

RESEARCH ARTICLE

Dynamical Analysis and Visualization of Tornadoes Time Series

António M. Lopes¹*, J.A. Tenreiro Machado²

1 Institute of Mechanical Engineering, Faculty of Engineering, University of Porto, Porto, Portugal, **2** Institute of Engineering, Polytechnic of Porto, Porto, Portugal

* These authors contributed equally to this work.

* aml@fe.up.pt



OPEN ACCESS

Citation: Lopes AM, Machado JT (2015) Dynamical Analysis and Visualization of Tornadoes Time Series. PLoS ONE 10(3): e0120260. doi:10.1371/journal.pone.0120260

Academic Editor: Dante R. Chialvo, National Scientific and Technical Research Council (CONICET), ARGENTINA

Received: October 6, 2014

Accepted: January 21, 2015

Published: March 19, 2015

Copyright: © 2015 Lopes, Machado. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in the paper is held by the National Oceanic and Atmospheric Administration (NOAA), National Weather Service, Storm Prediction Center. The data is available online at <http://www.spc.noaa.gov/>.

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In this paper we analyze the behavior of tornado time-series in the U.S. from the perspective of dynamical systems. A tornado is a violently rotating column of air extending from a cumulonimbus cloud down to the ground. Such phenomena reveal features that are well described by power law functions and unveil characteristics found in systems with long range memory effects. Tornado time series are viewed as the output of a complex system and are interpreted as a manifestation of its dynamics. Tornadoes are modeled as sequences of Dirac impulses with amplitude proportional to the events size. First, a collection of time series involving 64 years is analyzed in the frequency domain by means of the Fourier transform. The amplitude spectra are approximated by power law functions and their parameters are read as an underlying signature of the system dynamics. Second, it is adopted the concept of circular time and the collective behavior of tornadoes analyzed. Clustering techniques are then adopted to identify and visualize the emerging patterns.

Introduction

A tornado is a violently rotating column of air extending from a cumulonimbus cloud to the ground. Tornadoes may assume different shapes and sizes and, typically, appear as a funnel with the narrower end touching the ground. Tornadoes' damaging paths can surpass 1.5 km wide and 80 km long and the most violent events can cause huge destruction and fatalities [1]. Tornadoes are classified according to their intensity and damaging effects [2–3]. Earlier events use the Fujita scale [4–6], while most recent tornadoes are often classified using the enhanced Fujita scale, which was adopted in the United States on February 1, 2007 [7]. In both classification systems, the least damaging tornadoes are rated as F0, while the most damaging ones are rated as F5.

Often, Fujita scale relates the damage caused by a tornado to wind speed classes, according to the empirical equation [8]:

$$v(F) = 6.3(F + 2)^{3/2} \quad (1)$$

where v represents wind velocity and F is the Fujita scale value.

Pursuing a clear physical meaning, other indices have been used to quantify tornadoes intensity, namely the maximum horizontal wind speed, the maximum kinetic energy and the maximum energy flux density [8–10]. Schielicke and Névir [11] introduced the concept of atmospheric moment, M_a , analogous to the seismic moment that describes the strength of earthquakes. The value of M_a may be estimated by:

$$M_a \approx AL\bar{\rho}e \tag{2}$$

where, $A = \pi W^2/4$ is the circular area of the tornado on the ground, W and L denote the tornado width and path length, respectively, $\bar{\rho}$ represents the average air density (which is assumed to be equal to 1 kg/m^3) and e is the mass-specific kinetic energy, given by:

$$e = \frac{\langle v(F) \rangle^2}{2} \tag{3a}$$

$$\langle v(F) \rangle = 6.3(F + 2.5)^{3/2} \tag{3b}$$

Where $\langle v(F) \rangle$ denotes the average value of each wind velocity class, obtained by adding the constant 0.5 to each value of F .

Expression (2) implicitly assumes a constant tornado intensity and width, whereas those properties often change considerably during the tornado lifecycle. Despite this drawback, the atmospheric moment, M_a , has the advantage of being analogous to the seismic moment that describes the strength of earthquakes [11]. Such analogy allows comparing tornadoes and earthquakes, which have been widely studied in terms of their space-time statistical distributions [12].

Tornadoes reveal high variability in terms of intensity, geometric properties (e.g., width and path length) and temporal behavior [12], which is an indicator of the existence of fat tail distributions. In fact, tornadoes have features that are well described by power law (PL) functions [11–14] and unveil characteristics that are also found in complex systems [15–17]. This is the case of Earth’s atmosphere, which is a typical example of an open complex dissipative system, with external forcing caused by the differential heating of the solar radiation [11]. While there are no definitive conclusions about the physical significance of PL, some mechanisms generating such distributions are consistent with self-organized criticality (SOC) [18] and highly optimized tolerance (HOT) [19]. SOC is a process in which a system, by itself, converges to an ordered state, characterized by the emergence of a coherent global pattern created by interactions between low-level entities. The HOT conceptual framework explains the formation of scale-invariant patterns in complex systems. It emphasizes that PL behavior results from optimization of tradeoffs between system yield and tolerance to risk, which drives the system to a specific configuration. Other PL generative mechanisms are addressed by Newman [20].

Tornadoes have been investigated from complementary perspectives and using several statistical tools [8, 10, 21–22]. In the last years relevant studies have been published about this topic. In reference [22] it is shown that tornadoes intensities, expressed in terms of wind speed or in Fujita scale, can be described by Weibull distributions, for which the exponential distributions remain as special cases. Dotzek et al. [8] propose that tornado intensities, expressed in terms of the squared maximum horizontal wind speed, can be described by exponential distributions. Brooks [23] studies the statistical relationship between tornado path lengths and Fujita scale intensities. Schielicke and Névir [11] show that tornadoes exhibit PL behavior when intensities are measured by the atmospheric moment, M_a , given by expressions (2)–(4). In reference [16] severe tornadoes and tornado outbreaks are analyzed for the period 1982–2011. The touchdown path length, L , is used to measure tornadoes intensity. The authors demonstrate

that there is a strong linear scaling between the number of severe tornadoes and their total path length in that same year. Moreover, the noncumulative frequency-length statistics of severe tornado path lengths (i.e., $20 < L < 200$ km) is well approximated by an inverse PL function with exponent approximately equal to 3. Schielicke and Névir [12] compare the statistics of earthquakes and tornadoes, both in terms of intensity and temporal behavior. The results demonstrate that the statistics of tornadoes reveal PL behavior when the temporal resolution is high (i.e., between 10 and 60 min). Furthermore, it is shown that the distributions support the observation that tornadoes form clusters in time. Doswell et al. [24] show that tornado outbreaks are characterized by spatio-temporal clustering, meaning that long memory is present as an underlying feature. Moreover, the authors claim that the total path length of severe tornadoes in a convective day, L_D , is the preferred index to quantify the strength of a tornado outbreak. Verbout et al. [25] discuss the number of tornadoes above a given threshold in a convective day as a measure of the strength of a tornado outbreak.

In this paper we study the collective behavior of tornadoes from the perspective of dynamical systems. A public domain database containing the events occurred in the U.S. during the period 1950 up to 2013 is tackled. We model the occurrences as time series of Dirac impulses with amplitude proportional to the index quantifying the size of the events. In this perspective, we are considering the long term dynamics instead of modelling each individual event. The time series are viewed as the output of a dynamical system and, consequently, the data is interpreted as a manifestation of the system dynamics. First, we analyze the annual time series in the frequency domain, by means of the Fourier transform (FT). The amplitude spectra are approximated by PL functions and the parameters are viewed as a signature of the system characteristics. Second, we adopt the concept of circular time and we compute the circular correlation between the annual time series. Based on the maximum correlation and the corresponding time delay, we use clustering and visualization techniques to unveil patterns.

Bearing these ideas in mind, the paper is organized as follows. In section 2 we describe the data tackled in this work. In section 3 we process the tornadoes annual time series in the frequency domain, by means of the FT, and we interpret the results according to the characteristics of the spectra. In section 4 we correlate the annual time series and we propose clustering and visualization tools to capture hidden patterns. Finally, in section 5, we outline the main conclusions.

Dataset

We use the U.S. tornado database compiled by the National Oceanic and Atmospheric administration (NOAA), National Weather Service, Storm Prediction Center. Data is available online at <http://www.spc.noaa.gov/>, containing all reported U.S. tornadoes from year 1950 up to 2013. There are more than 58,000 tornadoes currently in the database. Each event includes, among other features, information about date and time (with one-minute time resolution), Fujita scale, total number of fatalities and injuries, geographic location (i.e., touchdown and liftoff latitude and longitude coordinates), tornado width and path length.

[Fig. 1](#) depicts the geographic location of the events occurred in the U.S., during the time period covered by the database. It can be seen that the East of the Rocky Mountains is particularly prone to tornadoes. This region includes the “Tornado Alley” (i.e., an area centered on Oklahoma, Kansas and northern Texas, and extending from Texas to Canada) and the “Dixie Alley” (i.e., the southern region of the U.S., namely the northern and central parts of Alabama and Mississippi).

[Fig. 2](#) represents the evolution of the number of occurrences versus year, where we can see an increasing trend. However, the statement that we should expect an increasing number of

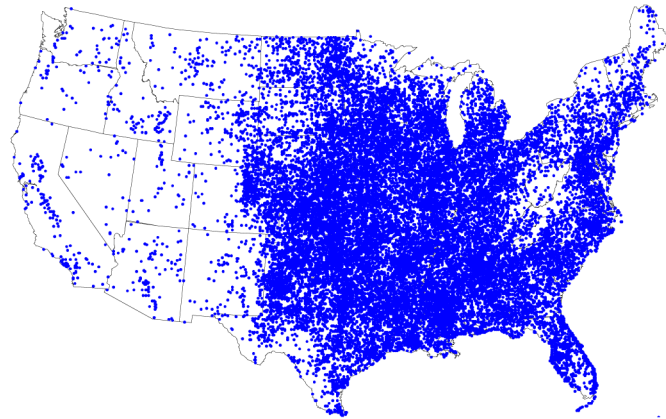


Fig 1. Geographic location of all reported tornadoes that occurred in the U.S., during the time period 1950–2013.

doi:10.1371/journal.pone.0120260.g001

events in the next years is not justified by this simple statistic, as we cannot ignore the fact that there has been a change in the detection capabilities and reporting of the events over time. In fact, in the last decades there has been a more systematic reporting of weak tornadoes over time [21, 23].

