Are tornadoes getting worse?

A statistical analysis

Table of Contents

INTRODUCTION 1
FIGURE 1 2
FIGURE 2 2
FIGURE 3 3
THE U.S. TORNADO ANALYSIS 4
FIGURE 4 4
TABLE 1 5
FIGURE 5 6
FIGURE 6 7
FIGURE 7 8
TABLE 2 8
FIGURE 8 9
FIGURE 9 10
FIGURE 10 11
FIGURE 11 12
TABLE 3 13
CONCLUSION 14
REFERENCES 14

Introduction

The Father's Day weekend, 2023 and the next ten days, were particularly difficult across greater Tulsa area, Oklahoma. The 100mph winds blew down trees and power poles and left 200,000 houses in power outage during the three-digit temperature season of the year [1]. Repair and cleanup costs are estimated to be more than US $2 million so far [2].

While searching for available cooling stations and internet hotspots near their damaged house, a couple of family friends came across a climate change related podcast claiming that for tornadoes "researchers haven't seen an increase in their frequency, and measurements…in the last decade, tornado activity has been moving towards the southeast, an area of the U.S. where we find a higher percentage of mobile homes, which don't stand a chance against a powerful twister." [3]. My friends wondered if there was any sound statistical analysis backing up these claims, and if extracting true facts from the national tornado data base would be a fun Father's Day game for me.

Fact checking if tornadoes in the U.S. are getting worse was a fun challenge, indeed. Additionally, it was useful because the public interest was at a peak not only in Oklahoma but also here in Texas [4]. Another goal of my statistical study was to model a simple equation that could predict tornado occurrences. In other words, how many tornadoes and of what magnitude will most likely occur across a thousand-mile sized geographic area during a specific season of the year, with a 95% confidence?

The challenge of understanding the tornado trends has been already taken on by many, from the weather officers E. J. Fawbush and R. C. Miller, who were the first leaders of the national Severe Weather Warning Center in 1951 [5], to the big data numerical simulations for tornado forecasts ran by the atmospheric physicists of today, such as Dr. Louis J. Wicker, an associate professor at Texas A&M University and a chief scientist at the National Severe Storms Laboratory in Norman, Oklahoma [6]. The tornado database [7] has recently become of interest for the Kaggle community [8] who generated a plethora of data exploratory visualizations.

A crucial study published in the *Nature Journal* links tornado production to satellite imagistic through an environmental covariate named Significant Tornado Parameter (STP); a significant number of the tornado events are linked to the tornadic supercells, a type of thunderstorm detectable through image processing algorithms. About one third of the Doppler radar detected supercells developed into touching-down tornadoes [9].

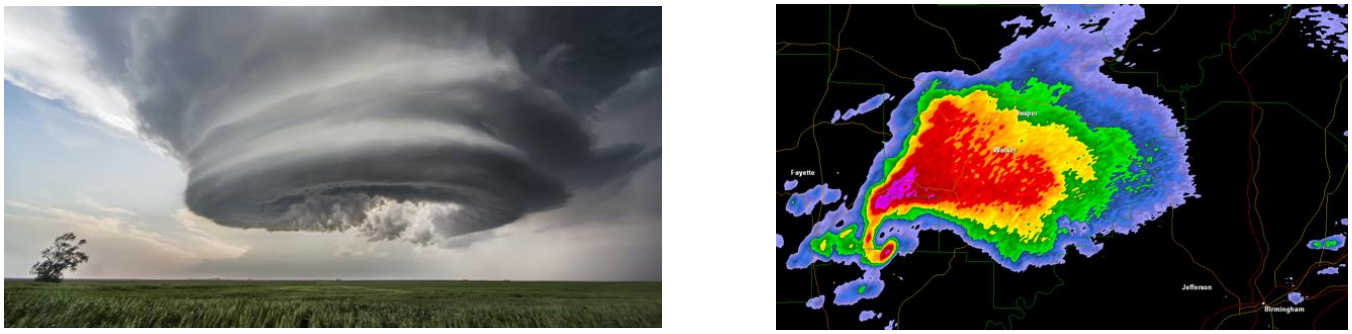

Figure 1 Tornadic supercell. Left panel: a supercell seen from the ground. Right panel: the radar image of a tornadic supercell with the hook like signature of the touchdown. *Sources: https://earth.stanford.edu/news/scientists-solve-mystery-supercell-storms-icy-plumes, https://www.ustornadoes.com/2013/02/14/understanding-basic-tornadic-radar-signatures/*

FIGURE 1 shows a supercell seen from the ground and the radar image of a tornadic supercell. What makes the radar image of a tornadic supercell different from a regular storm is its tailed shape and the hook like signature of the touchdown. Image processing can thus detect appreciably more tornadoes than the observers on the ground.

FIGURE 2 shows the anatomy of a typical supercell with a touchdown marked as “tornado,” and the real-life image of a supercell with two simultaneous touchdowns. One supercell can generate multiple tornado touchdowns. On the ground, this could create the impression of a unique hopping tornado. In fact, the touchdown funnel never changes direction, it always moves solidary with the cloud [10].

Figure 2 Left panel: cartoon detailing what makes a tornadic supercell different from a regular storm. Right panel: twin violent tornadoes in Nebraska, June 16, 2014. *Sources: https://www.wave3.com/story/38530243/behind-the-forecast-what-makes-a-supercell-super/, https://photolib.noaa.gov/Collections/National-Weather-Service/Other/emodule/627/eitem/17697*

Top of the Document

A particular interest in numerically modeling tornado events is shown by the insurance companies. One parameter employed in insurance premium calculations is the aggregated annual tornado loss function. Statistical models of the tornado risk and variability try to predict the actual losses, which in Oklahoma were US $1.27 billion during a 25-year period [11].

FIGURE 3 shows the Briarwood Elementary School and its neighborhood in Moore, Oklahoma, before and after the giant tornado on May 20, 2013. This Moore tornado was categorized as an EF5, a rare twister packing estimated wind speeds greater than 200 mph. It killed 25 people and created a swath of destruction, including 300 demolished homes. It raked a 14-mile path of destruction through the southern areas of Oklahoma City, causing $2 billion worth of damage [12].

Figure 3 Briarwood Elementary School (bottom of picture) was in the tornado's path. Compare before and after images. *Source:* https://www.bbc.com/news/world-us-canada-22608754

Tornadoes are classified based on the effects produced with the Fujita (F) scale, which was adopted in the U.S. in 1971. All tornadoes recorded in the national database before 1971 were retroactively assessed on the F-scale. The Fujita scale was updated in 2007 by introducing 28 distinct descriptors for tornado surveillance. The more precise tornado-effects scale is labeled as Enhanced Fujita (EF) [13].

An EFU tornado is one currently unclassified (U). Sometimes it takes months or years to survey the sites touched by tornadoes; clerical errors may appear, such as entering a length of 80 mi instead of 80 yards (0.8mi) [14]. Tornadoes were likely underreported before 1953, and before 1980 (before the geostationary operational environment satellites era) multiple tornadoes from the same supercell were recorded as one unique event. The tornado magnitudes assessed before 2007 were likely overstated, partially because there were no clear survey descriptors and partially because the constructions could get more easily damaged in the past.

