A Preliminary Approach to Investigate Recent Increases in New York Tornado Activity: A Wavelet Analysis and Forecasting Approach using R

Hubbs, D.[1], Gonzalez, X.[1], Rezki, A.[1], Tecusan, K.[1], Marsellos, A.E.[1]
[1]*Dept. Geology, Environment and Sustainability, Hofstra University, NY, USA.*

Abstract

When considering tornadoes, one's first thought likely veers towards the infamous "Tornado Alley" that is a region in the Midwest where tornadoes are notoriously frequent and severe. New York's tornado frequency, in contrast, is nowhere near as high. Despite this, historical records show an increase in tornado events within the NY State. This raised questions about the existence of a pattern tied to these events and subsequent geohazards intensity. Utilizing tornado data gathered from MPNNow's tornado database, we aimed to both investigate a potential pattern and forecast future tornado rates. Wavelet analysis detects frequency variations over time and after applying the Kolmogorov-Zurbenko (KZ) filter, which excludes high-frequency components, and reveals annual and multi-annual cycles. We modelled our data using the AutoRegressive Integrated Moving Average (ARIMA) function, allowing for a general forecast of potential future tornado frequencies. Geologically, tornadoes are known to affect multiple processes, including but not limited to soil erosion, sediment redistribution, and mass wasting. The existence of a multi-annual tornado-driving cycle would provide new insights in our understanding of tornadoes, which could prove useful in preparation and damage mitigation efforts.

Introduction

In the United States, tornadoes are most commonly associated with the notorious "Tornado Alley"[12], a region spanning through the Midwest which consistently experiences a high frequency of tornadoes. New York State, by contrast, is not typically regarded as a tornado hotspot. This low rate of occurrence has often led researchers to disregard the state in favor of those more prone, leaving the long-term tornado patterns in New York comparatively less studied.

Recently, however, scientists have documented a notable rise in tornado events within New York, with the increase in the trends now versus when the data was initially recorded. This observation has raised questions regarding the possibility of a cyclical or trend-related phenomena driving these events. Beyond direct, immediate impacts, tornadoes are known to affect numerous geological processes, including soil erosion, sediment redistribution, and mass wasting. As such, the confirmation of an existing, tornado-driving pattern could significantly advance our relatively limited understanding of tornadoes, aiding with preparational and risk mitigation related efforts.

To address this question, we analyzed historical tornado data from MPNNow's tornado database[15], which we supplemented with official reports from the National Oceanic and Atmospheric Administration (NOAA). To this data, we applied wavelet analysis, which detects frequency variations over time alongside the Kolmogorov-Zurbenko (KZ) filter, which removes low-frequency components such as recognized annual cycles from the time series. We also employed Fourier terms with the AutoRegressive Integrated Moving Average (ARIMA) modeling, which captured seasonalities that were then used to forecast future tornado activity.

Our study involved the pursuit of three distinct goals, which were:

1. To either determine or disprove the existence of a multi-annual tornado-driving cycle.
2. To, assuming the existence of the former, forecast future tornado frequencies.
3. To discuss the geological and environmental effects of increased tornado activity.

Methods

This study employs a multi-step approach to analyze and forecast tornado activity in New York, integrating data preprocessing, time series smoothing, wavelet analysis for time-variant periodicity detection, and ARIMA modeling enhanced with Fourier terms to capture seasonality in R programming language. Initially, tornado event data were retrieved from the website[15] The raw timestamps, recorded in the “Date/Time” column, were imported into a data frame using the R package *readr* and then converted into R date objects via *lubridate’s mdy_hm* function with the "*America/Chicago*" time zone.

The tornado event dataset was successfully filtered to include daily records from May 6, 1952, to September 10, 2024. Following the removal of duplicate dates and missing entries, a complete daily sequence was generated. Days with any tornadoes were marked with a value of one, and days with no tornado events were marked with a value of zero (Fig. 1A). This preprocessing ensured that our subsequent analyses operated on a continuous and reliable time series.