The severity of tornadoes, quantified by the total number of affected people (fatalities plus injuries) is represented in Fig. 3, in terms of the complementary cumulative distribution.

Fig. 4 depicts the complementary cumulative distributions of tornadoes path length, L , and width, W , revealing similar behavior in both cases. It should be noted that the use of fatalities

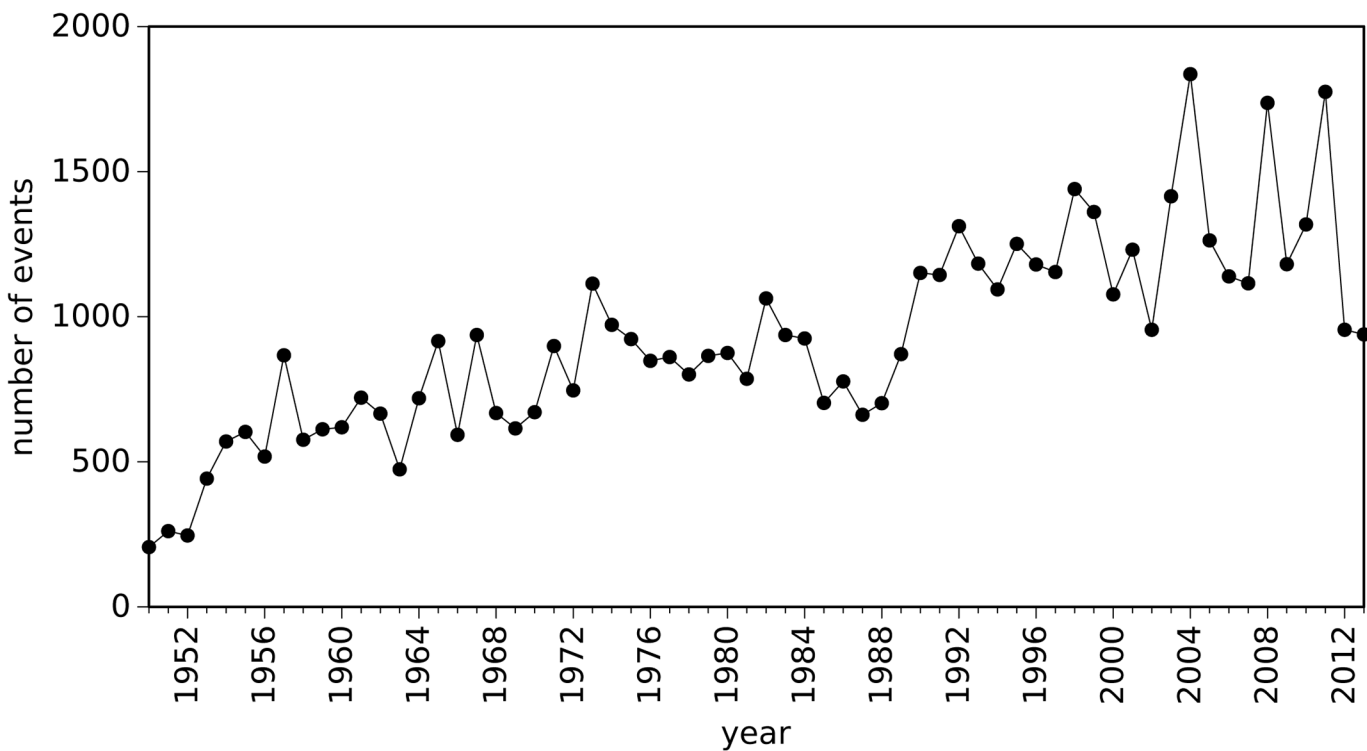


Fig 2. Evolution of the number of reported tornadoes that occurred in the U.S. versus year, during the time period 1950–2013.

doi:10.1371/journal.pone.0120260.g002

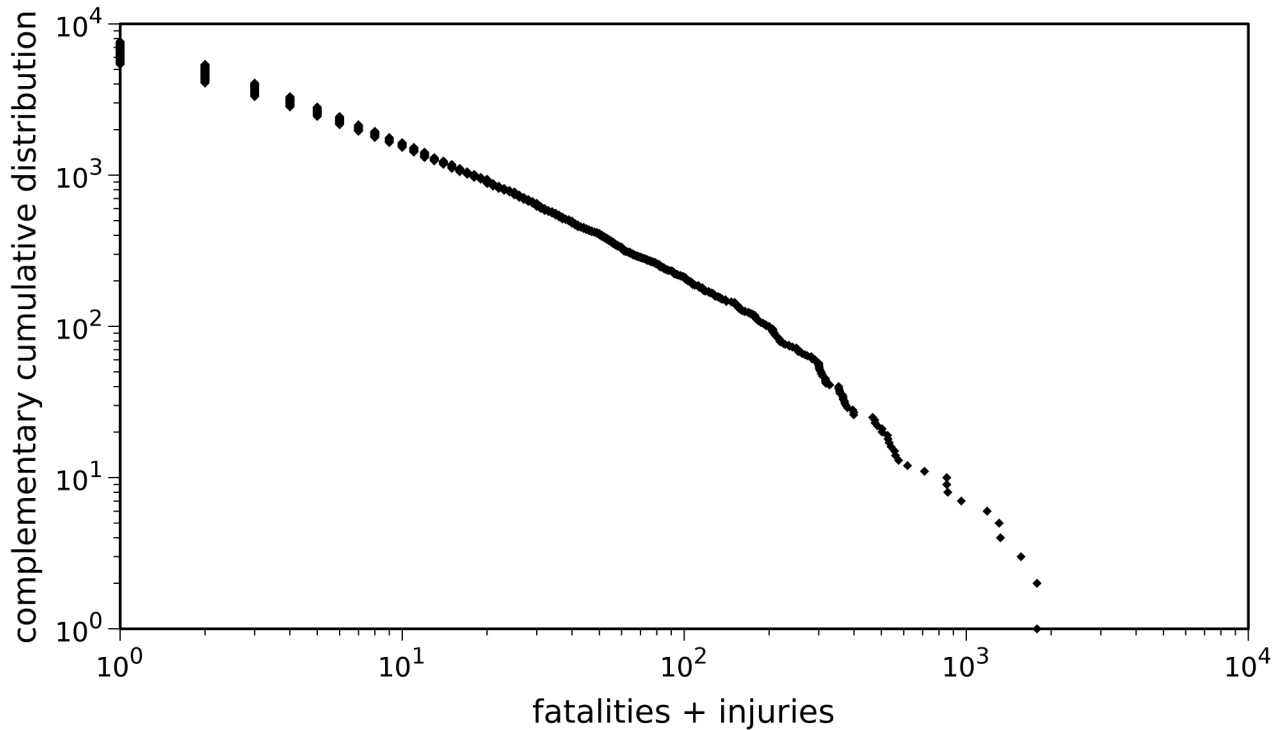


Fig 3. Complementary cumulative distribution of fatalities plus injuries. All reported tornadoes that occurred in the U.S., during the time period 1950–2013 are considered.

doi:10.1371/journal.pone.0120260.g003

and injuries as a measure of tornado intensity is questionable since it is highly influenced by many factors, namely population density, building codes, safety infrastructure, warning systems and awareness.

These direct statistics reveal limitations for characterizing tornadoes global behavior. In the next sections we adopt several complementary mathematical tools for a deeper understanding of the phenomena. The main tools used, namely FT, hierarchical clustering (HC) and

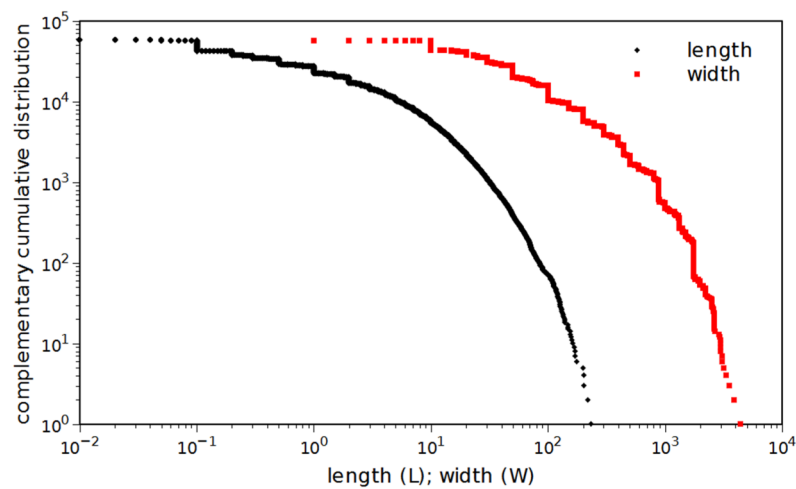


Fig 4. Complementary cumulative distribution of tornadoes path length, L , and width, W . All reported tornadoes that occurred in the U.S., during the time period 1950–2013 are considered.

doi:10.1371/journal.pone.0120260.g004

multidimensional scaling (MDS) rely on long time series that are processed in frequency or time domains. FT is a powerful and robust signal processing tool widely adopted in the area of dynamical systems. FT can “dilute” occasional signal artifacts into the final result. HC and MDS have relevant properties for the analysis of complex phenomena since they do not require a priori assumptions on the system nature. Furthermore, since HC and MDS maps do not need data in a uniform grid in time and space, they constitute valuable tools for studying the non-homogeneous data included in tornado catalogues.

Analysis of Tornado Time Series in the Frequency Domain

In this section we study tornadoes in the perspective of dynamical systems using the Fourier domain. The data is analyzed in an annual basis. For the i^{th} case, $i = 0, \dots, 63$ (cases $i = 0$ and $i = 63$ correspond to years 1950 and 2013, respectively), tornadoes are represented by:

$$x_i(t) = \sum_{k=1}^T A_{i,k} \delta(t - t_{i,k}) \tag{4}$$

resulting in a total of 64 time series. This means that tornadoes are modelled as Dirac impulses, where $A_{i,k}$ represents the size of the events, measured by an appropriate index, $t_{i,k}$ is the time instant of each occurrence (with one-minute resolution, referred to the beginning of year i), t represents time, T is the period of one year and $\delta(\cdot)$ corresponds to the Dirac delta function. With this approach, we are not modelling the dynamics of each particular event. Otherwise, we are describing the tornado global dynamics along several decades.