The analysis in the following sections of this report investigates possible significant trends in the tornado patterns in the U.S. from 1950's to this day. The questions the present study will answer are: (1) is there an increase in the annual number or strength of the tornadoes? (2) is there a shift of the geographic location of the tornadoes? (3) can a numerical model estimate the tornado occurrence in the greater Tulsa area?

Top of the Document

The U.S. Tornado Analysis

For this observational study, the entire database available at https://www.spc.noaa.gov/wcm/data/1950-2022_actual_tornadoes.csv [15] was employed and no entry was deleted. The data file has 29 variables (columns) all categorical (character class). Twenty variables (columns) were employed in the present analysis: 2, 3, 5, 8, 10, 11, and 16 to 29. This data set contains errors and oddities, some already corrected, some to be detected through cross validation with other sources [16].

The oldest tornado entries are from 01-03-1950, the most recent entries are from September 2022; during the almost 72 years of tornado data collection and classifying, this database recorded 68,701 tornadoes. Many of these tornadoes were observed as multiple segments, many traveled across multiple counties or multiple states.

The data exploration and the modeling were performed in R [17]. Besides *base* R, the libraries *MASS* and *lmtest* were employed. All outliers were kept in the models and interpreted. Two supplementary data sets were accessed for cross validating the information pertaining to the greater Tulsa area [18]. The statistical models were validated with the Oklahoma tornado database [19].

Question 1: is there an increase in the annual number or strength of the tornadoes?

The scatter plot of all the entries aggregated by calendar year in **Figure 4** displays three trend regions separated by distinct fracture lines. One change in trend occurred around 1963-1964 and the second around 1990-1991. Linear models fit the data separated for each temporal region (era) well.

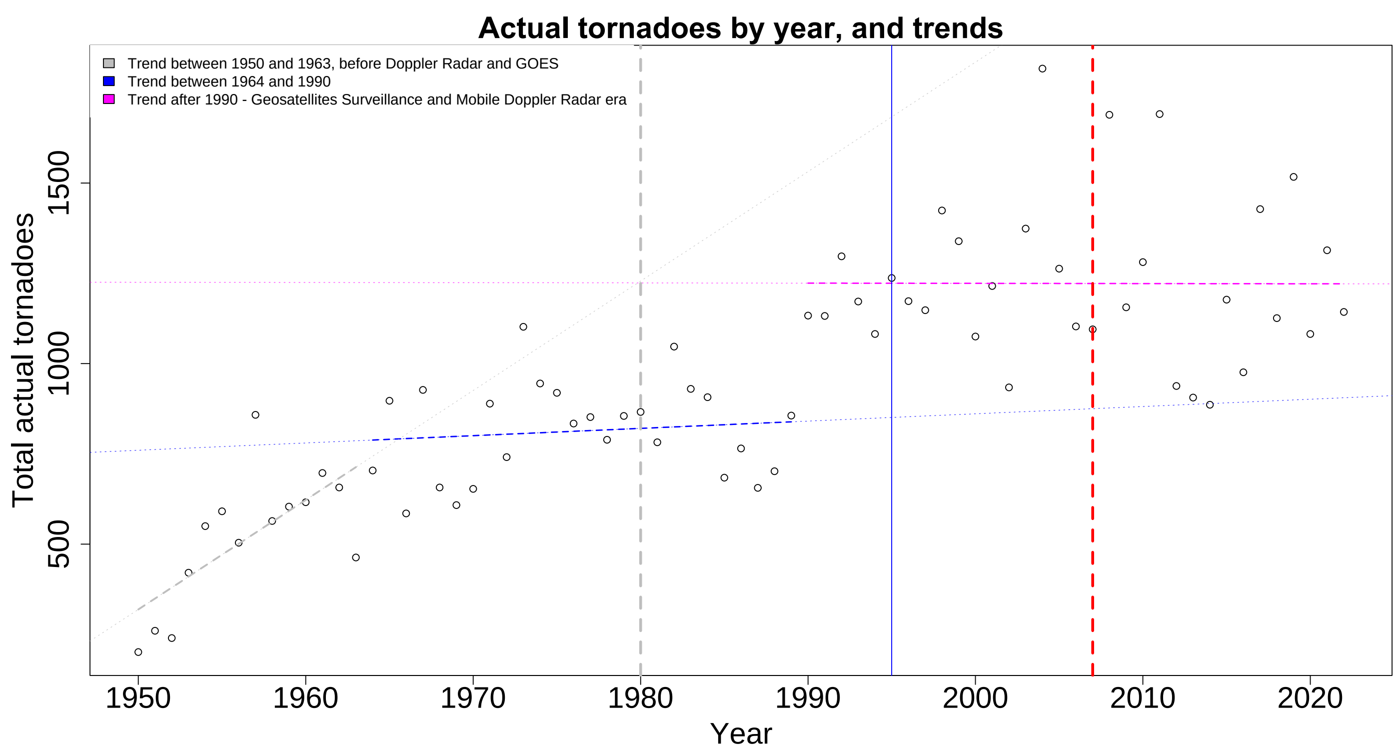

Figure 4 All tornadoes in the U.S. national database aggregated by calendar year. The vertical lines correspond to borderline years: 1980, the start of the GOES era; 1995, when the mobile Doppler radars were implemented by all states; 2007, the start of the EF scale. The linear trends fit the data from three technologically and logistically distinct eras: 1950 to 1963, 1964 to 1990, and 1990 to 2022.

Top of the Document

From 1950 to 1963, during the first 14 years of data collection, the number of recorded tornadoes increased steadily at a rate of 30 counts each year. This 14-year linear trend is statistically significant (p-value=0.007) but is not correlated with any known climatic change. It rather reflects the increase in personnel and operations involved in tornado detection, during a period of organizational and logistic development. The jump in tornado counts in 1963-1964, from a median of 557 tornadoes per year to 843 tornadoes yearly, most likely reflects an important technological change in the activity at the national storm center, perhaps within the newly formed Environmental Science Services Administration and the creation of the National Severe Storms Forecast Center.

During the 26-year period from 1964 to 1989, there was no significant change in the recorded tornado annual counts. The slope of the linear data-fit, of 2 counts per year, was not statistically significant (p-value=0.6). During this era, the Fujita scale was implemented and, starting with 1980, the Geostationary Operational Environmental Satellites (GOES) became operative.

The jump to a median of 1172 yearly recorded tornadoes after 1990 could be most likely causally connected with the generalization of the mobile Doppler radar units and with the computer assisted satellite image processing. During the 33-year period from 1990 to 2022, the annual tornado counts have a larger variability than in the precedent era but no linear trend, a zero rate of change. Marked on the plot in **FIGURE 4,** 1995 is the borderline year when all states had fully operational Doppler units, whereas 2007 is the borderline year when tornadoes started to be assessed with 28 clear descriptors on the EF scale.

In the U.S. national database, almost half of the recorded tornadoes are EF0 gales at less than 85mph, **TABLE 1**. A total of 756 tornadoes between 2016 and 2022 are EFU with no strength (magnitude) assigned. Less than 5% of the tornadoes are EF3 or greater, causing severe or devastating damage.