A preliminary plot of the raw daily event counts revealed notable short-term fluctuations and seasonality. These fluctuations underscored the need for smoothing techniques to highlight underlying trends.

Only the date component was retained to facilitate daily aggregation. Since our dataset records tornado events on a binary basis indicating whether an event occurred on a given day or not, the application of a 30-day rolling sum and subsequent moving averages was required to transform the binary dataset into a continuous representation of tornado frequency over time. Functions from *dplyr* and *lubridate* were used to generate a continuous date series and replace NA (non available tornado event) values in the event counts with zeros.

Following data acquisition and preprocessing, a 30-day rolling sum of daily tornado events was computed using the *rollapply* function from the *zoo* package, aggregating the data into a smoother representation over short periods. To remove higher-frequency noise such as seasonal cycles and short-term fluctuations, a *Kolmogorov–Zurbenko* (*KZ*) filter was applied as a superior moving average that preserves the ‘tails’ of the dataset.

Initially, a *KZ* filter with a 365-day window and three iterations smoothed the 30-day rolling sum to remove annual seasonal variability.[16] For subsequent wavelet analysis, a more advanced double-filtering approach was implemented: first, a *KZ* filter with parameters of 365 days at five iterations further removed annual signals (Fig. 1C), followed by a second *KZ* filter with a 7-year window and one iteration to eliminate longer-term phenomena such as *ENSO* effects (Fig. 1D). The sequential application of the (*KZ*) function yielded a heavily smoothed time series that formed the basis for further analysis.

Wavelet analysis was then performed to detect in the frequency domain dominant time-varying periodicities in the tornado time series. This method was chosen over a simple periodogram (such as TSA, periodogram function) because of its superior ability to capture nonstationary periodic signals. A new data frame was constructed containing the smoothed tornado event signal after double KZ filtering, and the *WaveletComp* package's *analyze.wavelet* function was employed with key parameters set to a 1-day time step (dt = 1), a resolution of 1/10 for each octave (dj = 1/10), and a period range from 365 days (1 year) to 10,950 days (30 years) to include all atmospheric long-term phenomena. Statistical significance was assessed through *p-values* generated from 10 simulations, and the wavelet power spectrum was visualized using *wt.image*, with an accompanying periodogram (a plot of average power versus period) where local peaks—indicative of dominant cycles such as a prominent 13 to 17-year cycle and possible ENSO-related signals were detected using the *scorepeak* package.

To forecast future tornado activity, the smoothed signal was converted into a time series object with a frequency of 365, reflecting daily data with annual seasonality. Stationarity was evaluated using *ndiffs* to determine the necessary order of differencing, with seasonal differencing (order = 1) enforced. Fourier terms, generated via the *fourier* function with eight harmonics, were incorporated as external regressors to better capture complex seasonal patterns. A Seasonal ARIMA model with non-seasonal order (2, d, 2) and seasonal order (1, 1, 1) was then fitted using the Arima() function, including the Fourier terms via the *xreg* parameter (lags), and was subsequently used to forecast tornado activity for the next 2 years (approximately 730 days). Diagnostic checks, including *ACF* and *PACF* plots of the residuals, ensured the adequacy of the model fit, while forecast confidence intervals were computed and visualized alongside the forecast trajectory.

Throughout the analysis, a comprehensive set of visualizations was generated to illustrate each stage (Fig. 1) from the raw and processed data (including daily tornado event counts, 30-day rolling sums, and *KZ-filtered* time series) to the wavelet spectrum and periodogram highlighting cyclical patterns, and finally to the ARIMA forecast plots (Fig. 2) with diagnostic *ACF* and *PACF* analyses. Vertical reference lines and custom axis ticks were added to the forecast plots to indicate notable dates, ensuring a clear interpretation of the time axis.

Results

Analysis of tornado events in New York, spanning from May 6, 1952 to September 10, 2024, revealed a long-term tornado frequency that exhibits an overall upward trend. Recent data indicates a significant increase, with the peaks present around 2023 and 2024 being noticeably higher than those of other, previously observed peaks. In addition, a dominant recurring cycle of approximately 13 to 17 years was observed, suggesting that fluctuations in tornado frequency follow a predictable long-term pattern rather than occurring at random.

