Modeling Tropical Cyclones in the North Atlantic Using an Extreme Value Framework

Abigail Kinaro | Kent Codding | Max Bahar
([Team 77])

Abstract

This project applies Extreme Value Theory (EVT) to model the intensities of North Atlantic tropical cyclones, focusing on peak wind speeds and minimum central pressure. By fitting the Generalized Extreme Value (GEV) distribution to the observed data, we find that extreme storm behavior is accurately characterized by the Fréchet distribution. Results reveal an increase in maximum wind speeds over the past 50 years, likely linked to global warming, and highlight the limitations of pre-satellite data. These findings underscore EVT's potential to enhance storm forecasting and provide valuable insights into the relationship between climate change and hurricane intensity.

1 | Introduction

With rising global temperatures, understanding and predicting extreme weather events like hurricanes and tropical storms is increasingly critical. Scientific research shows that a warmer Earth contributes to more intense storms, driven by higher sea surface temperatures and increased moisture in the air that fuel storm systems with greater energy[1]. Global warming intensifies the process of evaporation and moisture accumulation, creating stronger winds and potentially more severe storm outcomes. The impact of climate change on storm intensity is not only a pressing environmental concern but also a crucial area of study for disaster preparedness and mitigation efforts.

In addition to measures of wind speed, storm intensity can also be estimated by its central pressure, with lower central pressure generally corresponding to more powerful and destructive storms[2]. As storms draw energy from the warm ocean surface, the pressure within the storm system drops, enhancing wind speeds and overall storm severity. To model and predict extreme events in tropical cyclones, we turn to Extreme Value Theory (EVT), which provides a framework specifically suited for analyzing rare and significant deviations from typical values.

In this project, we apply the Generalized Extreme Value (GEV) distribution—an EVT-based approach—to real storm data from the North Atlantic. By fitting this distribution to storm characteristics, including maximum wind speeds and central pressure, we aim to evaluate its effectiveness in capturing the behavior of the most intense storms. Furthermore, we observe the behavior of the empirical data and the fitted GEV distributions over time, analyzing how the changing climate has affected tropical cyclone intensity over the past several decades.

Through this analysis, we aim to demonstrate the utility of EVT in modeling extreme weather events over time, contributing valuable insights for meteorology and disaster planning in the context of a changing climate. Understanding and improving our predictive capabilities for such events is essential to mitigate their economic, environmental, and societal impacts.

2 | Model Description

Our approach to modeling the distribution of maximum wind speeds and minimum central pressure of hurricanes and tropical storms relies on the Fisher-Tippett-Gnedenko theorem, which states that the maximum (or minimum) of a sample of independent, identically distributed variables can only converge in distribution to one of three possible distribution families: the Gumbel distribution, the Fréchet distribution, or the Weibull distribution. The standardized probability density function for the Generalized Extreme Value (GEV) distribution can be expressed as:

$$f(x, c) = exp(-\ (1 - cx)^{1/c})(1 - cx)^{1/c-1}$$

where $-\infty < x \leq 1/c$ if $c > 0$ and $1/c < x \leq \infty$ if $c < 0$. The parameter c is most commonly referred to as the shape parameter. If $c = 0$, the distribution follows a Gumbel distribution. If $c > 0$, the distribution follows a Weibull distribution. If $c < 0$, the distribution follows a Fréchet distribution. In addition to the shape parameter, we are also interested in the mean μ and standard deviation σ of the distribution. We hope to examine all three of these parameters over time to analyze the effect of global warming on the intensity of tropical cyclones. To do so, we will group tropical storms into 5-year time periods and fit the GEV distribution to each subset of data. In using 5-year windows, each period should have a sufficient number of tropical cyclones to adequately fit the GEV distribution.