Table 1 Annual tornado records

F or EF scale	U (unassigned)	0	1	2	3	4	5	Total
Counts	756	31,776	23,333	9,636	2,556	585	59	68,701
Percentages	1.1%	46.3%	34.0%	14.0%	3.7%	0.9%	0.1%	100.0%

By plotting the observed tornadoes separately for each magnitude and each detection technology era, **FIGURE 5**, one notices that the annual mean number of strong tornadoes has decreased since 1950-s, and it is only the counts for weak tornadoes of magnitude 0 and 1 that have scaled up. On one hand, the decrease in the annual number of tornadoes assessed as strong is most likely due to the change in tornado assessment, from the less objective F-scale to the more objective EF-scale. On the other hand, the construction technology has improved significantly in the areas with frequent tornadoes therefore a strong tornado causes less damages today than in 1950; this comparison does not apply to geographic areas where many are living in low safety homes, areas where tornadoes were unlikely to occur in the past.

The goodness of fit of the three era-conditional trends of the annual tornado occurrences can be estimated by plotting the residuals of the linear models, ***FIGURE 6***. A data trend is truly linear if the red lines in the left-hand side panels are practically horizontal, and if all the points align on the dotted oblique line of the “Normal Q-Q plots” in the right-hand side panels. As the red lines are not exactly horizontal, and the Q-Q plots have outliers, the trends for the annually collected tornado counts can only be approximated as linear.

Top of the Document

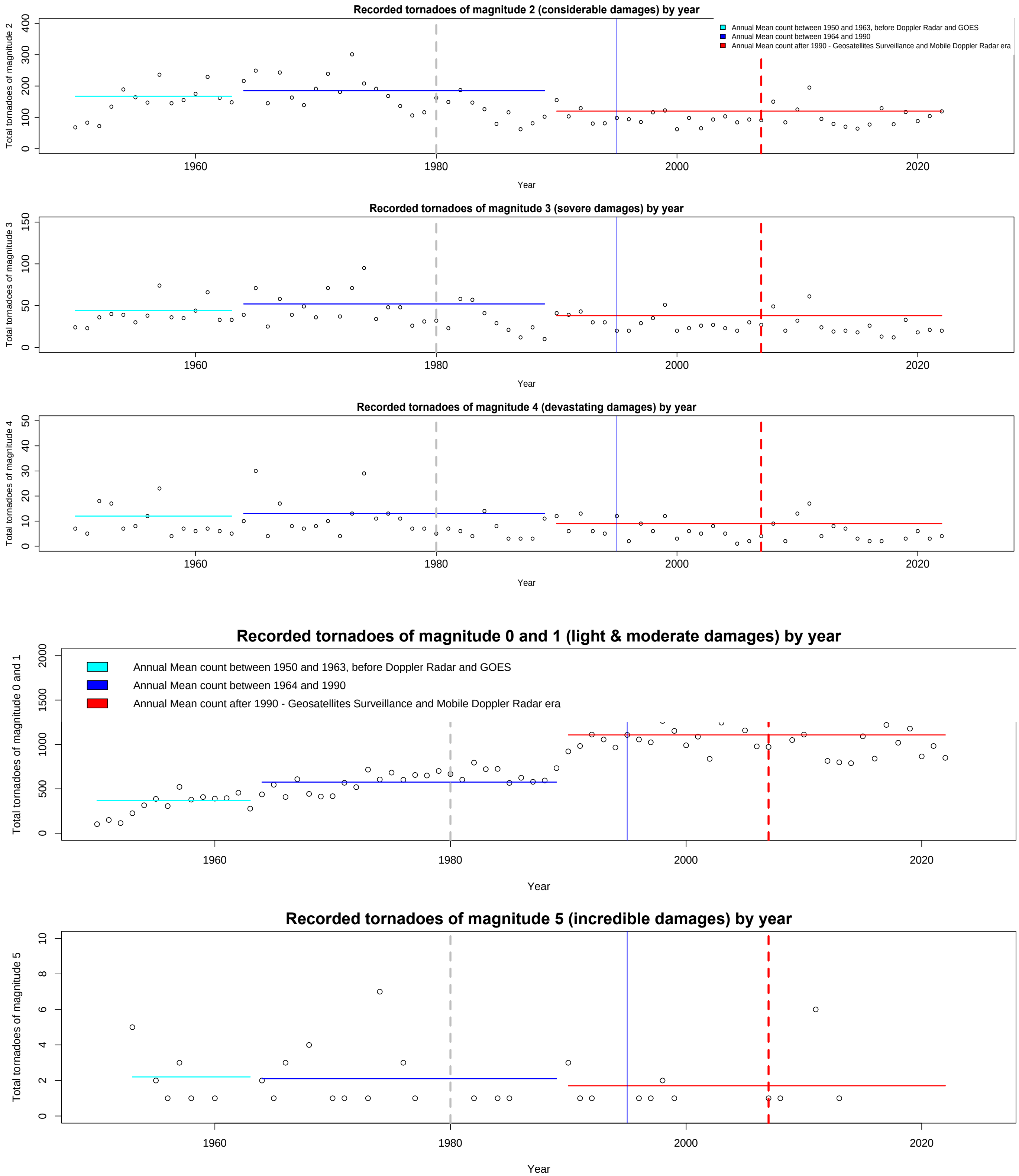

Figure 5 All tornadoes in the U.S. national database aggregated by F or EF magnitude and calendar year. The vertical lines correspond to borderline years: 1980, the start of the GOES era; 1995, when the mobile Doppler radars were implemented by all states; 2007, the start of the EF scale. The horizontal lines represent the average tornado count during three technologically and logistically distinct eras: 1950 to 1963, 1964 to 1990, and 1990 to 2022.