Secondary periodicities in the 2–7 year range were also detected, which may reflect the influence of interannual climate variability, such as ENSO effects. Forecast projections for the upcoming 2 years indicate that the current peak is likely to be followed by a gradual decline in tornado frequency; however, even the projected trough remains above the lower levels observed in previous cycles, implying a persistent elevation in baseline activity. The forecast confidence intervals (Fig. 3) further support the significance of the identified decadal cycle and the anticipated post-peak decline for future risk assessment.

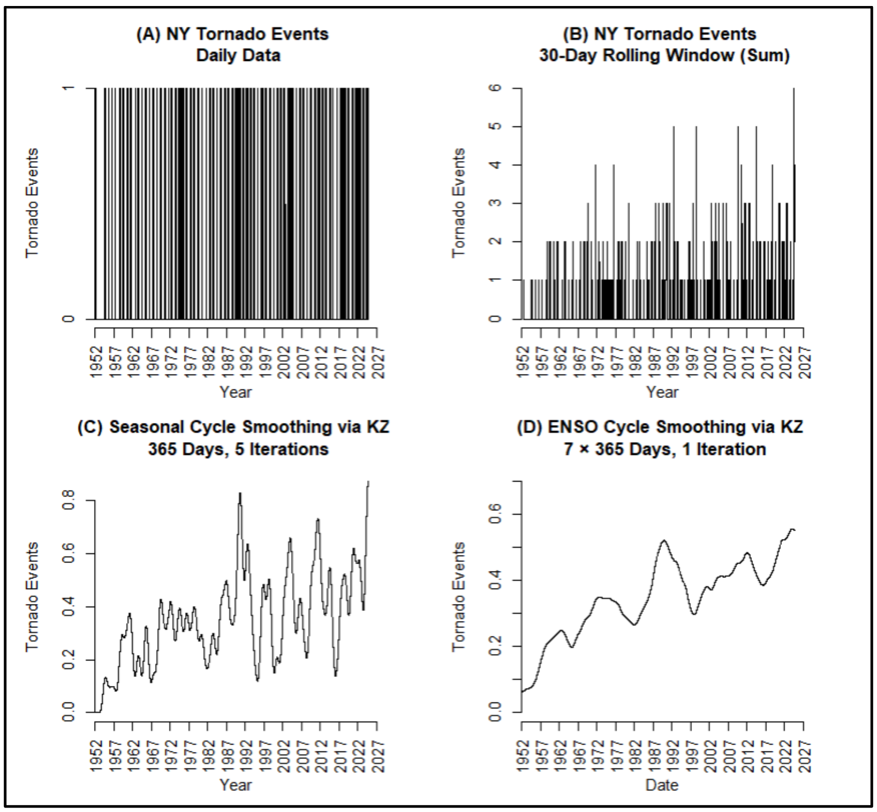

Figure 1: Time series analysis of tornado events in New York from 1952 to 2027. (A) Daily tornado occurrences. (B) 30-day rolling sum of tornado events. (C) Seasonal smoothing using the Kolmogorov-Zurbenko (KZ) filter with a 365-day window and five iterations. (D) Multi-year smoothing using the KZ filter with a 7-year window, highlighting long-term trends.