For this iteration of the project, we will be analyzing the behavior of tropical cyclones in the North Atlantic basin using data sourced from the National Oceanic and Atmospheric Administration's (NOAA) HURDAT2 dataset[3] which is accessed through the *tropycal* Python module[4]. The variables of interest are *vmax* which is the maximum 1 minute average wind speed of the tropical cyclone in knots at an elevation of 10 meters with unobstructed exposure and *mslp* which is the minimum central pressure of the tropical cyclone. The data contains tropical cyclones from 1851 to 2023, but we will only be using observations after 1979 as the quality of pre-satellite measurements before then are likely unreliable.

To evaluate the fit of the GEV distribution to the empirical data, we utilize the Kolmogorov–Smirnov test statistic, which compares the underlying distribution of a sample of data against a given distribution. This test measures the largest absolute difference between the two distribution functions over all possible values. This means lower values of the test statistic imply a better fit of the GEV distribution to the data. In the context of our analysis, we believe that values below 0.15 indicate an acceptable fit of the GEV distribution to the data.

For the final project, we plan to use additional data such as sea surface temperature or atmospheric temperature to examine how climate change has affected the behavior of extreme values in tropical cyclones. However, for this milestone, we will rely on the widely documented phenomenon that global average sea surface temperatures have risen consistently over the past 50 years[5]. Therefore, if we observe changes in the behavior of extreme values in wind speeds and central pressure, it is reasonable to speculate that these variations can be attributed to a warming ocean. In the future, we hope to leverage more robust statistical analyses to better understand the relationship between climate change and tropical cyclone intensity.

3 | Analysis

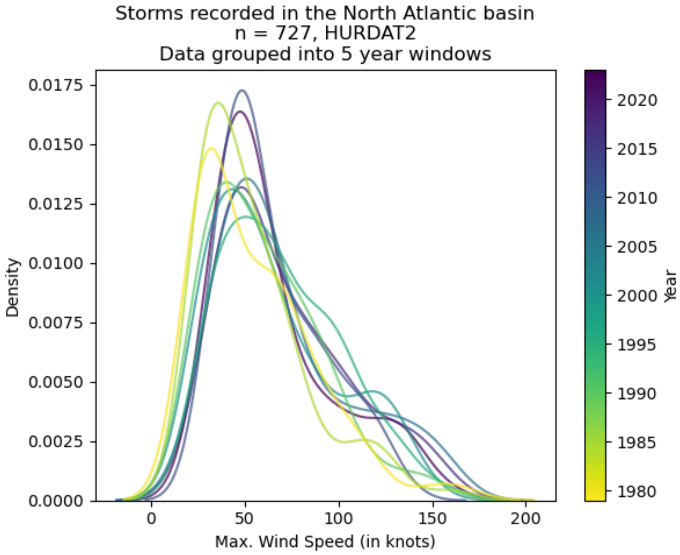

Figure 3.1

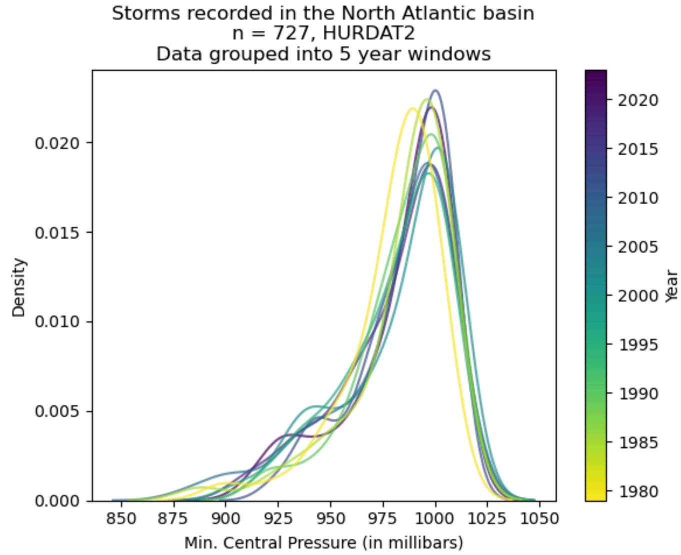

Figure 3.2

First, we examine the distribution of the extreme values in the observed data over time. Figure 3.1 shows the different distributions of tropical cyclone maximum wind speeds since the 1980s, with each kernel density plot representing 5-year periods of tropical storms. We can see that the right tail of the density plot has grown larger in more recent time periods, suggesting that it has become more likely for tropical cyclones in the North Atlantic to have higher maximum wind speeds.