Top of the Document

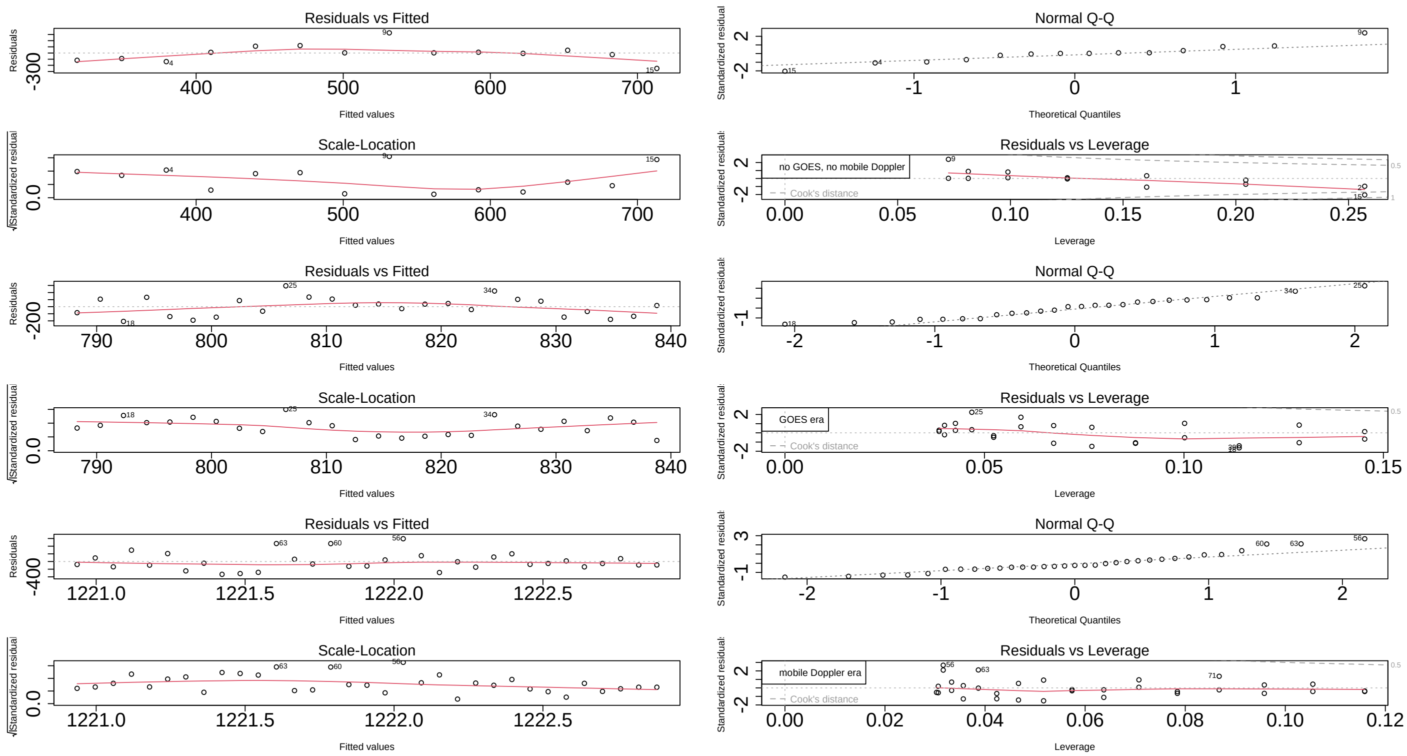

Figure 6 Graphical goodness of fit estimates of the three era-conditional linear trends of the annual tornado occurrences. The trend is truly linear if the red lines at the left-hand side are practically horizontal and all the points align on the dotted oblique line of the Normal Q-Q plots at the right-hand side.

The data aggregated on tornado strength and technological era are shown again in boxplot format, **Figure 7**. Boxplots are good at visualizing the variability of the data. The height of each blue box represents the interquartile range (IQR), from the bottom of the blue box (the 25th percentile) to the top of the box (the 75th percentile). The larger variability of the data recorded before 1990 may be attributed to the organizational growth of the national weather services, and the tornado detection technological improvements.

The horizontal thick line within the boxplot represents the median annual tornado count for each magnitude and era. The round dots represent outliers, which are defined as observations farther than (1.5)*IQR from the edge of the box. Some of the outlier tornado annual counts coincide with the historic tornado outbreaks and superoutbreaks such as those in 1965, 1974, 2008, and 2011 [20]. The whiskers, printed with dashed lines and terminated with horizontal fences, extend to the most extreme observed count that is not an outlier.

The presence of the outliers is an indication that the fluctuations around the mean of the annual tornado count are non-linear. It will be shown later in this report that tornadoes are random rare effects conditional on the geographic region, climatic conditions, and season of the year. For fixed tornado detection technology and database logistics, tornadoes are probabilistic events that can be modeled as a mixture of Poisson distributions.

Top of the Document

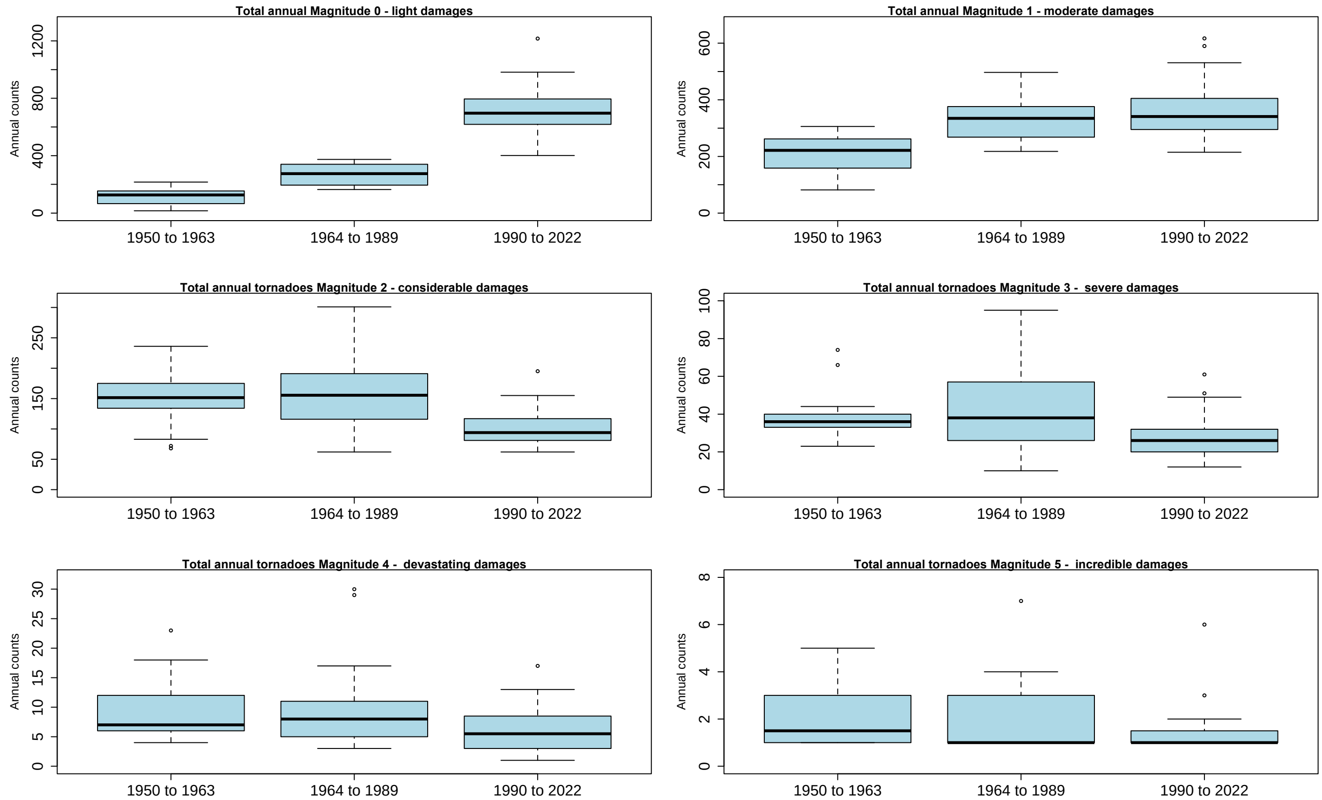

Figure 7 All tornadoes in the U.S. national database aggregated by magnitude and distinct time periods (eras). The bottom of each blue box represent the 25th percentile, respectively the 75th percentile for the top, of that collection of annual data. The height of each blue box represents the interquartile range (IQR). The horizontal thick line within the boxplot represents the median annual tornado count for each magnitude and era. The round dots represent outliers, which are farther than a (1.5)*IQR distance from the edge of the box.

Question 2: is there a shift of the geographic location of the tornadoes?

In order to assess a possible shift of the annual tornado locations, the U.S. geographic map was divided into eight areas or Zones, shown in TABLE 2. Out of these, Zone 6 is the South Great Plains region traditionally labeled as "the tornado alley" and totally or partially covers the states AR, MO, TX, LA, OK, MS, KS, and TN.

The - 90° longitude line lies close to the Mississippi River. Thus, Zone 6 includes the land West of Mississippi. Zone 4 is at the East of the -90° meridian, therefore East of Mississippi River. It totally or partially covers the states KY, MS, TN, FL, AL, AR, NC, GA, MO, IL, SC, VA, IN, and WV.

