A Statistical Analysis of Atlantic Tropical Cyclone Frequency and Severity Trends Using Poisson and Extreme Value Models

Yoshi Moxey

2026-04-10

Contents

1. Introduction

1.1 Background

Tropical cyclones are among the most destructive natural hazards, producing extreme winds, storm surges and heavy rainfall that can result in substantial economic damages and loss of life. Understanding the possible long-term changes in tropical cyclone activity is therefore an important problem in environmental statistics, climate science, and risk modelling. In particular, the question of whether storms are becoming more frequent or intense over time has received increasing attention in the context of climate change and global warming.

Storm behaviour is typically analyzed using two separate components: frequency and severity. Frequency refers to the number of storm events occurring in a given time period, while severity refers to the intensity of storms, typically measured by maximum sustained wind speed or minimum central pressure. These two components are often modeled separately because they are governed by different statistical processes. Frequency is commonly modelled using count processes such as Poisson or negative binomial models, while severity is modelled using continuous distributions, often from Extreme Value Theory. Separating frequency and severity allows for clearer interpretation and reduces noise that may arise when attempting to model both components simultaneously.

Extreme Value Theory (EVT) provides a statistical framework for modelling rare and extreme events such as extreme wind speeds. One commonly used method is the Peaks-Over-Threshold (POT) approach, which models observations exceeding a high threshold using the Generalized Pareto Distribution (GPD). This approach is widely used in fields such as environmental and climate sciences to model extreme events and estimated return levels.

In addition to modelling extremes, it is important to determine whether the statistical properties of storms are stationary over time. A stationary process has statistical properties that do not change over time, while a non-stationary process exhibits trends, shifts or structural changes. Detecting non-stationary is particularly important for long-term risk assessment, climate impact studies and insurance modelling, because models assuming stationarity may underestimate future risk if storm behaviour is changing over time.

1.2 Motivation

There is ongoing scientific and public interest in whether tropical cyclones are becoming more frequent or more intense over time. Changes in storm frequency would suggest changes in atmospheric and oceanic conditions that influence storm formation, while changes in storm intensity could significantly increase economic losses

and infrastructure risk even if the number of storms remains constant.

From a statistical point of view, typical cyclone data provides an opportunity to apply several important modeling techniques, including Poisson regression for count data, seasonal modelling using sine and cosine terms, extreme value modelling for rare events, and non-stationary distribution modelling.

Another motivation for this study is to determine the structure of any detected changes in storm behaviour. Changes may occur gradually or over time, may occur as discrete shifts between early and late periods, or may occur at a specific changepoint. Understanding whether changes are gradual or structural is important for climate modelling, insurance risk estimation, and disaster preparedness.

This research study therefore aims to combine frequency modelling, seasonal analysis, and extreme value modelling into a uniform statistical analysis of storm activity over time.

2. Data and Preprocessing

2.1 Dataset Description

The dataset used for this analysis is sourced from the NOAA's International Best Track Archive for Climate Stewardship (IBTrACS) data for tropical cyclones in the North Atlantic basin. The data includes records of cyclones over a period of 50 years, from 1975 to 2024, with 808 unique weather events which were recorded across multiple instances of their life cycles, for a total of 22,705 observations. R studio was the programming language used to facilitate all analyses and generate all visualizations, and the libraries used include knitr, ggplot2, dplyr, sf, rnaturalearth, rnaturalearthdata, viridis, patchwork, evir, extRemes, ismev, stargazer and webshot2.

Dataset Description:

- Name (Categorical) – storm name,
- Year (Categorical) – date of report - year,
- Month (Categorical) – date of report - month numeric,
- Day (Categorical) – date of report - day,
- Hour (Categorical) – hour of report (UTC),
- Lat (Continuous) – location of storm center - latitude,
- Long (Continuous) – location of storm center - latitude,
- Status (Categorical) – storm classification (Levels: tropical storm, tropical depression, hurricane, etc),
- Category (Categorical) – Saffir-Simpson hurricane category calculated from wind speed
  - Levels:
    - NA: Not a hurricane
    - 1: 64+ knots
    - 2: 83+ knots
    - 3: 96+ knots
    - 4: 113+ knots
    - 5: 137+ knots,
- Wind (Continuous)[1] – Storm's maximum sustained wind speed (in knots),
- Pressure (Continuous) – Air pressure at the storm's center (in millibars).

[1] data was discretized into multiples of 5; jitter applied prior to severity analysis

2.2 Cleaning and Preprocessing

The dataset was mutated to include a column grouping cyclones into five year bins. From the original, cleaned dataset, a secondary dataset containing the observation with the highest recorded wind speed of each storm was generated.

Missing or placeholder values in the dataset were replaced with standard missing value notation (NA) prior to analysis. Each storm contains multiple observations recorded at varying time intervals. To analyze storm severity, observations were aggregated to obtain one representative value per storm. For each storm, the following summary statistics were computed:

- maximum recorded wind speed,
- minimum central pressure,
- mean latitude and longitude,
- month and year of occurrence of maximum wind speed.

Maximum wind speed was used as the primary measure of storm severity. Storm frequency analysis used unique storm identifiers to ensure that each storm was only counted once.

3. Methods

3.1 Exploratory Data Analysis

An Exploratory Data Analysis (EDA) was conducted to identify spatial, seasonal and temporal patterns in storm activity prior to formal modelling.