The signal $x_i(t)$, representative of the collective dynamics of tornadoes, is processed by means of the Fourier transform and the amplitudes of the frequency spectra are approximated by PL functions.

In analytical terms, for the time series signal $x_i(t)$, we have:

$$F\{x_i(t)\} = X_i(j\omega) = \int_{-\infty}^{+\infty} x_i(t) \cdot e^{-j\omega t} \cdot dt \tag{5}$$

where F represents the Fourier operator, ω denotes the angular frequency and $j = \sqrt{-1}$.

The PL approximation is given by:

$$|F\{x_i(t)\}| = |X_i(j\omega)| \approx a_i \cdot \omega^{-b_i}, \quad a_i \in \mathfrak{R}^+, \quad b_i \in \mathfrak{R} \tag{6}$$

The values obtained for b_i reveal underlying characteristics of the systems dynamics.

For fitting the model (6) to the experimental data we tried the least squares algorithm, which is known for generating systematic errors under certain conditions [26], the maximum likelihood estimator (MLE) and related statistical tests [26–27] and a genetic algorithm (GA). We obtained identical values for the PL exponents in the three methods. However, as our experimental data is highly “noisy”, the MLE method yields poor statistical significance p -values. Therefore, no definitive significance tests are made and no absolute conclusions about PL behavior can be drawn.

We present the results obtained with a standard GA, with elitism, crossover within all population and 5% mutation probability. A population of 5,000 individuals and 1,000 iterations are adopted [28]. For the fitness function we use the Canberra distance between the $N = 1000$

experimental data points, $P = p_1, p_2, \dots, p_N$, and the model, $Q = q_1, q_2, \dots, q_N$:

$$J(P, q) = \frac{1}{N} \sum_{i=1}^N \frac{|p_i - q_i|}{|p_i| + |q_i|} \tag{7}$$

Expression (7) leads to good results since, by calculating the ratio between the difference and the sum of two values, it is possible to capture the relative error of the adjustment, avoiding the undesirable effects that occur when using the standard Euclidean norm due to the simultaneous presence of large and small values.

Two indices are used to measure tornadoes intensity: tornado path length, L , and tornado width, W . For example, Fig. 5 depicts the time series of year 2005, $x_{55}(t)$, where amplitude A_{55} represents the corresponding tornadoes path length, L_{55} . Fig. 6 shows the magnitude of the Fourier transform, $|F\{x_{55}(t)\}|$, and the PL approximation. The resulting parameters are $(a_{55}, b_{55}) = (3179.7, 0.330)$, revealing a fractional value of b_{55} . Alternatively, for A_{55} representing tornadoes width, W_{55} , the results are qualitatively similar, yielding parameters $(a_{55}, b_{55}) = (149759, 0.280)$. PL behavior is a characteristic exhibited by long memory systems. Furthermore, Eq. (6) establishes a direct relationship between PL behavior and fractional Brownian motion [29–30], which is a signature of complexity.

Adopting the procedure described above, the PL parameters are calculated for all years $i = 0, \dots, 63$. Furthermore, the PL and the exponential approximations are compared. Fig. 7 depicts the fitness function (7) obtained for the PL and the exponential models, both for tornado path

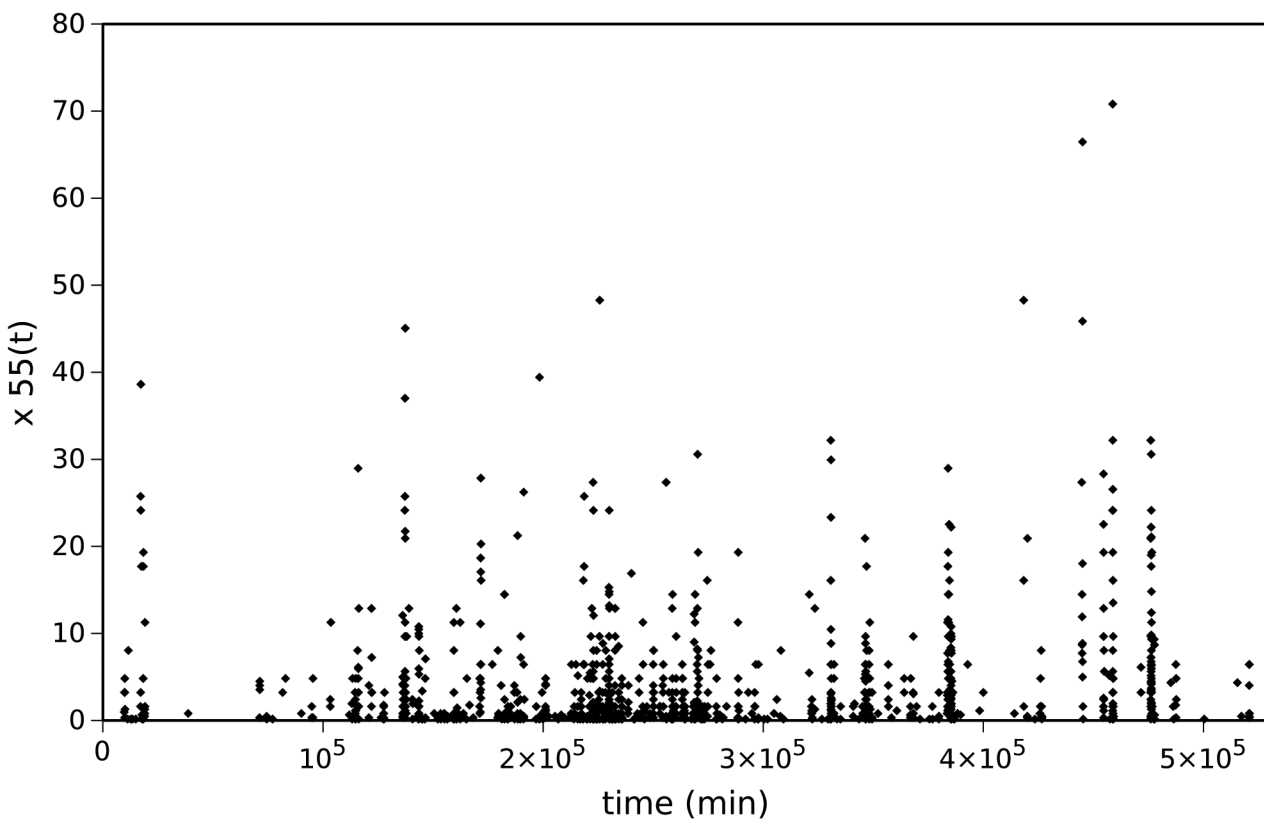


Fig 5. Time series, $x_{55}(t)$, representation of tornadoes path length, L (km), in year 2005.

doi:10.1371/journal.pone.0120260.g005

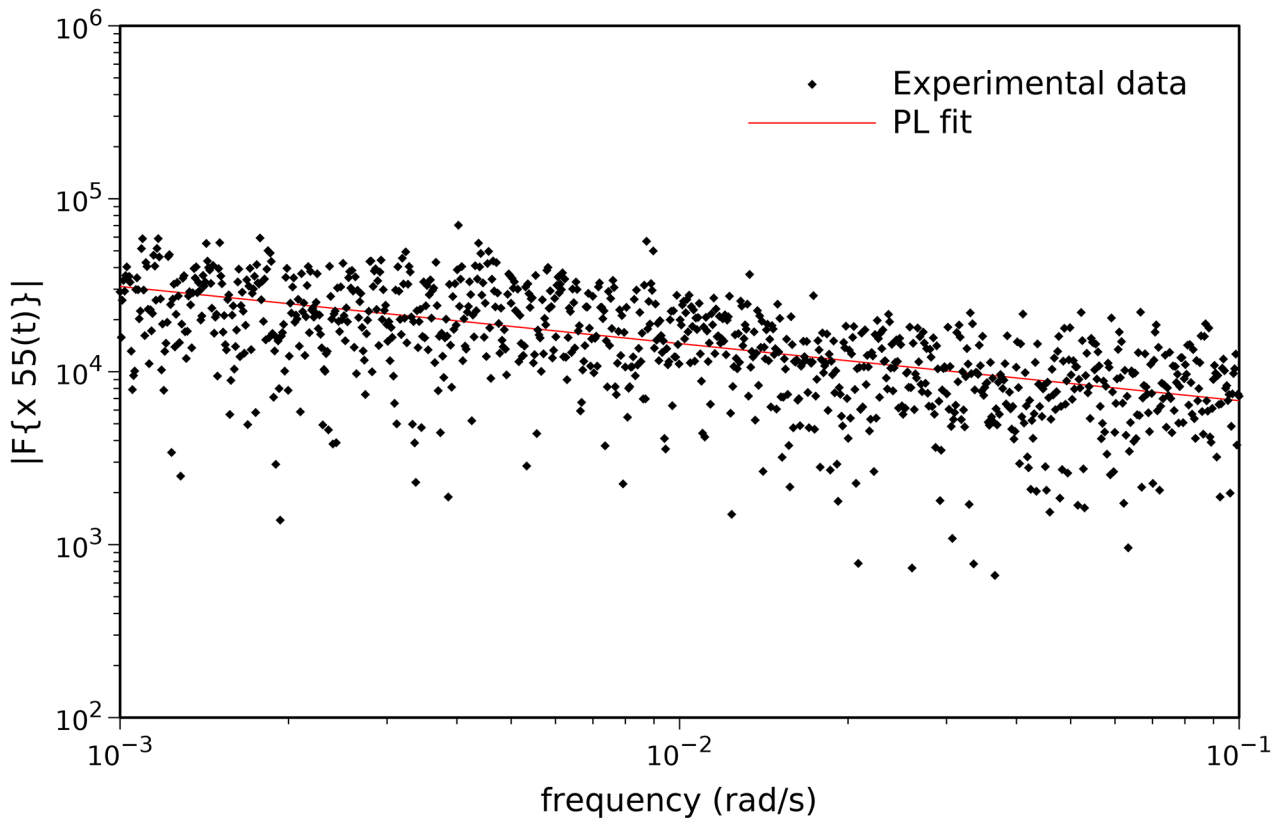


Fig 6. Magnitude of the Fourier transform, $|F\{x_{55}(t)\}|$, and PL approximation for the time series representation of tornadoes path length, L , in year 2005.