On the other hand, the kernel density plots in Figure 3.2 are less straightforward in showing trends in minimum central pressure of tropical storms. The behavior of the left tail over time suggests that there were a higher proportion of tropical cyclones with a minimum central pressure lower than 900 millibars in the past. However, the proportion of tropical cyclones with minimum central pressure of around 925 millibars has grown in recent years. This change suggests that while extreme low pressures are less common now, moderately low pressures have become more likely.

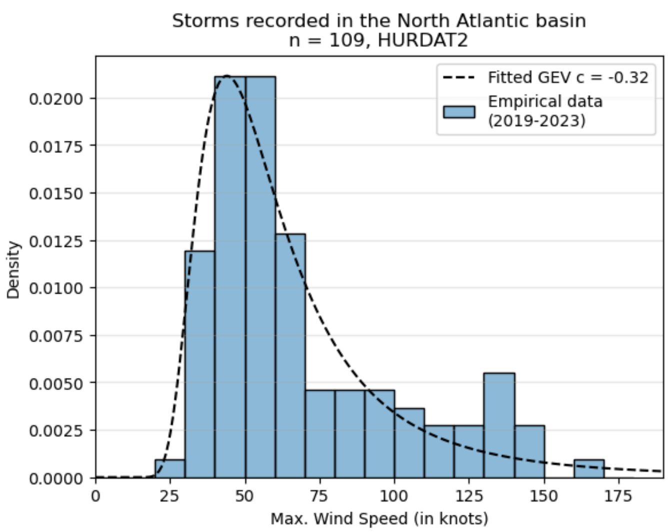

Figure 3.3

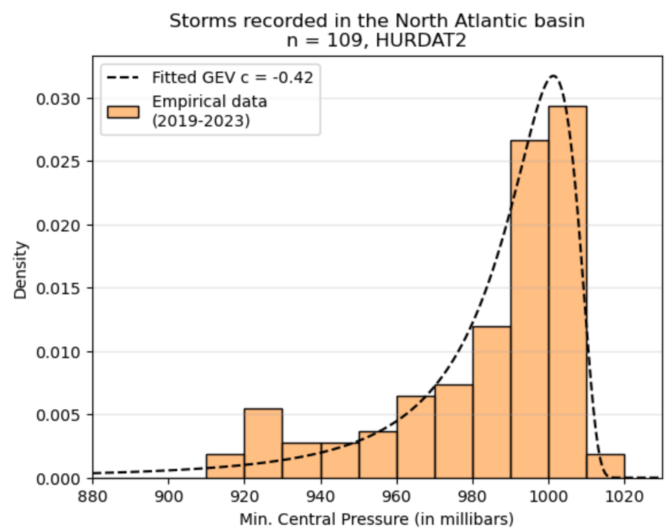

Figure 3.4

Next, we can visually compare the distribution of the empirical data with the probability density function with the fitted parameters to see if extreme values for wind speeds and central pressure of tropical cyclones can be characterized by the GEV distribution. In Figure 3.3, we can see that the GEV distribution fits relatively well to maximum wind speed data from the past 5 years, suggesting that the selected model can accurately model the behavior of extremes in tropical cyclone wind speeds. Specifically, we obtain a negative shape parameter ($c < 0$) when fitting the GEV distribution to the empirical data, indicating that maximum wind speeds are sampled from a Fréchet distribution. This distribution is characterized by a heavy tail, suggesting that very high wind speeds in tropical cyclones are rare but possible events in the North Atlantic. However, it is worth noting that maximum wind speed distributions from certain time periods appear not to fit well to the GEV distribution. This will be discussed further in the discussion section below.