Table 2 Zones within the U.S. for tornado location trend analysis

Zone	West of -100° meridian	From -90° to -100°	From -80° to -90°	East of -80° meridian
North of 38° N latitude	**7**	**5**	**3**	**1**
South of 38° N latitude	**8**	**6**	**4**	**2**

Top of the Document

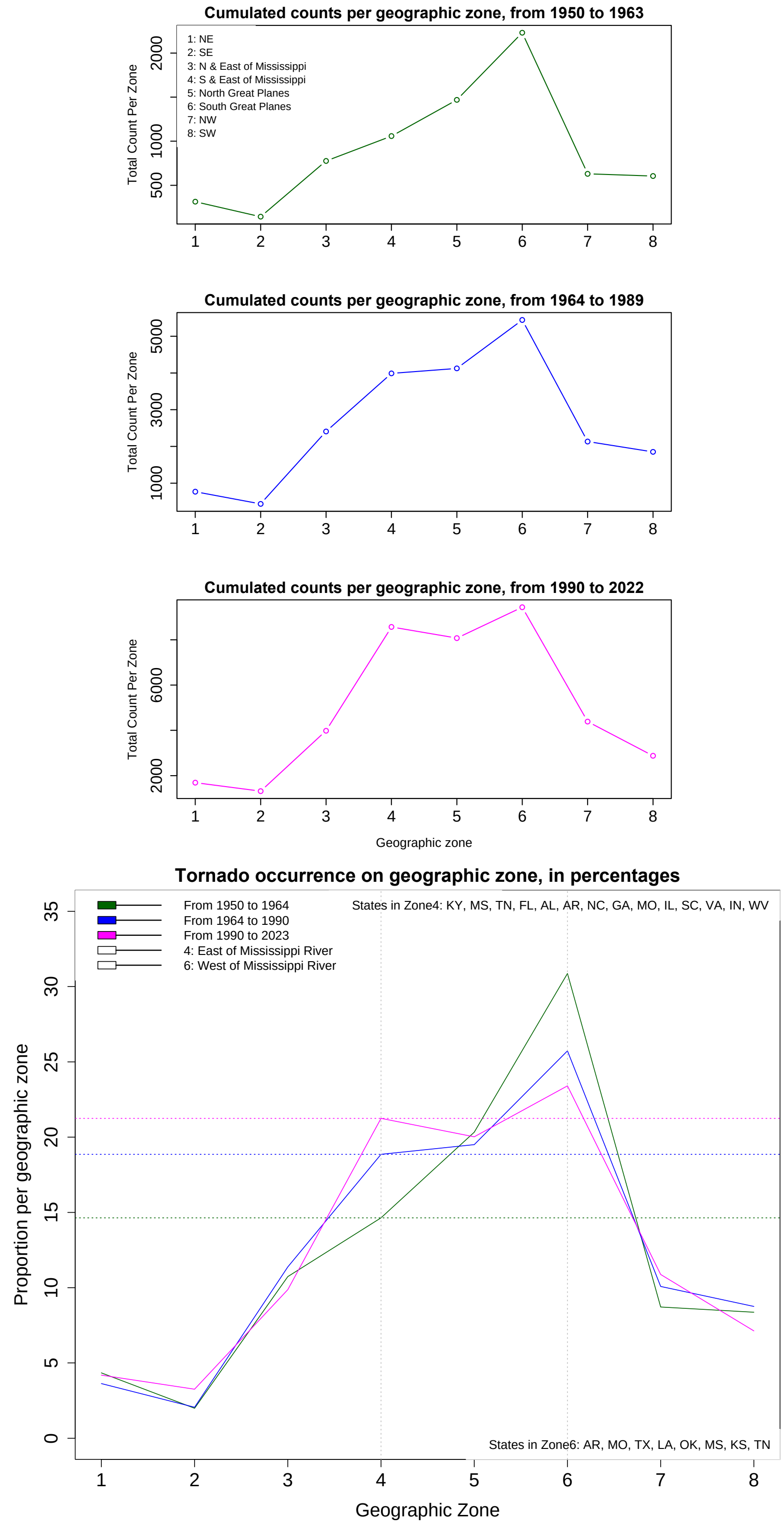

Figure 8 Tornado occurrences in eight distinct geographic zones within the U.S. during three eras.
On the abscissa, the numbers represent the geographic zones labeled from East to West and from North to South. On the ordinate, in the upper panels: total tornado counts per zone during that era. On the ordinate, lowest bottom panel: proportion of tornadoes in that zone during each era.

[Top of the Document]

The zonal percentage (proportion) was calculated by dividing the number of tornadoes observed in that zone during one era to the total number of tornadoes of that era. A change in the zonal percentage across eras implies a change in the tornadoes geographic pattern.

FIGURE 8 presents visual comparisons of the zonal counts and of the zonal percentages. It reveals a decrease of the proportion of tornado events in Zone 6 (the tornado alley) and an increase of the proportion of tornado events in Zone 4, at East of the Mississippi River. The shift of the tornadoes' geographical location from 1950's to today was checked for statistical significance with the Pearson's Chi-squared test and with Fisher's exact test for count data. Both tests confirmed a significant change in the tornado counts corresponding to the zones 4 and 6, from 1950 to 2022 (p-value <0.0005). The data revealed a statistically significant eastward shift of the tornadoes' geographic location. More research should be done, and other sources should be accessed, for verifying if the eastward shift is a real effect and not an artifact due to underreporting in the states east of the Mississippi river before 1990.

Question 3: can a numerical model estimate tornado occurrence in the greater Tulsa area?

A statistical model for the greater Tulsa area was built based on the data recorded during the mobile geosatellite surveillance and mobile Doppler radar era between 1990 and 2022 from seven adjacent counties in Oklahoma: Tulsa, Okmulgee, Wagoner, Rodgers, Osage, Pawnee, and Creek. The tornadoes initiated in these counties do not always cross the metro area but the storms accompanying the tornadic supercells could affect large parts of the city. For example, the two tornadoes in Rodgers County on June 16, 2023, caused damage and power outages of the eastern side of the city.

The annual distribution of the tornado events is illustrated in ***FIGURE 9.*** The histogram emphasizes that the annual totals vary randomly. The minimum annual total of 0 was reached in 2002; the maximum of 29 was reached in 2019. The 50^{th} percentile (the median) annual count was 7, and the standard deviation from the mean 6.4. The boxplot at the right side is skewed toward large counts which means that for half of the years there are more than seven tornadoes in the greater Tulsa area. The year 2019 represents an outlier, a rare case during the 33-year period analyzed.