doi:10.1371/journal.pone.0120260.g006

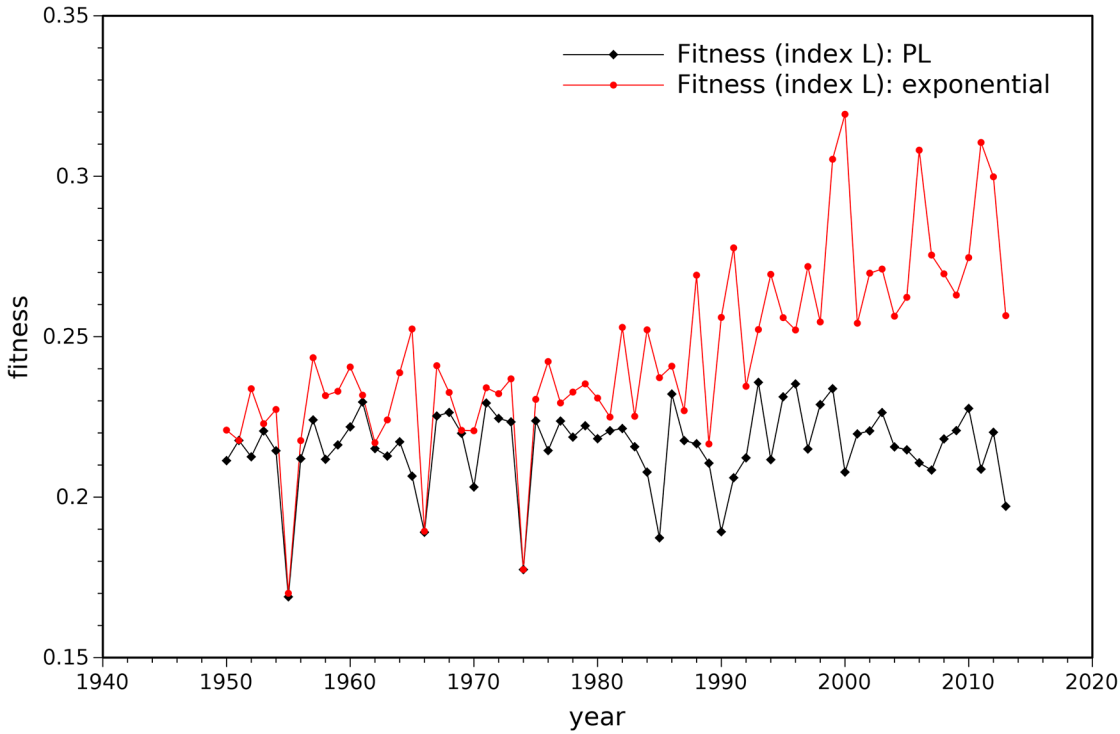
length, L , and tornado width, W , indices. It can be seen that, for all years, the PL model fits well into the experimental data.

Fig. 8 depicts the (a_i, b_i) loci when considering the tornado path length, L , and the tornado width, W . A larger (smaller) diameter of the bubbles means a larger (smaller) value of the fitness function of the PL fit. We can observe some kind of “line alignment” regularity in the parameters. The mention to such behavior emerged recently in the scientific literature in phenomena with memory and fractal features, but a definitive justification is still to be found [31–34].

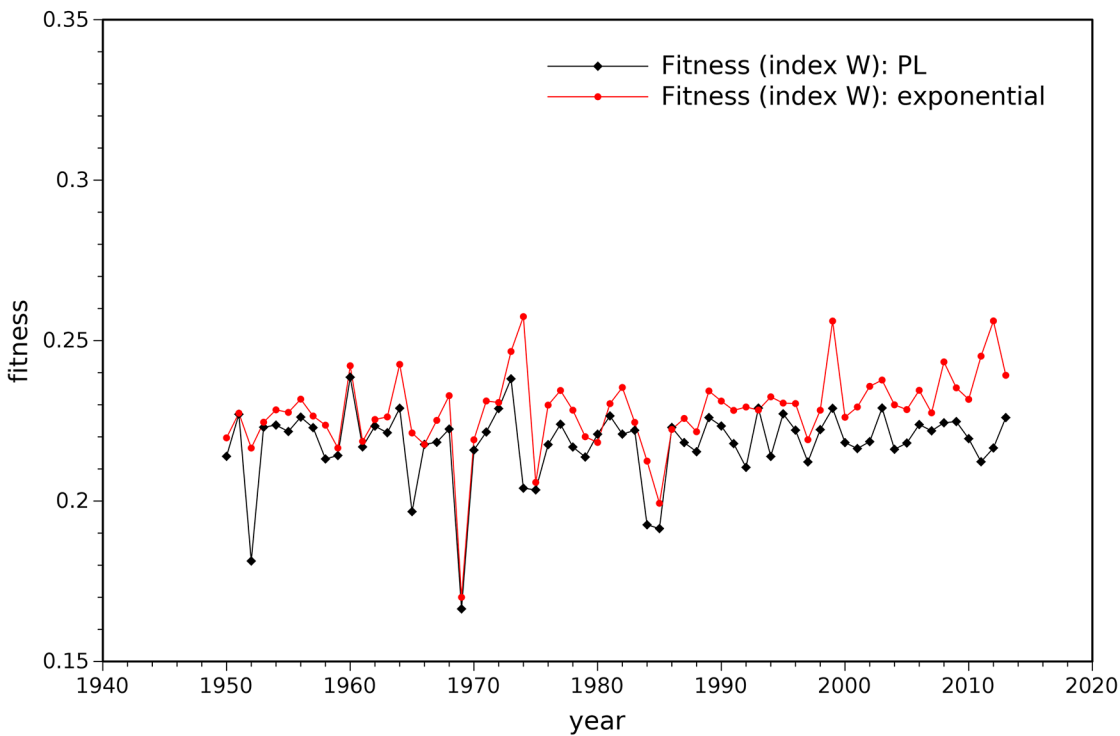
The results show that tornadoes collective behavior exhibit correlations and characteristic patterns. However, complementary analysis is needed to reveal deeper characteristics of the phenomena.

Analysis of Tornado Time Series in the Time Domain

In this section we analyze the tornado time series, $x_i(t)$, $i = 0, \dots, 63$, in the time domain. In subsection 4.1 we introduce the concept of circular time. In subsection 4.2 we compute the circular correlation [35] between the time series and we analyze the temporal behavior exhibited by tornadoes. In subsection 4.3, based on the maximum circular correlation values and the corresponding time delays, we use clustering and visualization tools to unveil patterns.



(a)



(b)

Fig 7. Fitness function values obtained for the PL and the exponential models: (a) tornado path length index, L ; (b) tornado width index, W .

doi:10.1371/journal.pone.0120260.g007

The flow of time

The adoption of a non-Cartesian time scale is motivated by the “one-year periodicity” observed in tornado time series, where December is close to January, and not the opposite, as a Cartesian scale implicitly assumes. On the other hand, a given month within one year is expected to be similar (i.e., close in some sense) to the same month in most of other years, or, at least, to the same months in the closest years.

Bearing these ideas in mind, we consider that time evolves along an Archimedean spiral, defined by:

$$\theta = 2\pi \cdot \left(\frac{t}{T} + i\right), r = p + q \cdot \theta \tag{8}$$

where (r, θ) denotes the radius and angle coordinates, respectively, and $i = 0, \dots, 63$, represents the year. For simplicity, we consider $p = q = 1$. We represent the tornadoes by points, where coordinates (r, θ) correspond to the circular (spiral) time of the event. We interpolate the points, associating a color to the intensity, measured by L (in log units), and we depict the events on a contour plot. With this representation we preserve the proximities of seasons both within each year and between distinct years. This means, for example, that January of year 2006 is close to December of year 2005, and that January 2005 is close to January 2004 and January 2006.

On the other hand, the major drawback is that initial years (in the center) have a small graphical representation, while recent years (in the outside region) have a larger area.

Fig. 9 illustrates the temporal evolution and intensity of the events occurred in the U.S. during the time period 1950–2013. On the left half circle (i.e., for the months from April up to September) we can see many small (in time) events, while, on the right half circle, the events are larger. Higher intensity events are observed in the fourth and first trimester, in that order.

Since the results shown in Fig. 9 close resemble those of chaotic systems, we first analyze the chart by means of the fractal dimension. The fractal dimension is a signature of complex systems, composed by many parts that interact with each other at different scales in time and space, which is the case of the atmospheric phenomena that trigger tornadoes.

We calculate the fractal dimension of sets of points (in a black and white picture) according with a variable threshold level. First, we divide the whole set of events into $l = 16$ groups, as a function of their intensity level, measured by L (in log units). Second, we generate the respective monochrome images corresponding to the temporal evolution of the events (i.e., black and

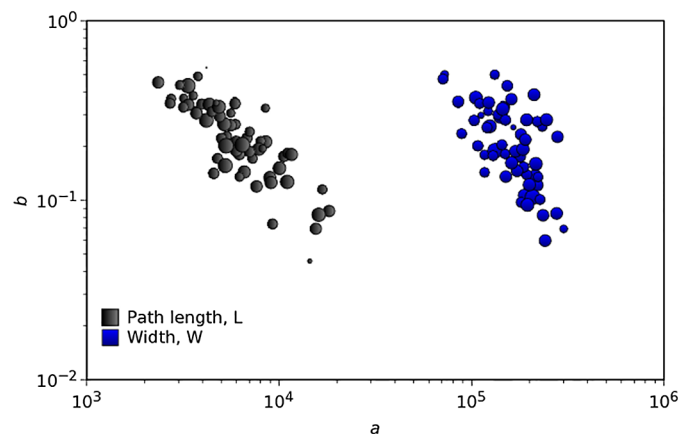


Fig 8. Loci of the PL parameters (a, b) for all years, $i = 0, \dots, 63$, considering the tornadoes path length, L , and width, W , to quantify intensity. A larger (smaller) diameter of the bubbles means a larger (smaller) value of the fitness function of the PL fit.