In fitting the GEV distribution to minimum central pressure data, we flip the values by multiplying with -1 to capture the behavior of extreme values in the minimum or left tail of the distribution. The function we are using to fit, *scipy.stats.genextreme.fit(...)*, only models the *maxima* in extreme value distributions[6]. After transforming the data, we find a negative shape parameter for minimum central pressure, as shown in Figure 3.4, indicating that extreme values for this measurement of tropical cyclone intensity are also sampled from a Fréchet distribution. This suggests that, structurally, both measures of tropical cyclone intensity have similar statistical properties, where rare events that deviate significantly from typical values are possible.

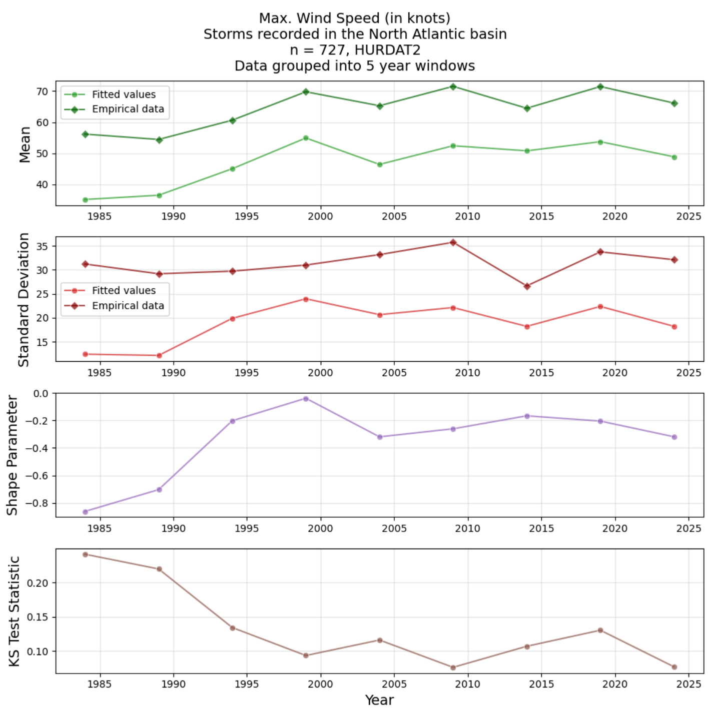

Figure 3.5

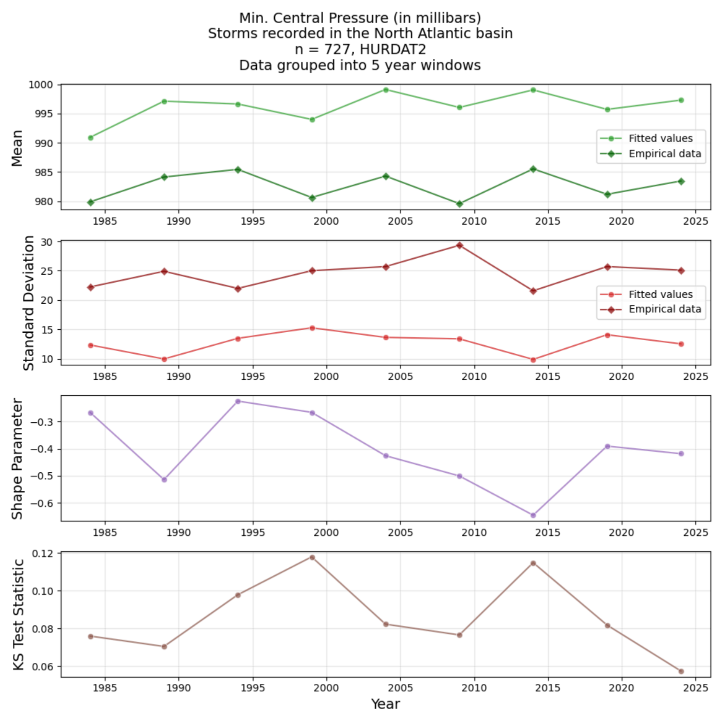

Figure 3.6

To understand the behavior of extreme values in the tropical cyclone data over time, we fit maximum wind speed and minimum central pressure from 5-year windows to the GEV distribution and plot the parameters for each subset of data. We also plot the Kolmogorov–Smirnov test statistic to quantify how good the observed data fits to the GEV distribution in each time period. For maximum wind speeds, we can see in Figure 3.5 that both the mean and standard deviation of the fitted GEV distributions are consistently higher than in the observed data, which suggests that the GEV distribution does not perfectly fit the empirical data. There does seem to be an increase in the mean maximum wind speeds over the past few decades, increasing from 56 knots in 1979-1984 to 70 knots in 1994-2004 for the empirical mean. The mean of maximum wind speed plateaus in more recent time periods, remaining consistently higher than 1979-1984 levels at values of 65 knots or larger. This increase indicates that intense tropical cyclones are indeed becoming more common when examining data from the past 45 years. It is worth noting that mean maximum wind speeds seem to be higher when examining tropical cyclone seasons further in the past. However, we believe that this is due to biases in data collection, which we explore in the discussion section.

There does not seem to be any meaningful trends in the standard deviation of maximum wind speeds, suggesting that the variability of extreme wind speeds has remained mostly the same. The magnitude of the shape parameter is higher in the past, implying that the distribution of maximum wind speeds had a heavier tail and that tropical cyclones with extreme wind speeds were more likely in the past. However, this could be attributed to the bad fit of the data to the GEV distribution, as indicated by the high Kolmogorov–Smirnov test statistic for the first two time periods which are both greater than 0.2. We believe this poor fit is likely due to an optimization error with the *scipy.stat.genextreme.fit(...)* method, we investigate this further in the discussion section. The shape parameters for the time windows examined are all negative, implying that the Fréchet distribution fits best to maximum wind speed data.

