later helpful for policy making and insurance analysis. The inherent characteristics of each individual LTO-day in our study can be later linked to the unique characteristics of its affected counties (e.g., the built-in area and population exposed to risk, infrastructures, local adaptation system and warnings) and used for studying socioeconomic risks.

As discussed in the introduction, several studies suggested that the chance of having more clustered tornadoes is increasing [4, 8]. These studies drew their conclusion mainly by exploring the trends in the tornado counts or tornado days in different geographical regions. The spatial analysis in the previous works primarily focused on counting the frequency of events per geographical grids. In our study, we examined the spatial characteristics using a statistical learning method (clustering), i.e., using an unsupervised learning approach for automatic calibration of the changes. Our analysis not only shows the centroid of LTO clusters during two district periods, but also provided statistical significance test on detected changes in the distribution of landfalls and spatial variance of clusters.

6. Conclusions

We investigated the (E)F2⁺ tornado outbreaks, called large tornado outbreaks - LTOs, which simultaneously struck several US counties in a day. Seasonality analysis of the LTO days revealed that most LTOs (69 percent) happened in March-April-May, and there was no record during August. Wavelet transformation of the LTO time series illustrated significant power at the decadal time scale during the 1970–1980 and a strong inter-annual cycle potentially relevant to El-Niño Southern Oscillation.

Spatial analysis of LTOs reveals that their nucleus has been shifting to the Southeast during the recent 31-years compared to the earlier records. The cluster variance has significantly decreased in both latitude and longitude direction, which suggests a decrease in spatial dispersion. Considering the relocation of LTOs towards the Southeast and Dixie Alley, we can conclude that the risk of tornado outbreaks is becoming more concentrated in these regions. We are currently investigating the potential connection with local environmental factors such as convective available potential energy and other meteorological variables responsible for the change in LTO spatial characteristics [7, 9]. It is also likely that there is a connection between US hurricanes and tornadoes' geographical shift to the Southeast. The key contributing factors in the formation of tornado is warm temperatures and wind shear. According to several studies [43, 44], tornadoes spawning from a hurricane upon landfall is common. For instance, Novlan and Gray [43] reported that 25% of US hurricanes during 1950–1972 spawned tornadoes. A more recent study [45] discussed the environmental ingredients for tornadoes within hurricane Ivan in 2004, which produced around 120 tornadoes across the nine states. However, additional analysis and supporting data is required to draw firm conclusions about the links between geographical shifts of tornadoes and hurricanes in the southeast.

Using an exponential probability model for the LTO return time, we identified the arrival rate of LTOs during 40 successive 31-year periods. Results showed that LTOs became less frequent between 1950–1980 and 1977–2007. The findings regarding the inter-arrival rates of LTOs are presented here for the first time. Additional investigation is needed to explore the connection between large-scale and synoptic variability and the return period of LTOs.

This study's results agree with previous studies on the topic while provide additional valuable information for risk assessment of large tornado outbreaks at the US county scale.

Acknowledgments

This work is supported by the U S Department of Energy Early CAREER Award #DE-SC0018124.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/https://www.spc.noaa.gov/wcm/#data>.

ORCID iDs

Niloufar Nouri  <https://orcid.org/0000-0001-9247-7962>

Naresh Devineni  <https://orcid.org/0000-0003-0311-8752>

References

- [1] Gensini V A and Brooks H E 2018 Spatial trends in united states tornado frequency *npj Climate and Atmospheric Science* **1** 1–5
- [2] Smith A B and Matthews J L 2015 Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates *Nat. Hazards* **77** 1829–51
- [3] Tippett M K and Cohen J E 2016 Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity *Nat. Commun.* **7** 10668
- [4] Brooks H E, Carbin G W and Marsh P T 2014 Increased variability of tornado occurrence in the United States *Science* **346** 349–52
- [5] Farney T J and Dixon P G 2015 Variability of tornado climatology across the continental United States *Int. J. Climatol.* **35** 2993–3006
- [6] Guo L, Wang K and Bluestein H B 2016 Variability of tornado occurrence over the continental United States since 1950 *J. Geophys. Res.* **121** 6943–53
- [7] Moore T W and McGuire M P 2019 Using the standard deviational ellipse to document changes to the spatial dispersion of seasonal tornado activity in the United States *npj Climate and Atmospheric Science* **2** 21
- [8] Elsner J B, Elsner S C and Jagger T H 2015 The increasing efficiency of tornado days in the United States *Clim. Dyn.* **45** 651–59
- [9] Agee E, Larson J, Childs S and Marmo A 2016 Spatial redistribution of U.S. Tornado activity between 1954 and 2013 *Journal of Applied Meteorology and Climatology* **55** 1681–97
- [10] Moore T W and DeBoer T A 2019 A review and analysis of possible changes to the climatology of tornadoes in the United States *Progress in Physical Geography: Earth and Environment* **43** 365–90
- [11] Tippett M K, Lepore C and Cohen J E 2016 More tornadoes in the most extreme U S tornado outbreaks *Science* **354** 1419–23
- [12] Ćwik P, McPherson R A and Brooks H E 2021 What is a Tornado Outbreak? Perspectives through time. *Bulletin of the American Meteorological Society* **102** 817–35
- [13] Galway J G 1977 Some climatological aspects of tornado outbreaks *Mon. Weather Rev.* **105** 477–8
- [14] Grazulis T P 1993 *Significant Tornadoes 1680-1991* (United States: Environmental Films)
- [15] Concannon P R, Brooks H E and Doswell C A 2000 Climatological Risk of Strong and Violent Tornadoes in the United States *Second Conf. On Environmental Applications (Long Beach, California)* (American Meteorological Society) pp 212–9 https://www.nssl.noaa.gov/users/brooks/public_html/concannon/
- [16] Verbout S M, Brooks H E, Leslie L M and Schultz D M 2006 Evolution of the U S tornado database: 1954–2003 *Weather and Forecasting* **21** 86–93
- [17] Anderson C J, Wikle C K, Zhou Q and Royle J A 2007 Population influences on tornado reports in the United States *Weather and Forecasting* **22** 571–9
- [18] Brooks H E 2004 On the relationship of Tornado path length and width to intensity *Weather and Forecasting* **19** 310–9
- [19] Coleman T A, Dixon P G, Coleman T A and Dixon P G 2014 An objective analysis of tornado risk in the united states *Weather and Forecasting* **29** 366–76
- [20] Dunham M H 2003 *Data Mining Introductory and Advanced Topics* (Upper Saddle River, NJ): Prentice Hall/Pearson Education)
- [21] Santhanam T and Velmurugan T 2010 Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points *Journal of Computer Science* **6** 363–8
- [22] Madhulatha T S 2012 An overview on clustering methods *Journal of Engineering* **2** 719–25
- [23] Clarke M R B, Duda R O and Hart P E 1974 Pattern classification and scene analysis *Journal of the Royal Statistical Society. Series A (General)* **137** 442
- [24] Kaoungku N, Suksut K, Chanklan R, Kerdprasop K and Kerdprasop N 2018 The silhouette width criterion for clustering and association mining to select image features *International Journal of Machine Learning and Computing* **8** 69–73
- [25] Elsner J B, Jagger T H, Widen H M and Chavas D R 2014 Daily tornado frequency distributions in the United States *Environ. Res. Lett.* **9** 024018
- [26] Long J A, Stoy P C and Gerken T 2018 Tornado seasonality in the southeastern United States *Weather and Climate Extremes* **20** 81–91