doi:10.1371/journal.pone.0120260.g008

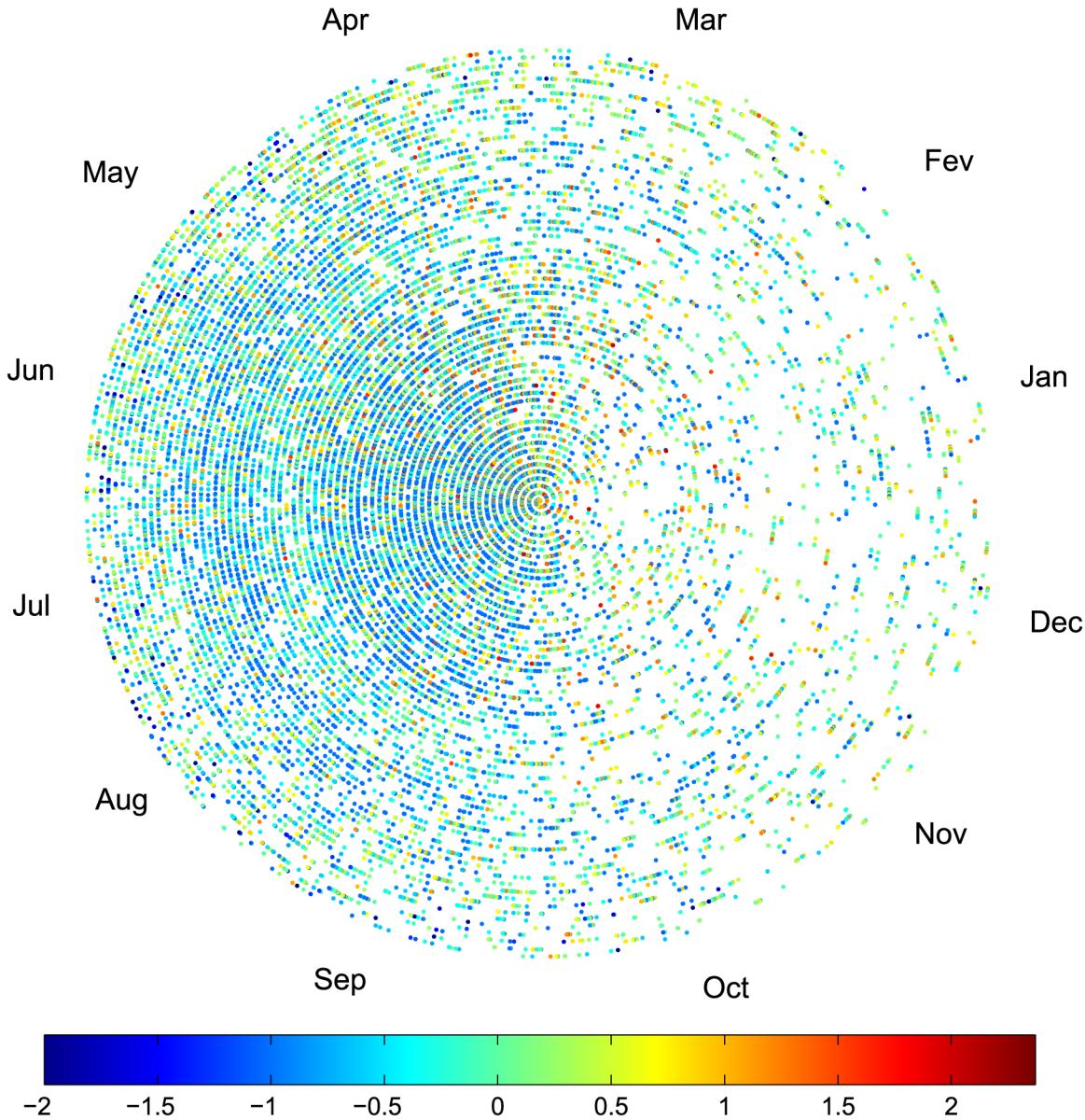


Fig 9. Temporal evolution and intensity (in log units) of the events occurred in the U.S. during the time period 1950–2013. The coordinates (r, θ) represent the circular (spiral) time of the events and the colors represent the intensity, measured by L (in log units).

doi:10.1371/journal.pone.0120260.g009

white images, similar to Fig. 9). Third, we adopt the box-counting method for estimating the fractal dimension of the plots [36–37]. The value for l establishes a compromise between good resolution and a descriptive number of events in each group.

Given a set S in a n -dimensional space and considering that $N_\epsilon(S)$ is the minimum number of n -dimensional cubes of side-length $\epsilon > 0$ needed to cover S , if there is a number d so that $N_\epsilon(S) \sim 1/\epsilon^d$ as $\epsilon \rightarrow 0$, then the box counting dimension of S is:

$$d = -\lim_{\epsilon \rightarrow 0} \frac{\ln[N_\epsilon(S)]}{\ln(\epsilon)}, \quad \epsilon > 0 \tag{9}$$

In our case S consists of monochrome images and small values of ϵ are reached by accessing the images at the pixel level.

In Fig. 10 we depict d for the 16 groups of data. It can be seen that for the first three groups the fractal dimension is small, probably due to missing data corresponding to unreported weak events. For mid-range intensity values d is almost constant. The fractal dimension decays for stronger events, since those are less common.

In conclusion, this graphical technique leads to an intuitive representation of data, but does not reveal clear patterns and relationships between sets of data. This means that more advanced processing techniques must be developed.

Time series correlation

Inspired on the previous sub-section and in the quasi-periodicity of time for the phenomena under study, here we propose analytical techniques based on the concept of “circular time”.

The circular correlation function, $c_{ij}(\tau)$, compares $x_i(t)$ with the time delayed signal $x_j(t-\tau)$, where time, t , is measured with respect to the beginning of the corresponding year. In analytical terms we have:

$$c_{ij}(\tau) = \frac{\sum_{k=1}^T x_i(t_k) \cdot x_j(t_k - \tau)}{\sqrt{\sum_{k=1}^T x_i(t_k)^2 \cdot \sum_{k=1}^T x_j(t_k - \tau)^2}}, \quad i, j = 0, \dots, 63 \tag{10}$$

For each case, the delayed signal $x_j(t-\tau)$ is treated as a circular buffer. The time delays, denoted as τ_{ij}^* , correspond to the time at which the circular correlation has its maximum value, $c_{ij}^* = c_{ij}(\tau_{ij}^*)$.

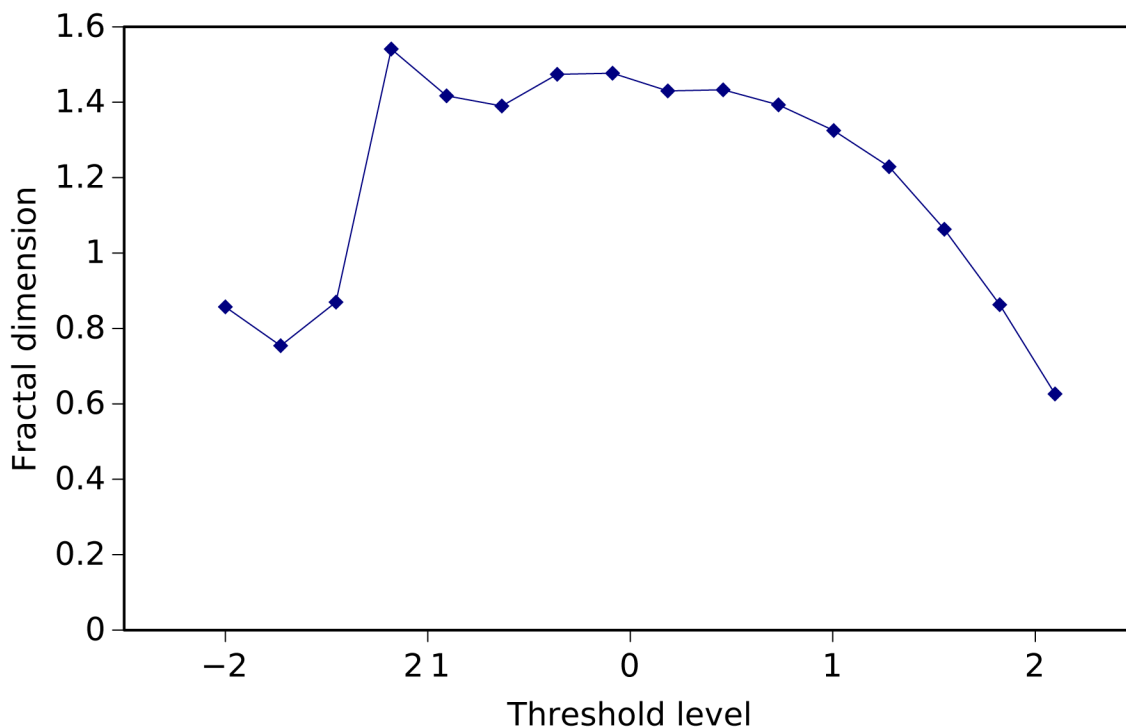


Fig 10. Evolution of the fractal dimension, d , for the 16 groups of data obtained according to the intensity level of the events, measured by L (in log units). The box counting method is used.