On the other hand, Figure 3.6 shows that for minimum central pressure, the mean of the fitted GEV distribution is consistently higher than the empirical mean, while the fitted standard deviation is consistently lower. Based on the empirical data, there does not appear to be any significant trends in minimum central pressure, with means oscillating between 980 and 985 millibars and standard deviation consistently being between 10 to 15 millibars. We do, however, see the shape parameter of the fitted GEV distribution decrease from the time period of 1989-1994 to 2009-2014, implying that the likelihood of tropical cyclones with extremely low central pressure grew during this time. While it's possible that this phenomenon is due to poor fit, the Kolmogorov–Smirnov test statistic is relatively low, remaining between 0.07 and 0.12, which implies that the data fit relatively well to the GEV distribution. Similar to maximum wind speed, all the shape parameters are negative, suggesting that minimum central pressure data from all time periods fit to the Fréchet distribution.

4 | Discussion

Our results confirm that the Generalized Extreme Value distribution can be used to describe empirical distributions of maximum wind speeds and minimum central pressure recorded from tropical cyclones in the North Atlantic. Analysis of the Kolmogorov–Smirnov (KS) test statistic proved that the GEV distribution can accurately characterize these two metrics for the majority of tropical cyclone seasons. In addition to the KS tests, Figures 3.3 and 3.4 visually illustrate that maximum wind speed data and minimum central pressure from the past 5 years fits well to the Fréchet distribution which is characterized by a heavy tail. This implies that extremely high wind speeds and low central pressure are rare but possible. Examining the empirical data and GEV parameters over time reveals that shape parameters are consistently negative, implying that extreme values for both maximum wind speed and minimum central pressure from any time period follow a Fréchet distribution. Furthermore, average maximum wind speeds have increased since the 1970s, which suggests that the warming climate may have increased the intensity of tropical cyclones in the North Atlantic. However, additional work is needed to establish a causal relationship between the two phenomena.