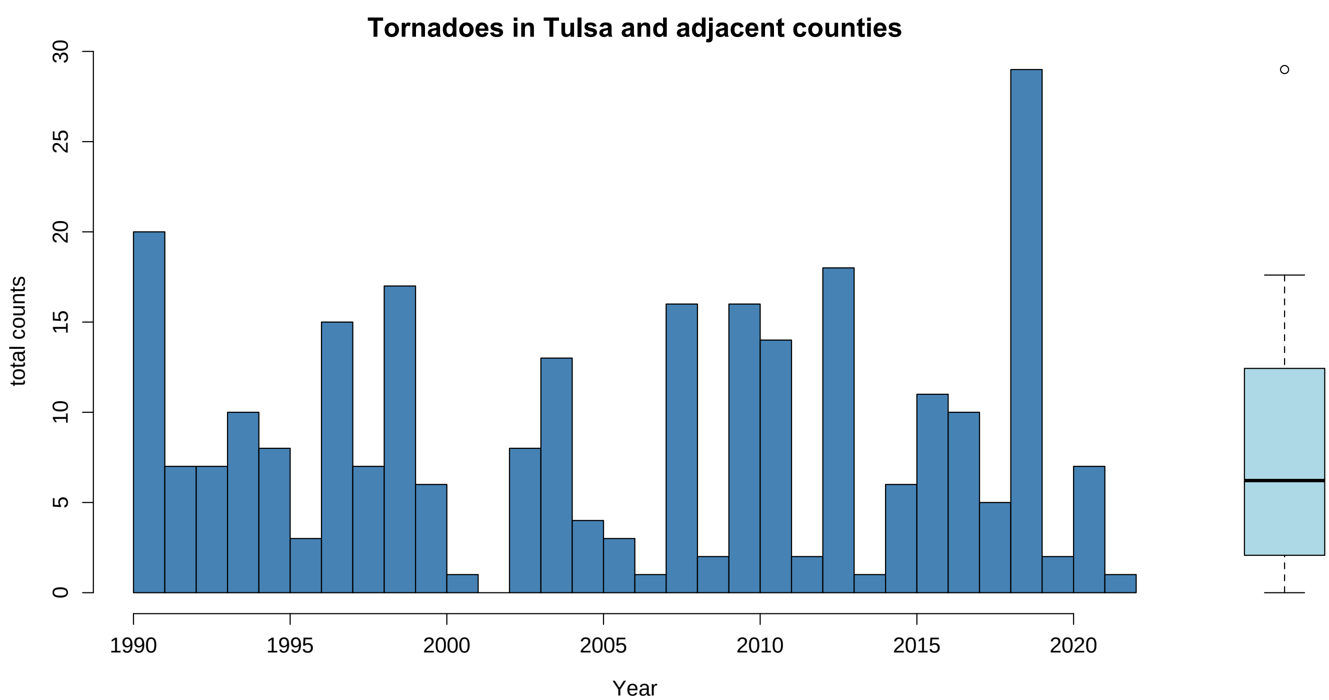

Figure 9 Total tornado occurrences in the greater Tulsa area (seven counties) for each year of the mobile geosatellite surveillance and mobile Doppler radar era. On the boxplot at the right side, the median mark is at 7 tornadoes per year and the outlier represents 2019 with its 29 tornadoes in greater Tulsa area.

[Top of the Document]

Out of 270 entries in the 33-year data-subset, six recent tornadoes are marked as EFU (unassigned magnitude). These EFU tornadoes have lengths ranging from 0.1mi to 2.0mi; none of them have recorded damage or casualties. It was preferable to keep these six entries in the statistical model therefore their strength was guessed (imputation) from their length. Although there is no simple linear relationship between tornado length and tornado strength, it made sense[1] to consider the four tornadoes shorter than 0.8mi as EF0, and the other two as EF1 during the data modeling.

Tulsa tornadoes are seasonal events with maximum counts, magnitudes, and lengths occurring usually in April and May. **FIGURE 10** illustrates, in its upper panel, the totals for each tornado magnitude across the twelve calendar months, cumulative for the 33 years of observations. The bottom panel in **FIGURE 10** illustrates the "tornado seasons", which were considered for statistical modeling.

In order to achieve statistical power for data modeling, the monthly counts were assigned to five *tornado seasons*. April and May are *tornado seasons* by themselves; June-July, August through October, and November through March are the other *seasons*. Each *season* cumulatively amassed at least 14 tornadoes in 33 years. Because the data-subset contained only five EF3 and four EF4 tornadoes, the strong tornado events were cumulated in the category "EF3 and higher."

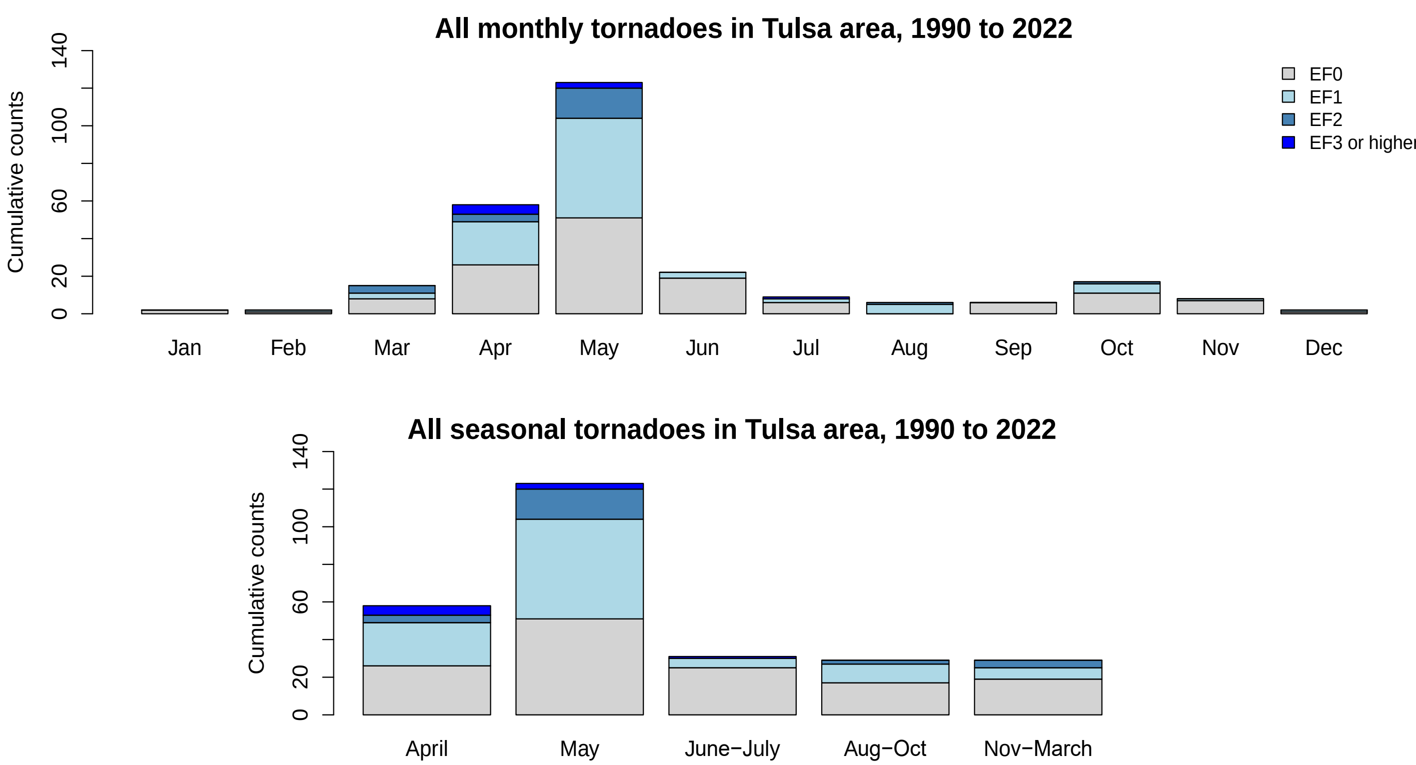

Figure 10 The monthly total of the tornado occurrences in the greater Tulsa area (seven counties) cumulated for each magnitude for 33 years. The box sizes are proportional with the total counts.