doi:10.1371/journal.pone.0120260.g010

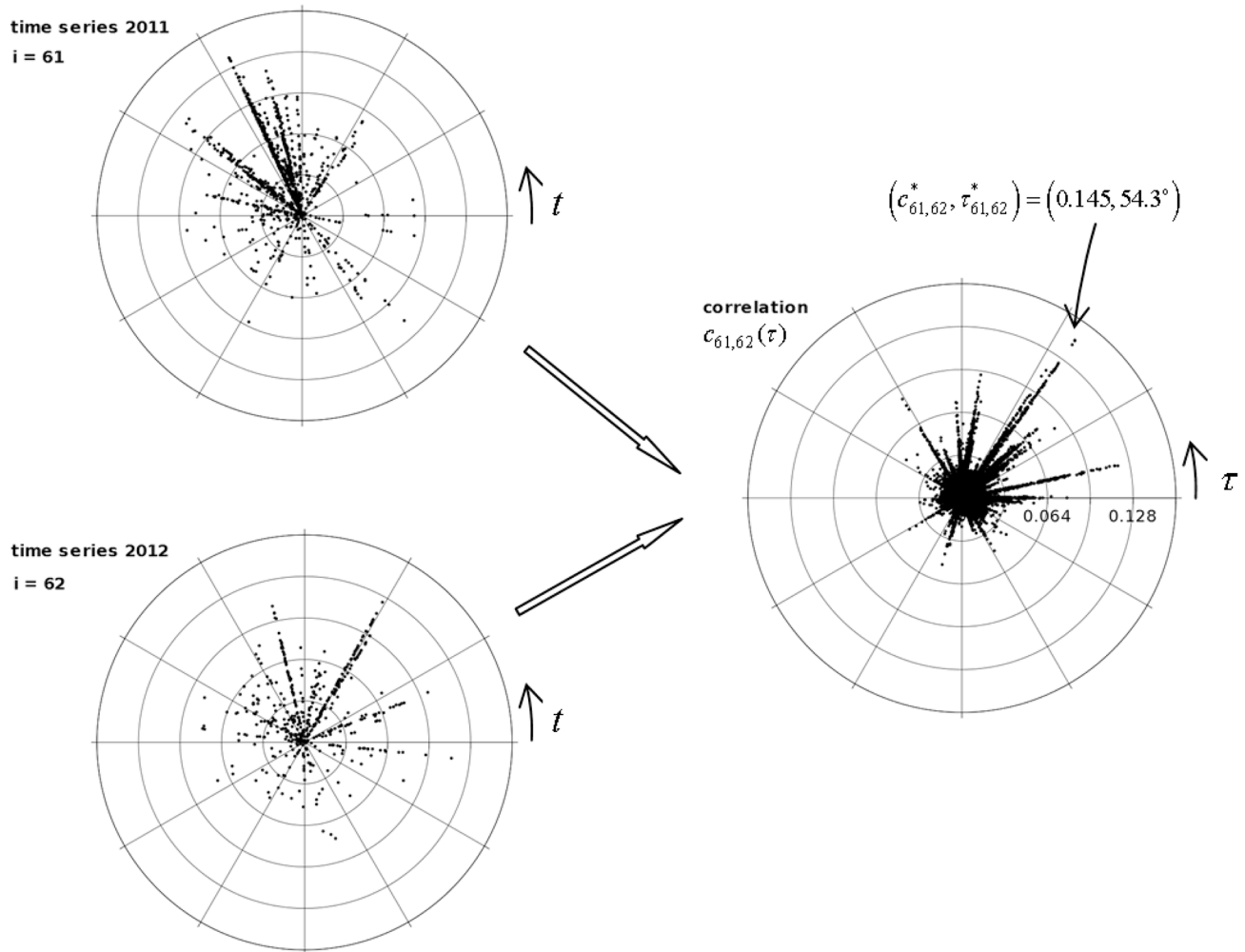


Fig 11. Tornado time series $x_{61}(t)$ and $x_{62}(t)$ (on the left) and correlation function, $c_{ij}(\tau)$, (on the right) between years 2011 and 2012.

doi:10.1371/journal.pone.0120260.g011

For example, Fig. 11 (on the left part) depicts the time series $x_{61}(t)$ and $x_{62}(t)$ (i.e., years 2011 and 2012), respectively. In both graphs the coordinated (r, θ) represent intensity (i.e., tornadoes path length, L , in logarithmic units) and circular time, respectively. For each time series, time t is referred to the beginning of the corresponding year, yielding $\theta = 2\pi \cdot t \cdot T^{-1}$. Fig. 11 (on the right part) shows the correlation function, $c_{61,62}(\tau)$. Points (r, θ) correspond to correlation value, $c_{61,62}(\tau)$, and (circular) time delay, τ , respectively. For this case, we obtain the maximum correlation $c_{61,62}^* = 0.145$ and time delay $\tau_{61,62}^* = 79500$ minutes (i.e., 55.2 days, in the time series, or 54.3° , in the polar plot).

We compute the parameters (c_{ij}^*, τ_{ij}^*) for all pairs of years $(i, j) = 0, \dots, 63$, using the procedure described above. The results are illustrated in Fig. 12A-B for, c_{ij}^* and τ_{ij}^* , respectively. The values for $i = j$ (i.e., those cases representing autocorrelation) are removed from Fig. 12A in order to avoid saturating the chart.

These maps compare all annual time series in a global perspective. However, the results are difficult to interpret and are most suitable for pairwise comparisons.

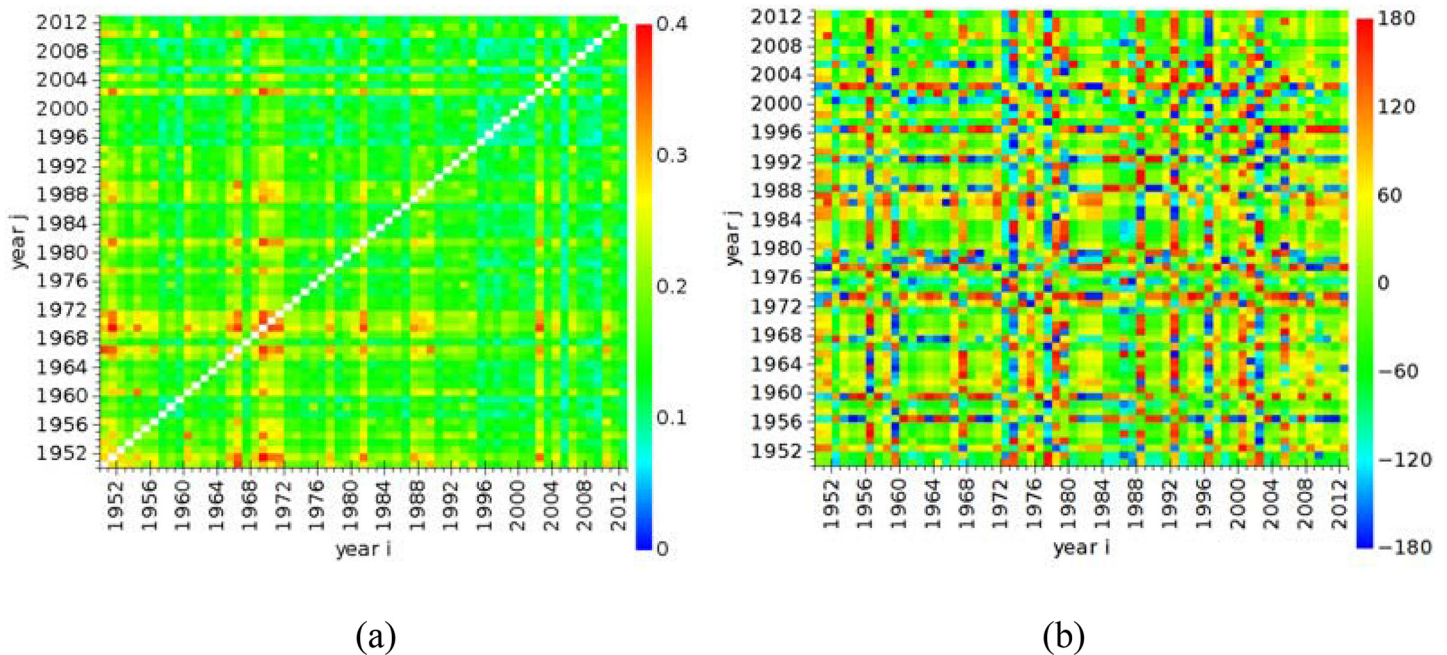


Fig 12. Matrix plots: (a) maximum circular correlation, c_{ij}^* ; (b) time delay, τ_{ij}^* , between all pairs of years $(i, j) = 0, \dots, 63$.

doi:10.1371/journal.pone.0120260.g012

Clustering and pattern visualization

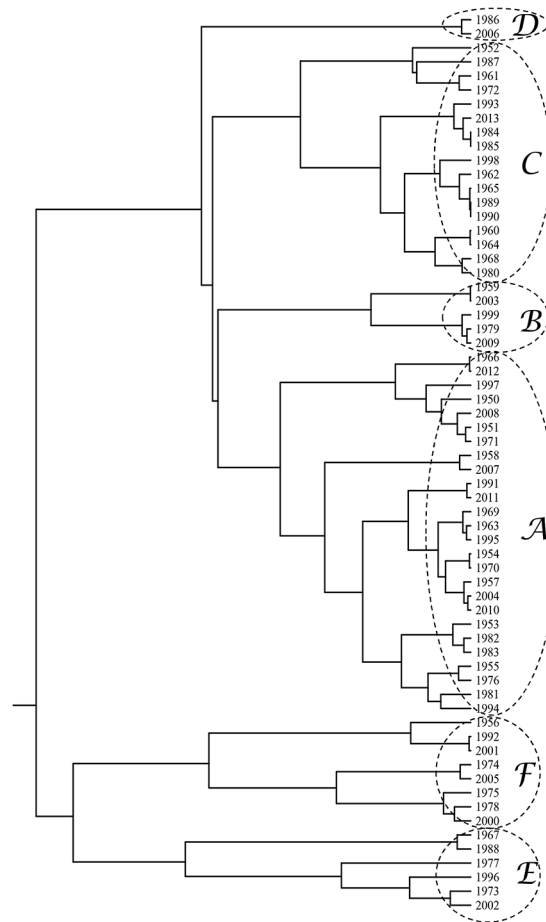
In this subsection, we adopt HC and MDS techniques to reveal and to visualize tornado patterns.

Hierarchical clustering. HC is a statistical technique for data analysis [38]. HC aims to build a hierarchy of clusters, in such a way that objects in the same cluster are similar to each other. In agglomerative clustering each object starts in its own singleton cluster and, at each step, the algorithm merges the two most similar (in some sense) clusters. In divisive clustering, all objects start in one single cluster and, at each step, the algorithm removes the “outsiders” from the least cohesive cluster. The results of HC are presented in a dendrogram. Clusters are combined (for agglomerative) or split (for divisive) based on a measure of their dissimilarity. The algorithm uses a metric, to quantify the distance between pairs of objects, and a linkage criterion, to quantify the dissimilarity between clusters. The metric will influence the composition of the clusters and must be chosen carefully. The Euclidean and Manhattan distances and the maximum, minimum and average linkages are often used.