Some unanticipated findings include mean maximum wind speeds being higher for 5-year time windows before the 1970s as shown in Figure 8.3 in the appendix and the lack of clear trends in minimum central pressure. Higher mean maximum wind speeds in the past contradicts the theory that the consistent warming of the Earth observed in the last several decades increases tropical cyclone intensity. Our team believes that this discrepancy can be explained by the biased data collection in the HURDAT2 dataset. Before the 1970s, tropical cyclone data relied on direct observations which lacked the accuracy and consistency of modern satellite technology[7]. According to the *Tropycal* module's github page, "Satellite data was generally available before [1979] but on a more limited basis."[8] and documentation of the HURDAT2 dataset cites the late 1960s as the "pre-satellite era"[5]. The lack of satellite observations in the past may have caused weaker tropical cyclones to be systematically excluded, causing the intensity of tropical cyclones to appear as though it has decreased. Additionally, there does not seem to be any consistent trends in minimum central pressure over the last few decades, despite an upward trend in maximum wind speeds. We are not sure why this is; theoretically, lower central pressure should translate to more intense storms. To shed light on this discrepancy between theory and data, additional research is needed to examine the precise relationship between central pressure and wind speed.

Thus far, some shortcomings of the analysis include a poor fit for certain time periods like 1969-1973 and 1974-1978 as shown in Figure 8.1 and 8.2 in the appendix. While the aforementioned data quality issues with observations before the 1970s may have affected the fit of the data to the GEV distribution, we believe that the poor fit is primarily due to optimization issues with *scipy.stats.genextreme.fit()*. According to the *Scipy* documentation, the function performs Maximum Likelihood Estimation to fit the data to the GEV distribution, which could lead to convergence problems when initial parameter estimates provide a poor estimate of the data. Our team conducted several tests where we passed initial guesses for the parameters based on the empirical data. Unfortunately, these experiments did not yield any meaningful results. In future iterations of the library, we may explore the usage of different optimizers to better fit the GEV distribution to the data.

We are most interested in extending our current analysis by pairing the tropical cyclone data from NOAA with other climate datasets to predict tropical cyclone intensity based on predictors such as sea surface temperature, vertical wind shear, and relative air humidity. Then, we could use a climate model to examine how these predictors might change in the future, allowing us to predict tropical cyclone intensity and behavior based on different climate projections. We believe the main challenge to this approach is the need to harness large amounts of raster data to examine how predictors like sea surface temperature have changed in specific locations at specific times. To maximize the accuracy of our predictions, we hope to focus our analysis on specific regions, such as the Gulf of Mexico.

5 | Conclusion

Our findings suggest that Extreme Value Theory (EVT) can be used to accurately model both maximum wind speed and minimum central pressure of tropical cyclones, potentially enabling better forecasting of extreme events and enhancing disaster response strategies. We utilized tropical cyclone data from 1979 onward, measured via satellite, to minimize biases that are likely present in earlier data. The GEV distribution fits contemporary data relatively accurately both visually and statistically, as measured by the Kolmogorov–Smirnov test statistic, though it fails to fit certain periods due to optimization issues. In analyzing the behavior of the distribution over time, we find that 5-year averages of maximum wind speeds have increased over the past 50 years, suggesting that the changing climate may have contributed to increased tropical cyclone intensity. In contrast, minimum central pressure measurements remain relatively stable, and further research is needed to understand why. To extend our analysis, we plan to incorporate additional variables such as sea surface temperature to help predict tropical cyclone intensity. We hope this will allow us to provide further evidence of a causal relationship between climate change and extreme tropical cyclone intensity.