Top of the Document

[1] In the entire national database, the tornadoes shorter than 0.8mi are EF0 in a proportion of 66%, whereas less than 30% of the tornadoes shorter than 2mi have magnitudes EF2 or higher.

Tornadoes are probabilistic rare events conditioned by the season of the year and by the amount of energy. With the assumption that the detection technology and the database logistics were consistently maintained at the same level between 1990 and 2022, the annual tornado entries can be modeled as a mixture of Poisson distributions. The annual tornado events can be statistically described as negative binomial counts [21].

Several negative binomial regressions, linear and non-linear, were run first on monthly counts, then on seasonal counts, then on seasonal counts conditioned on magnitude. The resulting statistical models were computationally compared to one another through the maximum likelihood ratio. For graphical comparisons, the quantiles (percentiles) of the actual data were plotted against the quantiles predicted by each model.

FIGURE 11 represents the graphical goodness of fit of the two most accurate statistical models considered. The unconditional model (blue triangles) has two parameters, the conditional model (magenta discs) has eight parameters. The closer model-estimated points are to the blue diagonal line, the more accurate that model is.

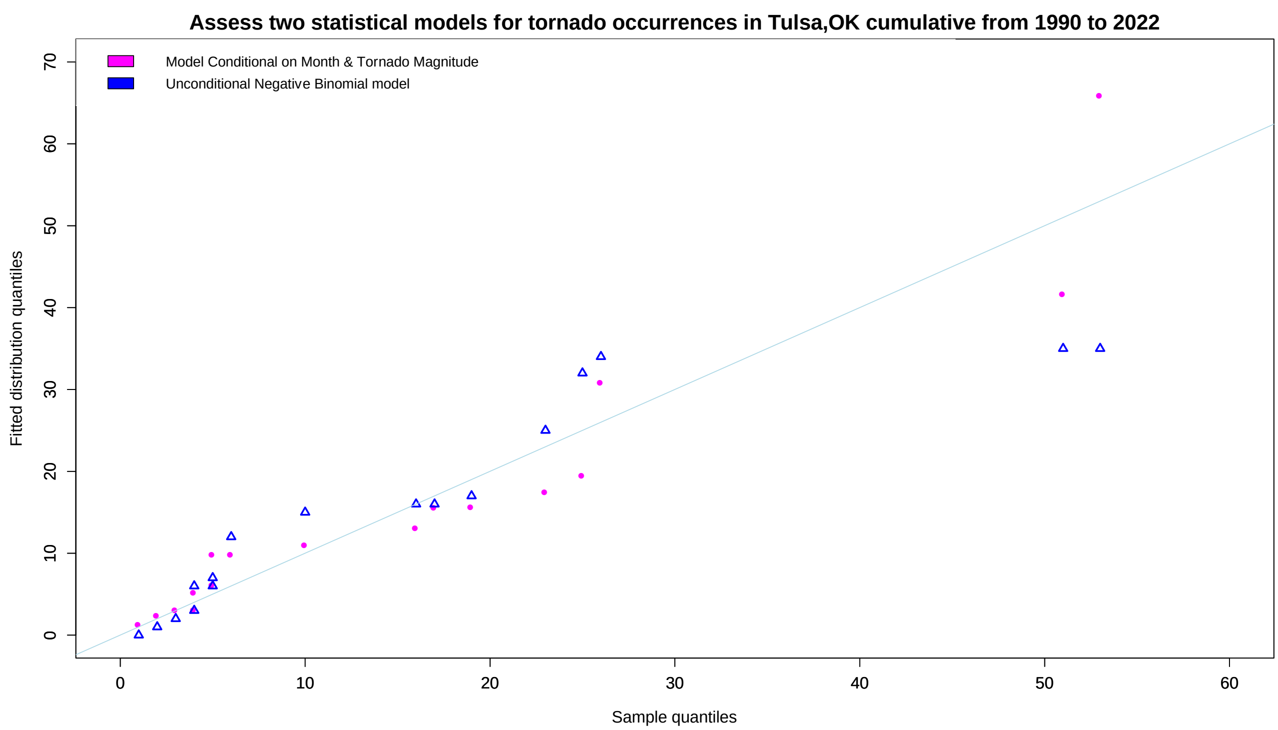

Figure 11 Graphical comparison between observed and estimated tornado counts, for two constructed statistical models. The closer the model-estimated points are to the blue diagonal line, the more accurate the model is.

The conditional model that estimated the tornado counts most accurately was retained. It is a log-linear model; in other words, it is built on the relationship between the logarithm of the counts and the two conditions predicting the tornado occurrences: *season* and *EF magnitude*. The best model has the equation:

$$log(y)= 3.4324 +(0.7568)*I_{May}+ (-0.5661) *I_{June\text{-}July} + (-0.6813) *I_{Aug\text{-}Oct} + (-0.6813)*I_{Nov\text{-}March} + \\ + (-0.4583)*I_{EF1} + (-1.6112)*I_{EF2} + (-2.5292)*I_{EF3+} \qquad \text{(Eq.1)}$$

where y is the estimated cumulative count for 33 years

I_{May} , $I_{June\text{-}July}$, $I_{Aug\text{-}Oct}$, $I_{Nov\text{-}March}$ are indicators that take the value 1 during that season and 0 otherwise

I_{EF1}, I_{EF2}, I_{EF3+} indicators that take the value 1 for tornadoes of that EF magnitude and 0 otherwise.

There is no indicator for the season *April* and strength *EF0*, but a reference value of 3.4324 associated with this case.

Top of the Document

Note that there is a positive coefficient multiplying the indicator for May, I_{May}, meaning that it is more likely to observe EF0 tornadoes in May than in April. Note that there is a negative coefficient multiplying the indicator for EF1, I_{EF1}, meaning that in any *season* it is less likely to observe EF1 tornadoes than EF0. Also, note that all the other coefficients are negative meaning that during *seasons* other than *April*, and for strengths other than EF0, tornado occurrence is less likely.

The coefficients in *(Eq.1)* are all statistically significant at 0.05 alpha-level or better. The numerical coefficients of the log-linear model *(Eq.1)* are not exact but have standard errors that must be accounted for. Generally, the probabilistic models estimate a numerical interval (range) and not a point value. It is customary to rely on the 95% confidence intervals generated by the statistical software instead of punching numbers on the calculator.

The output of the statistical software was rounded to the nearest smaller integer and listed in **TABLE 3.** The *min* and *max* values listed represent the ends of the 95% confidence interval. The meaning of the 95% confidence interval is: if one is to count all tornadoes in 100 identical *seasons*, these counts will be 95 times within the interval listed in **TABLE 3**. The annual count estimation of 2 to 18 tornadoes annually for 95 out of 100 years is consistent with the actual count distribution in **FIGURE 9**, which has 27/33 =82% of the annual counts observed in this range.

The statistical model does not specifically estimate the day and the time when a tornado will occur. For finer temporal estimations, more complex information other than that in the data file analyzed, and more environmental covariates other than the geographic location and the season of the year, are needed. The estimation of the annual tornado events across the seven counties around Tulsa could be extended to future years only if the characteristics of the 1990-2022 period remain unchanged. As the environmental characteristics are not known in totality, the data analysis and the model need to be updated yearly before trying any future estimation.