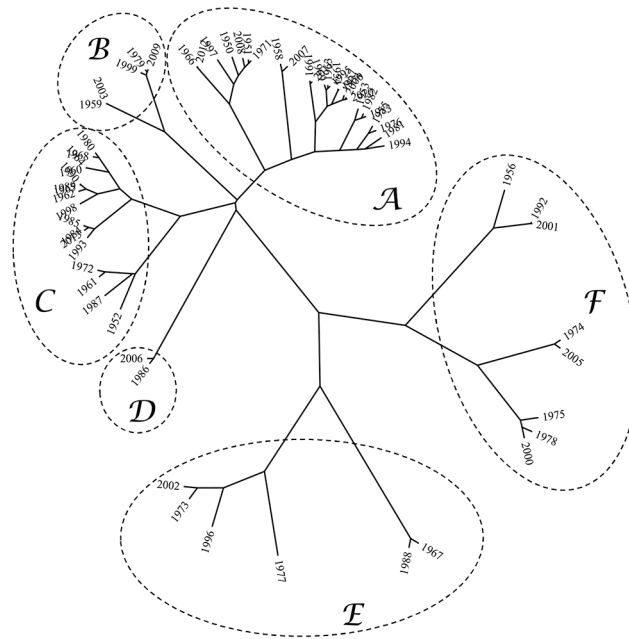
To assess the clustering we use the cophenetic correlation coefficient [39], which is a measure of how well the cluster tree generated by the linkage function preserves the pairwise distances between the original unmodeled data points. If the clustering is valid, the linking of objects in the cluster tree has a strong correlation with the distances between objects in the original data set. The closer the value of the cophenetic correlation coefficient is to 1, the more accurately the clustering reflects the original data. The result is often plotted in a Shepard graph comparing the original and the cophenetic distances. The better the clustering the closer to the 45 degree line the points will lie.

We feed the HC algorithm with the 64×64 matrix $C = [|\tau_{ij}^*|]$ and we adopt the successive (agglomerative) clustering and average-linkage methods.

Fig. 13A and 13B depict the corresponding dendrogram and visualization tree, respectively. As can be seen, two main patterns emerge forming a “dipole”, where each part is composed by



(a)



(b)

Fig 13. HC graphs: (a) dendrogram; (b) visualization tree. The matrix $C = [|\tau_{ij}^*|]$ is used.

doi:10.1371/journal.pone.0120260.g013

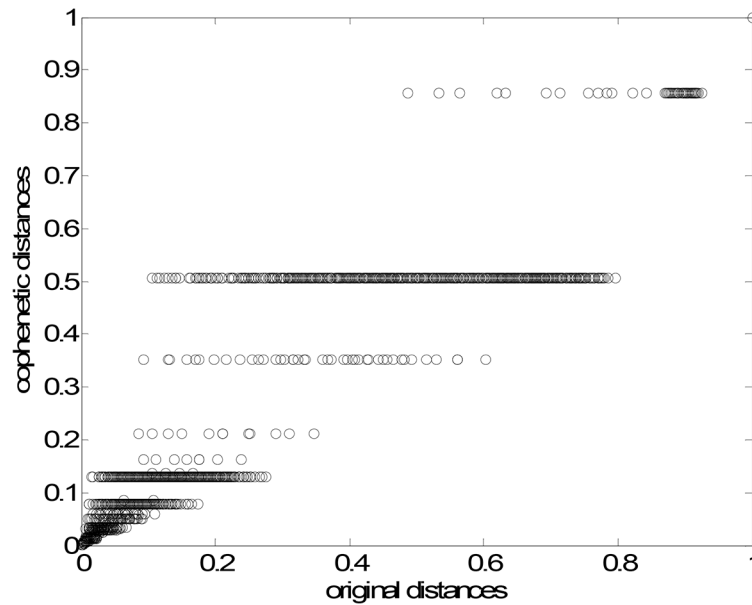


Fig 14. Shepard plot of the HC. The cophenetic correlation coefficient is equal to 0.93.

doi:10.1371/journal.pone.0120260.g014

the smaller clusters {A, B, C, D} and {E, F}, respectively. Fig. 14 represents the Shepard plot for the HC, reflecting an accurate clustering of the original data, where the cophenetic correlation coefficient is equal to 0.93.

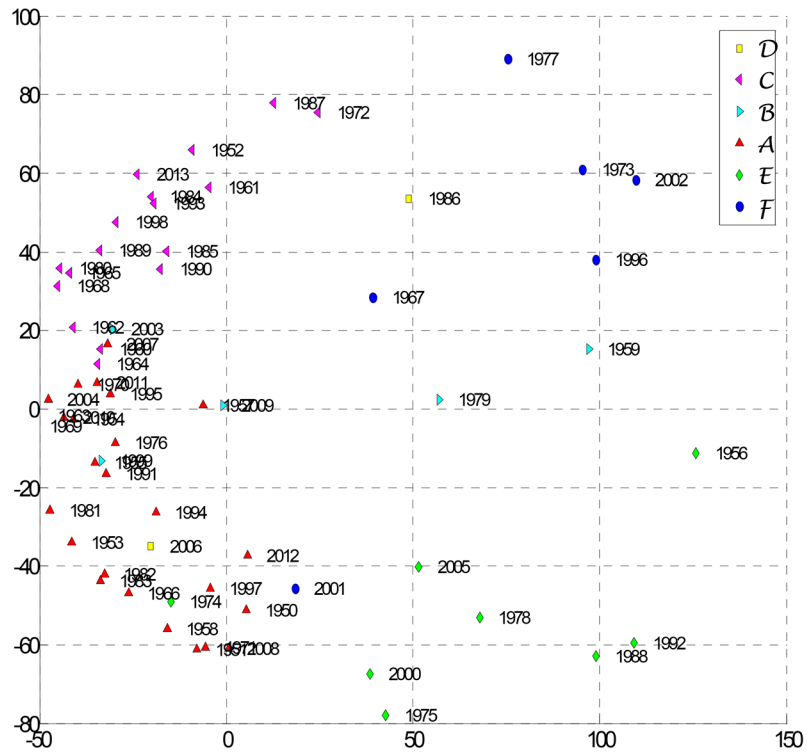
Multidimensional scaling. MDS is a statistical technique for visualizing data [40]. MDS generates maps, where objects perceived to be similar to each other are placed in clusters. The algorithm requires a $p \times p$ symmetric matrix, C , of item to item similarities (or, alternatively, of similarities), where p is the number of objects. MDS assigns a point to each item in a m -dimensional space and arranges the objects in order to reproduce the observed similarities (or, alternatively, dissimilarities). This is achieved by evaluating different configurations, while maximizing the goodness-of-fit. The raw stress, S , is often used to evaluate how well a particular configuration reproduces matrix C :

$$S = [d_{ij} - f(\delta_{ij})]^2 \tag{11}$$

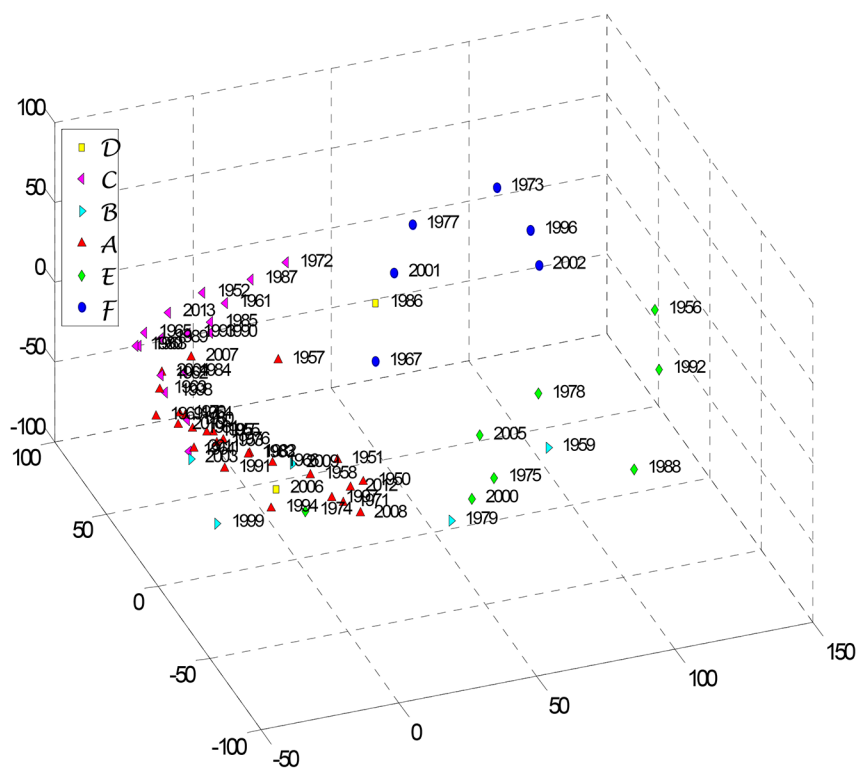
where, d_{ij} and δ_{ij} correspond to the reproduced and to the observed distances, respectively. The function $f(\cdot)$ represents a transformation applied to the input data. The smaller the stress, S , the better is the fit between the observed and the reproduced data.

For $m = 2$ or $m = 3$ the generated map can be easily interpreted. The actual orientation of axes in the final solution is arbitrary. MDS maps are insensitive to rotations and translations. In fact, MDS maps interpretation is based on the emerging clusters and relative distances, rather than on the absolute coordinates or shapes. The measure for constructing matrix C depends of the researcher’s choice.