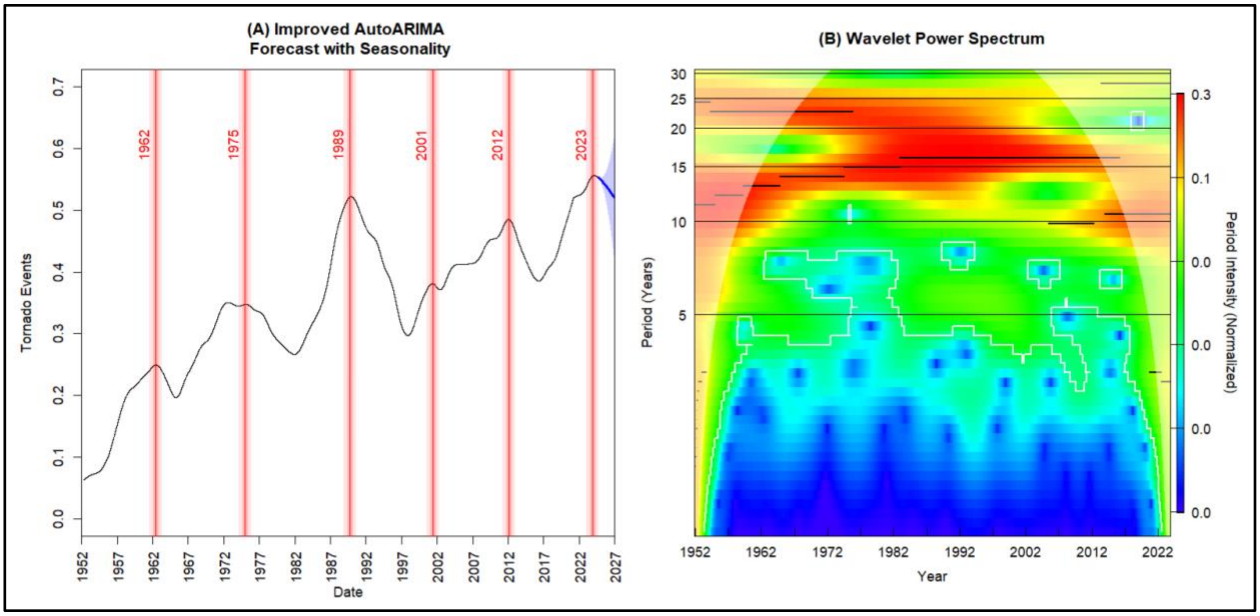

Figure 2. (A) Periodogram of Average Wavelet Power for New York Tornado Events. (B) Wavelet Power Spectrum of New York Tornado Events.

The wavelet analysis shows significant periodicity in tornado events, with strong peaks every 13 to 17 years, indicating a recurring cycle in the number of tornado events. The wavelet power spectrum highlights a consistent 13 to 17-year cycle in tornado activity over time, with areas of high power (red) indicating periods of stronger recurrence, and low power (blue) indicating periods of little to no recurrence (right; Fig. 2). AutoARIMA forecasts tornado activity up to 2027 (left; Fig. 2), demonstrating a likely decrease in tornado activity, as implied by the visual trend.

Discussion and Conclusions

Our analysis identified a recurring 13-17 year mostly periodicity in tornado occurrences. Additionally, we observed a steady increase in the localized minima and maxima over time, suggesting a long-term upward trend in tornado activity. Shorter decadal modulations were also detected, likely influenced by interannual climate variability. Specifically, ENSO cycles (2-7 years) appear to impact shorter timescales, contributing to multiple localized peaks and valleys between the broader 13-17 year periodic patterns, as shown in Figure 1D.

Tornadoes are known to be more frequent during La Niña phases. While La Niña's effects on the Northeast are generally weaker compared to other regions, our findings suggest a possible correlation, as the detected minima and maxima in tornado activity coincide with both La Niña and El Niño cycles[10].

Beyond their meteorological significance, tornadoes have substantial geological and environmental impacts. In regions such as Long Island, their high winds can accelerate soil erosion, particularly in areas with loose glacial till (silt, sand, and gravel). This erosion can lead to the displacement of nutrient-rich topsoil, potentially causing localized nutrient deficiencies. Tornadoes can also destabilize slopes, increasing the likelihood of landslides. Furthermore, tornadoes are frequently accompanied by heavy rainfall and severe thunderstorms, which exacerbate flooding in areas with inadequate drainage systems, increasing the risk of water contamination. When tornadoes pass over bodies of water, they may generate waves or even transform into waterspouts, disturbing bottom sediments and creating additional environmental concerns[11].