- [27] Cleveland W S 1981 LOWESS: a program for smoothing scatterplots by robust locally weighted regression *The American Statistician* **35** 54
- [28] Daubechies I 1990 The wavelet transform, time-frequency localization and signal analysis *IEEE Trans. Inf. Theory* **36** 961–1005
- [29] Torrence C and Compo G P 1998 A practical guide to wavelet analysis *Bull. Am. Meteorol. Soc.* **79** 61–78
- [30] Kessler E 1983 *The Thunderstorm in Human Affairs* 1 2nd ed. (Norman, Oklahoma: University of Oklahoma Press)
- [31] Risk Management Solutions 2004 *Analysis and Reconstruction of the 1974 Tornado Super Outbreak, RMS Special Report* https://forms2.rms.com/rs/729-DJX-565/images/scs_1974_tornado_super_outbreak.pdf
- [32] National Climatic Data Center 2011 *NOAA Storm Data- 2011 Annual Summaries* <https://www.ncdc.noaa.gov/stormevents/>
- [33] Venables W N and Ripley B D 2002 *Modern Applied Statistics With S* 4th ed. (New York: Springer)
- [34] Silverman B W 2017 *Density Estimation for Statistics and Data Analysis* 1st edn (Boca Raton, FL: Taylor & Francis Group) (<https://doi.org/10.1201/9781315140919>)
- [35] Lee C C 2012 Utilizing synoptic climatological methods to assess the impacts of climate change on future tornado-favorable environments *Nat. Hazards* (<https://doi.org/10.1007/s11069-011-9998-y>)
- [36] Gensini V A and Mote T L 2015 Downscaled estimates of late 21st century severe weather from CCSM3 *Clim. Change* (<https://doi.org/10.1007/s10584-014-1320-z>)
- [37] Kooperberg C and Stone C J 1991 A study of logspline density estimation *Comput. Stat. Data Anal.* **12** 327–47
- [38] Mantua N J and Hare S R 2002 The pacific decadal oscillation *J. Oceanogr.* **58** 35–44 (https://waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/deltaflow/docs/exhibits/nmfs/spprt_docs/nmfs_exh4_mantua_hare_2002.pdf)
- [39] Spencer R W 2011 *The Tornado—Pacific Decadal Oscillation Connection* <http://drroyspencer.com/2011/05/the-tornado-pacific-decadal-oscillation-connection/>
- [40] Elsner J B, Jagger T H and Fricker T 2016 Statistical models for tornado climatology: long and short-term views *PLoS One* **11** e0166895
- [41] Nouri N, Devineni N, Were V and Khanbilvardi R 2021 Explaining the trends and variability in the United States tornado records using climate teleconnections and shifts in observational practices *Sci. Rep.* **11** 1741
- [42] Hurrell J W, Kushnir Y and Visbeck M 2001 The North Atlantic Oscillation *Science* **291** 603–5
- [43] Novlan D J and Gray W M 1974 Hurricane-spawned tornadoes *Mon. Weather Rev.* **102** 476–88
- [44] Mccaul E W 1993 *Observations and Simulations of Hurricane-Spawned Tornadoic Storms* 79 (American Geophysical Union (AGU)) (10.1029/GM079P0119)
- [45] Baker A K, Parker M D and Eastin M D 2009 Environmental ingredients for supercells and tornadoes within hurricane ivan *Weather and Forecasting* **24** 223–44
- [46] Storm Prediction Center 2019 Severe Weather Database Files <https://www.spc.noaa.gov/wcm/#data>