Table 3 Estimation table: 95% CI computer estimated number of seasonal tornadoes occurrences.

Season	EF0		EF1		EF2		EF3+		Season total	
	min	max	min	max	min	max	min	max	min	max
April	1	1	0	1	0	0	0	0	1	2
May	1	4	0	4	0	1	0	1	1	10
June-July	0	1	0	1	0	0	0	0	0	2
Aug-Oct	0	1	0	1	0	0	0	0	0	2
Nov-March	0	1	0	1	0	0	0	0	0	2
Annual total	2	8	0	8	0	1	0	1	2	18

Calculation examples:

(I) With 95% confidence, how many tornadoes and of what magnitude will occur during June-July 2023 in the greater Tulsa area?

From **TABLE 3**, the seasonal total will be from 0 to 2 tornadoes, with a maximum of one EFO and one EF1. Cross-validation with the 2023 Oklahoma Tornadoes [19]: there were two EF1 tornadoes in June, in Rodgers County, and none in July. The observed total was in the 95% confidence interval, but one observed tornado was stronger than estimated.

(II) With 95% confidence, how many tornadoes and of what magnitude will occur from Nov 1st, 2022, to March 31st, 2023, in the greater Tulsa area?

From **TABLE 3**, the seasonal total will be from 0 to 2 tornadoes, with a maximum of one EFO and one EF1. It is highly unlikely that tornadoes stronger than EF1 would occur during this season. Cross-validation with [19]: there was one EF1 tornado in Jan. 2023, in Osage County. The observed events were all in the 95% confidence estimated interval.

(III) With 95% confidence, how many tornadoes and of what magnitude will occur from Nov 1st, 2022, to Oct 31st, 2023, in the greater Tulsa area?

From **TABLE 3**, the annual total will be between 2 and 18 tornadoes. There will be a minimum of two EFO, and a maximum of one EF3+, tornadoes.

Top of the Document

Conclusion

Fact checking if the tornadoes in the U.S. are getting worse revealed that the claim "researchers haven't seen an increase in their [tornado] frequency, and measurements" is incorrectly formulated. There is an increase in the number of annually recorded EF0 and EF1 tornadoes since the 1950's, but this change appears to be most likely a consequence of the advancements in detection technology and data recording within the U.S. There is solid evidence of three distinct eras of tornado data recording. For confirming trends across these three different eras, supplementary complex information from other sources is needed.

The database analyzed offered statistically significant evidence of an eastward shift of the tornadoes' geographic locations, from 1950's to 2020's. If this location drift is a real effect and not an artifact due to the underreporting in the states east of the Mississippi river before 1990, then there is and will be an increase of the tornado damage within states such as Louisiana and Mississippi.

A negative binomial log-linear model was built for estimating the tornado occurrences in seven Oklahoman adjacent counties: Tulsa, Okmulgee, Wagoner, Rodgers, Osage, Pawnee, and Creek. Although not all the tornadoes initiated in these counties have traversed the metro, the storms accompanying them have affected, and could affect in the future, a large part of the population in the greater Tulsa area.

The statistical model for the greater Tulsa area estimates the tornado events occurring during long time intervals (month, season, or year). This model was validated with the data collected between 1990 and 2022. The model does not specifically estimate the day when a tornado will occur. For finer temporal estimations, more complex information than that stored in the database analyzed, and more environmental covariates other than the geographic location and the season of the year, are needed.

The model estimating the annual tornado events across the greater Tulsa area could be employed for forecast if and only if the characteristics of the 1990-2022 period remain unchanged in the future. As many of the environmental characteristics are not known, and as new observations are added periodically into the tornado database, it is recommended to update yearly the model presented in this report.

People living in the greater Tulsa area could comment that the information offered by this statistical model is superfluous, that statements such as "it is highly unlikely that tornadoes stronger than EF1 would occur during the November-March season" are common knowledge, and that the statistical analysis in this report has no practical significance. An answer to this criticism is that a statistical model offers viable data-backed algorithms, which can be implemented into unsupervised intelligent devices (AI). The statistical model is of practical importance for weather monitoring automation, and for detecting tornado pattern changes in the future. Although this model cannot substitute the tornado warnings, it offers useful reference and it helps with planning and tornado preparedness.

Top of the Document

References

[1] SPC Preliminary reports for 06/17/23 https://www.spc.noaa.gov/exper/archive/event.php?date=20230617
[2] Storms cause damage across Tulsa https://www.newson6.com/june-17-storms
[3] Are tornadoes in the US getting worse? https://www.bbc.com/reel/video/p0frj4m5/are-tornadoes-in-the-us-getting-worse-
[4] Governor Abbott issues disaster declaration for Matador tornado response https://gov.texas.gov/news/post/governor-abbott-issues-disaster-declaration-for-matador-tornado-response
[5] S. Corfidi, A brief history of the Storm Prediction Center https://www.spc.noaa.gov/history/early.html
[6] Interview with Dr. Lou Wicker, https://www.youtube.com/watch?v=moOQfl-c6yA
[7] U.S. Tornadoes (1950-2022), accessed on 06-20-2023 https://www.spc.noaa.gov/wcm/#data

[8] Seven decades of tornadic activity in the United States https://www.kaggle.com/datasets/danbraswell/us-tornado-dataset-1950-2021
[9] Gensini, V.A., Brooks, H.E., Spatial trends in United States tornado frequency. *npj Clim Atmos Sci* **1**, 38 (2018). https://doi.org/10.1038/s41612-018-0048-2
[10] Tornado FAQ https://www.spc.noaa.gov/faq/tornado/
[11] Romanic, D et al, Oklahoma tornado risk and variability: A statistical model. *j* ijdrr 16: 19-32 (2016). https://doi.org/10.1016/j.ijdrr.2016.01.011
[12] 2013 Moore, Oklahoma, tornado https://www.worldvision.org/disaster-relief-news-stories/2013-moore-oklahoma-tornado-facts
[13] The Enhanced Fujita scale https://www.weather.gov/oun/efscale
[14] Gall, M. et al, When do losses count? Six Fallacies of Natural Hazards Loss Data https://www.spc.noaa.gov/wcm/data/2008bams.pdf
[15] U.S. TORNADOES (1950-2022) Single tracks. No state segments or continuing county info, accessed on 06-20-2023 https://www.spc.noaa.gov/wcm/data/1950-2022_actual_tornadoes.csv
[16] Errors, oddities and artifacts in U.S. tornado data, 1995-2001. https://www.spc.noaa.gov/publications/edwards/oddities.pdf
[17] The Comprehensive R Archive Network https://cran.r-project.org/
[18] Tulsa County, OK Tornadoes (1875-Present), accessed on 06-28-2023. https://www.weather.gov/oun/tornadodata-county-ok-tulsa
[19] Oklahoma Tornadoes, accessed on 06-28-2023. https://www.weather.gov/oun/tornadodata-ok-2022 ; https://www.weather.gov/oun/tornadodata-ok-2023
[20] Corfidi, S. (2013). "Noteworthy North American Tornado Outbreaks; https://en.wikipedia.org/wiki/Tornado_outbreak_sequence
[21] Getting started with negative binomial regression https://data.library.virginia.edu/getting-started-with-negative-binomial-regression-modeling/; https://en.wikipedia.org/wiki/Negative_binomial_distribution#Statistical_inference

Top of the Document