For accessing the quality of the MDS map there are adopted the stress and Shepard plots. The stress plot, representing S versus the number of dimensions m of the MDS, leads to a monotonic decreasing chart. Therefore, we choose the “best” dimension m as a compromise between stress reduction and number of dimensions for the MDS. The Shepard diagram depicts the reproduced distances, for a particular value m of the MDS dimension, versus the observed input data (distances). Therefore, in the Shepard diagram, a narrow scatter around a 45



(a)



(b)

Fig 15. MDS maps: (a) 2-dimensional; (b) 3-dimensional. The matrix $C = [|\tau_{ij}^*|]$ is used.

doi:10.1371/journal.pone.0120260.g015

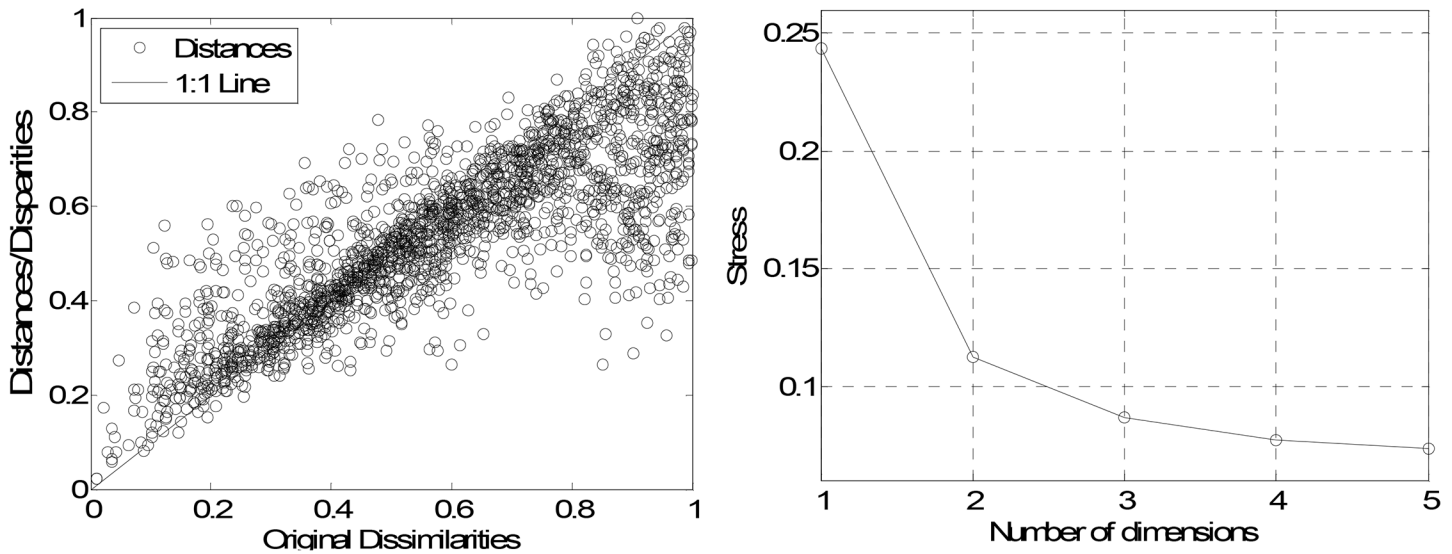


Fig 16. MDS assessment graphs: (a) Shepard map for the 3-dimensional MSD map; (b) stress map.

doi:10.1371/journal.pone.0120260.g016

degree line indicates a good fit of the distances to the dissimilarities, while a large scatter indicates a lack of fit, for that value m .

As before we feed the MDS with the 64×64 matrix $C = [|\tau_{ij}^*|]$ and generate the 2- and 3-dimensional maps (Fig. 15A and 15B, respectively), where we can identify the “dipole” mentioned previously. To assess the quality of the MDS maps we plot the Shepard and the stress diagrams. The Shepard graph (Fig. 16A) shows the points distributed around the 45 degree line, which means a good fit of the distances to the dissimilarities. The stress plot (Fig. 16B) reveals that a 3-dimensional space describes well the data.

The results obtained for the distinct visualization techniques (Figs. 13 and 15) reveal clusters and, consequently, patterns in time. Nevertheless, the results are far from a simple type of periodicity. This means that either the time series length is insufficient to grab some period, or that some complex long run behavior is present and needs to be further analyzed with complementary techniques. Embedding other variables, namely trajectories, pressures, temperatures and humidity may help getting a clear picture of the global pattern.

Conclusion

This paper analyzed the collective behavior of tornadoes from the perspective of dynamical systems, due to the fact that such phenomena reveal characteristics found in complex systems. A public domain database of U.S. tornadoes occurred in the time period 1950–2013 was adopted. The events were modeled as time series of Dirac impulses proportional to their intensity. In a first approach, we proposed the Fourier transform was applied to characterize tornadoes behavior. A second approach adopted the concept of circular time and correlation. Hierarchical clustering and multidimensional techniques were used to identify and visualize patterns. It is clear the emergence of clusters corresponding to complex dynamical effects. In fact, the emerging patterns have resemblances with those of chaotic systems that lead to a poor predictability. Furthermore, measuring a richer set of variables and recording of longer time series might be necessary to establish a solid basis of computer data analysis. These approaches confirm that tornado dynamics is complex and exhibits long memory characteristics. The results may

inspire still other approaches that, perhaps in combination with those presented, may yield important insights to better understand the phenomenon.

Author Contributions

Conceived and designed the experiments: AML JTM. Performed the experiments: AML JTM. Analyzed the data: AML JTM. Contributed reagents/materials/analysis tools: AML JTM. Wrote the paper: AML JTM.

References

1. Simmons KM, Sutter D, Pielke R. Normalized tornado damage in the United States: 1950–2011. *Environmental Hazards* 2013; 12: 132–147.
2. Edwards R, LaDue JG, Ferree JT, Scharfenberg K, Maier C, Coulbourne WL. Tornado Intensity Estimation: Past, Present, and Future. *Bulletin of the American Meteorological Society* 2013; 94: 641–653.
3. Feuerstein B, Groenemeijer P, Dirksen E, Hubrig M, Holzer AM, Dotzek N. Towards an improved wind speed scale and damage description adapted for Central Europe. *Atmospheric Research* 2011; 100: 547–564. doi: [10.1007/s00392-011-0289-0](https://doi.org/10.1007/s00392-011-0289-0) PMID: [21327842](https://pubmed.ncbi.nlm.nih.gov/21327842/)
4. Fujita TT. Proposed characterization of tornadoes and hurricanes by area and intensity. SMRP Research Paper 91 1971.
5. Fujita TT. Tornadoes and downbursts in the context of generalized planetary scales. *J. Atmos. Sci.* 1981; 38: 1511–1534.
6. Fujita TT, Pearson AD. Results of FPP classification of 1971 and 1972 tornadoes. Eighth Conf. on Severe Local Storms, Denver, CO, Amer. Meteor. Soc.; 1973: 142–145.
7. Potter S. Fine-tuning Fujita. *Weatherwise* 2007; 60: 64–71.
8. Dotzek N, Kurgansky MV, Grieser J, Feuerstein B, Névir P. Observational evidence for exponential tornado intensity distributions over specific kinetic energy. *Geophys. Res. Lett.* 2005; 32: L24813.
9. Dotzek N. Derivation of physically motivated wind speed scales. *Atmos. Res.* 2009; 93: 564–574.
10. Kurgansky MV. The statistical distribution of intense moist-convective, spiral vortices in the atmosphere. *Dokl. Earth Sci.* 2000; 371: 408–410.
11. Schielicke L, Névir P. Introduction of an atmospheric moment combining Eulerian and Lagrangian aspects of vortices: Application to tornadoes. *Atmos. Res.* 2011; 100: 357–365.
12. Schielicke L, Névir P. Comprehensive analysis of tornado statistics in comparison to earthquakes: intensity and temporal behavior. *Nonlin. Processes Geophys.* 2013; 20: 47–57.
13. Malamud BD, Turcotte DL. Statistics of severe tornadoes and severe tornado outbreaks. *Atmos. Chem. Phys.* 2012; 12: 8459–8473.
14. Pinto CMA, Lopes AM, Tenreiro Machado JA. A review of power laws in real life phenomena. *Communications in Nonlinear Science and Numerical Simulations* 2012; 17: 3558–3578.
15. Tenreiro Machado J. Accessing complexity from genome information. *Communications in Nonlinear Science and Numerical Simulation* 2012; 17: 2237–2243.
16. Turcotte DL, Malamud BD. Earthquakes as a complex system. *International Geophysics Series* 2002; 81(A): 209–228.
17. Sornette D, Pisarenko V. Fractal plate tectonics. *Geophysical Research Letters* 2003; 30.
18. Bak P, Tang C, Wiesenfeld K. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters* 1987; 59: 381–384. PMID: [10035754](https://pubmed.ncbi.nlm.nih.gov/10035754/)
19. Carlson JM, Doyle J. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E* 1999; 60: 1412. PMID: [11969901](https://pubmed.ncbi.nlm.nih.gov/11969901/)
20. Newman ME. Power laws, Pareto distributions and Zipf's law. *Contemporary physics* 2005; 46: 323–351.
21. Brooks HE, Doswell CA. Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.* 2001; 56: 191–201. PMID: [11719250](https://pubmed.ncbi.nlm.nih.gov/11719250/)
22. Dotzek N, Grieser J, Brooks HE. Statistical modeling of tornado intensity distributions. *Atmos. Res.* 2003; 67–68; 163–187.
23. Brooks HE. On the Relationship of Tornado Path Length and Width to Intensity. *Weather Forecast.* 2004; 19: 310–319.

24. Doswell CA, Edwards R, Thompson RL, Hart JA, Crosbie KC. A simple and flexible method for ranking severe weather events. *Weather Forecast*. 2006; 21: 939–951.
25. Verbout SM, Brooks HE, Leslie LM, Schultz DM. Evolution of the U.S. tornado database: 1954–2003. *Weather Forecast*. 2006; 21: 86–93.
26. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM review* 2009; 51: 661–703.
27. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* 1989; 307–333.
28. Tenreiro Machado J. Fractional order modelling of fractional-order holds. *Nonlinear Dynamics* 2012; 70: 789–796.
29. Mandelbrot BB, Van Ness JW. Fractional Brownian motions, fractional noises and applications. *SIAM Rev*. 1968; 10: 422–437.
30. Beran J. *Statistics for Long-Memory Processes*. New York: CRC Press; 1994.
31. West BJ, West D. Fractional dynamics of allometry. *Fractional Calculus and Applied Analysis* 2012; 15: 70–96.
32. Tenreiro Machado JA, Costa AC, Quelhas MD. Can power laws help us understand gene and proteome information?. *Advances in Mathematical Physics* 2013.
33. West BJ. Colloquium: Fractional calculus view of complexity: A tutorial. *Reviews of Modern Physics* 2014; 86: 1169–1184.
34. Tenreiro Machado JA, Lopes AM. The Persistence of Memory. *Nonlinear Dynamics* 2014; doi: [10.1007/s11071-014-1645-1](https://doi.org/10.1007/s11071-014-1645-1)
35. Fisher NI. *Statistical analysis of circular data*. New York: Cambridge University Press; 1995.
36. Berry MV. Diffractals. *Journal of Physics* 1979; A12: 781–97.
37. Schroeder M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. New York: W.H. Freeman; 1991.
38. Jain A, Dubes R. *Algorithms for Clustering Data*. Englewood Cliffs: Prentice-Hall; 1988.
39. Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods. *Taxon* 1962; 11: 33–40.
40. Cox TF, Cox MA. *Multidimensional scaling*. CRC Press; 2000.