Using ARIMA modeling, we generated a forecast of tornado activity extending to 2027. Our results suggest a decrease in tornado frequency following the current peak. Figure 2A illustrates this prediction, where the blue line represents the projected trend, while the surrounding shaded region, a 95% confidence interval, demonstrates the model's reliability. This suggests that while fluctuations will continue, the general uptrend is predictable with a cyclical nature in tornado activity, with a periodicity of 13-17 years.

References

1. Roesch, A., & Schmidbauer, H. (2018). *WaveletComp: Computational Wavelet Analysis*. R package version 1.1. https://CRAN.R-project.org/package=WaveletComp
2. Chan, K. S., & Ripley, B. (2022). *TSA: Time Series Analysis*. R package version 1.3. https://CRAN.R-project.org/package=TSA
3. Wickham, H., Hester, J., & Bryan, J. (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5. https://CRAN.R-project.org/package=readr
4. Spinu, V., Grolemund, G., & Wickham, H. (2023). *lubridate: Make Dealing with Dates a Little Easier*. R package version 1.9.3. https://CRAN.R-project.org/package=lubridate
5. Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://CRAN.R-project.org/package=dply
6. Zhou, H., & Wong, W. K. (2021). *kza: Kolmogorov-Zurbenko Adaptive Filters*. R package version 4.2. https://CRAN.R-project.org/package=kza
7. Zeileis, A., & Grothendieck, G. (2005). *zoo: S3 Infrastructure for Regular and Irregular Time Series*. Journal of Statistical Software, 14(6), 1–27. https://doi.org/10.18637/jss.v014.i06
8. Hyndman, R., & Khandakar, Y. (2008). *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03
9. Chang, W., & Park, W. (2024). *scorepeak: Scoring and Ranking Peaks of Curves*. R package version 1.1.0. https://CRAN.R-project.org/package=scorepeak
10. National Oceanic and Atmospheric Administration. (2021, March 18). *El Niño, La Niña, and tornadoes*. NOAA Climate.gov. https://www.climate.gov/news-features/blogs/enso/enso-and-tornadoes
11. Enviroliteracy. (2024). *How does a tornado affect the environment?* Enviroliteracy. https://enviroliteracy.org/how-does-a-tornado-affect-the-environment/?utm_source=chatgpt.com
12. Kuhne, M. (n.d.). *What is Tornado Alley?* AccuWeather. Retrieved March 31, 2025, from

https://www.accuweather.com/en/weather-news/what-is-tornado-alley-2/432271
13. Scibior, V. (2024, September 11). *Record setting tornado season across the United States and New York State.* AccuWeather. Retrieved from https://cnycentral.com/weather/weather-wisdom/record-setting-tornado-season-across-the-united-states-and-new-york-state
14. Chen, Yansong, Chengyi Zhao, and Hai Zhi. *"Analysis of ENSO Event Intensity Changes and Time–Frequency Characteristic Since 1875." Atmosphere*, vol. 15, no. 12, 2024, https://doi.org/10.3390/atmos1512XXXX.
15. MPNNow. (n.d.). *Tornado archive: New York.* Retrieved March 31, 2025, from https://data.mpnnow.com/tornado-archive/new-york/
16. National Severe Storms Laboratory. (n.d.). *Tornadoes*. National Oceanic and Atmospheric Administration. Retrieved March 31, 2025, from https://www.nssl.noaa.gov/education/svrwx101/tornadoes/#:~:text=The%20peak%20%E2%80%9Ctornado%20season%E2%80%9D%20for,is%20in%20June%20or%20July.

Credit Authorship Contribution Statement

Hubbs, D.: Code production, figure construction, formatting, and captions, primary editing

Gonzalez, X.: Discussion, conclusion, literature review, general editing

Rezki, A.: Abstract, introduction, general editing

Tecusan, K.: Abstract, introduction, general editing

Marsellos, A.E.: Supervision, Code production, literature review, methods and results