6 | References

1. “A force of nature: Hurricanes in a changing climate.” *NASA Science*. science.nasa.gov/earth/climate-change/a-force-of-nature-hurricanes-in-a-changing-climate. Accessed 15 Nov. 2024.

2. Sparks, Nathan, and Ralf Toumi. “The Dependence of Tropical Cyclone Pressure Tendency on Size.” *Geophysical Research Letters*, vol. 49, no. 19, Oct. 2022, https://doi.org/10.1029/2022gl098926.

3. Landsea, C. W. and J. L. Franklin. “Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format.” 2013, https://www.nhc.noaa.gov/data/hurdat/hurdat2-format-atl-1851-2021.pdf.

4. “Tropycal.” *Tropycal*, tropycal.github.io/tropycal/index.html. Accessed 14 Nov. 2024.

5. US EPA. “Climate Change Indicators: Sea Surface Temperature | US EPA.” *US EPA*, 18 Dec. 2016, www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature. Accessed 14 Nov. 2024.

6. “Scipy.Stats.Genextreme.” *Scipy.Stats.Genextreme - SciPy v1.14.1 Manual*, docs.scipy.org/doc/scipy/reference/generated/scipy.stats.genextreme.html. Accessed 14 Nov. 2024.

7. Hagen, Andrew B., et al. “A Reanalysis of the 1944–53 Atlantic Hurricane Seasons—The First Decade of Aircraft Reconnaissance*.” *Journal of Climate*, vol. 25, no. 13, Feb. 2012, pp. 4441–60. https://doi.org/10.1175/jcli-d-11-00419.1.

8. “Data Sources - tropycal.” *Tropycal,* https://tropycal.github.io/tropycal/data.html. Accessed 15 Nov. 2024.

7 | Attribution of Effort

Abigail Kinaro

- Project report: Abstract, introduction, various edits and improvements.
- Implementation of evaluation method to quantify differences in the empirical cumulative distribution function and fitted cumulative distribution function.

Kent Codding

- Project report: Discussion, summary/conclusion, various edits and improvements.
- Implementation of *pytest* test suite of the library: regression tests to ensure all method outputs are consistent, validation tests to ensure all method outputs are valid.

Max Bahar

- Project report: Model description, analysis, plots, various edits and improvements.
- Implementation of the fetch method to download data from HURDAT2/IBTrACS, fit method to fit the GEV distribution to the empirical data, score method to obtain Kolmogorov–Smirnov test statistics for the GEV distribution's fit, plotting methods to plot the distribution of the data and GEV parameters over time.
- Documented the library with type hinting, docstring, and examples in the readme.

8 | Appendix

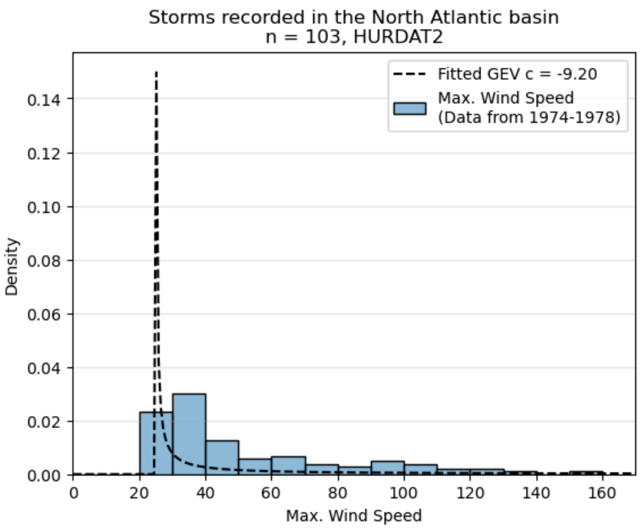

Figure 8.1

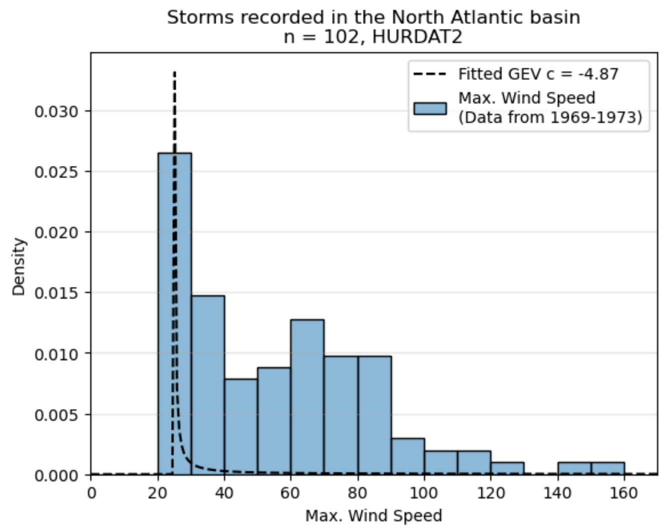

Figure 8.2

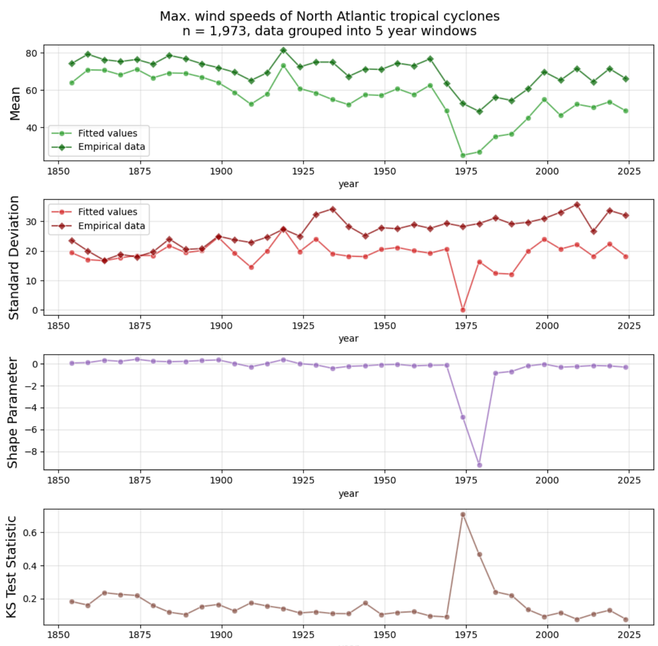

Figure 8.3

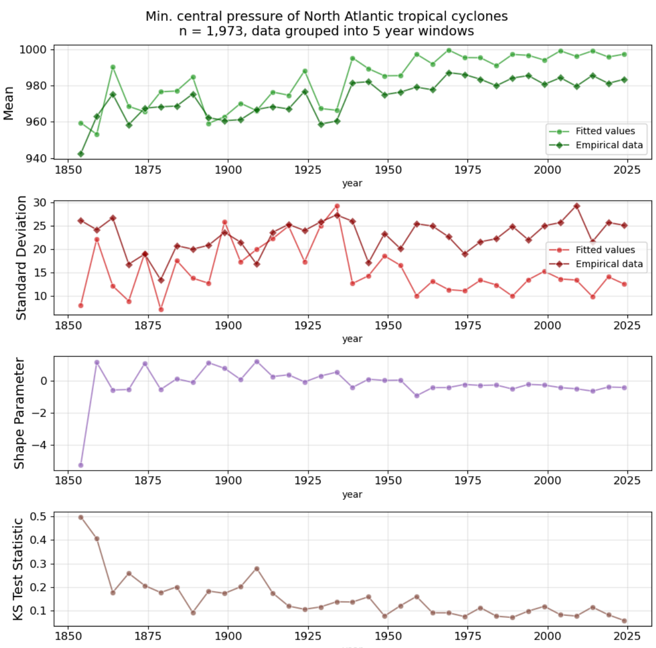

Figure 8.4