The following exploratory methods were used:

- Geographical plotting of storm paths using latitude and longitude,
- Storm origin location mapping,
- Histogram of storm counts per year,
- Monthly storm frequency histograms,
- Boxplots of storm wind speed to examine tail behaviour,
- Linear trend plots of storm severity over time.

This EDA provided preliminary evidence of strong seasonal frequency patterns and bounded-tailed severity distributions, motivating the use of Poisson regression for frequency and EVT for severity.

3.2 Storm Frequency Modelling

Storm frequency was modelled using Poisson regression models applied to annual and monthly storm counts. The modeling process followed a structured sequence to test for stationarity, seasonality, seasonal changes over time, and structural changes in storm frequency. Subsequent models were compared to the stationary model via Likelihood Ratio Test to evaluate their statistical significance.

3.2.1 Poisson Process Model for Annual Storm Counts

Let N_t denote the number of storms occurring in year t. Annual storm counts were assumed to follow a Poisson distribution:

$$N_t \sim \text{Poisson}(\lambda_t),\ E[N_t] = \lambda_t,$$

where λ_t is the expected number of storms in year t.

To test whether storm frequency is stationary over time, two models were considered:

Stationary model:

$$\lambda_t = \lambda.$$

Time trend model:

$$\log(\lambda_t) = \beta_0 + \beta_1 t.$$

This formulation corresponds to a Poisson regression model with a log link function.

An LRT was used to compare the stationary and time trend models. The likelihood ratio statistic is defined as

$$\text{LRT} = -2(l_0 - l_1),$$

where

- l_0 = log-likelihood of the stationary model.
- l_1 = log-likelihood of the time trend model.

Under the null hypothesis that the simpler model is sufficient, the LRT statistic follows approximately a chi-square distribution with degrees of freedom (df) equal to the difference in the number of parameters between models.

3.2.2 Seasonal Modelling of Storm Frequency

To model the seasonal variation in storm frequency, monthly storm counts were modelled using a Poisson regression model with sinusoidal terms

$$\log(\lambda_{t,m}) = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi m}{12}\right) + \beta_3 \cos\left(\frac{2\pi m}{12}\right).$$

This sinusoidal formation captures the periodic seasonal variation in storm frequency. Seasonality was tested by comparing the stationary annual model with the seasonal monthly model.

3.2.3 Stability of Seasonal Pattern Over Time

To determine whether the seasonal pattern of storm occurrences changed over time, interaction terms between time and the seasonal sine and cosine terms were added:

$$\log(\lambda_{t,m}) = \beta_0 + \beta_1 t + \beta_2 \sin(\frac{2\pi m}{12}) + \beta_3 \cos(\frac{2\pi m}{12}) + \beta_4 t \sin(\frac{2\pi m}{12}) + \beta_5 t \cos(\frac{2\pi m}{12}).$$

This model was then compared to the model without interaction terms to determine whether there was a meaningful seasonal pattern change over time.

3.2.4 Shoulder Season Expansion Test

The Atlantic Hurricane Season spans the period from June 1 to November 30 each year. For the purposes of this study, months were grouped into:

- Peak Season (August, September, October),
- Shoulder Season (June, November),
- Off Season. A Poisson regression model with interaction between period (early vs late) and season type was fitted:

$$\log(\lambda_t) = \beta_0 + \beta_1 \text{Period} + \beta_2 \text{Season type} + \beta_3 \text{Period} \times \text{Season type}.$$

A significant interaction would indicate that shoulder season storm frequency has changed at a different rate compared to peak season.

3.2.5 Changepoint Analysis for Frequency

To test for structural changes in storm frequency, a changepoint model was fitted. The model allows for storm frequency to change after an unknown changepoint year τ:

$$\log(\lambda_t) = \beta_0 + \beta_1 I(t > \tau),$$

where $I(t > \tau)$ is an indicator function equal to 1 if the year is after the changepoint and 0 otherwise.

Models with different candidate changepoint years were fitted and compared using log-likelihood, AIC, and BIC. This approach determines whether storm frequency experienced a structural shift rather than a gradual shift.

3.3 Extreme Value Modelling of Storm Severity

Storm severity was modelled using EVT with maximum record storm wind speed as a proxy, according to the metric of the Saffir-Simpson Hurricane Wind Scale classification. Rather than modelling the entire wind speed distribution, EVT focuses on the behaviour of extreme observations; therefore, the POT approach, which models observations exceeding a prespecified threshold, was implemented.

Let X denote the maximum recorded wind speed and u denote the chosen threshold. Exceedance were defined as:

$$Y = X - u \text{ , given } X > u.$$

For a high enough threshold, u, exceedances approximately follow a generalized Pareto distribution with cumulative distribution function:

$$F(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}},$$

for $y > 0$, where

- $\sigma > 0$ is the scale parameter.
- ξ is the shape parameter.

The shape parameter determines the tail behaviour of the distribution, and how quickly the probabilities decrease for extreme events.

- $\xi > 0$: Indicates a heavy tailed distribution.
- $\xi \approx 0$: Indicates an exponential/light tail.
- $\xi < 0$: Indicates a bounded tail with a finite upper bound, where extreme values cannot exceed a certain threshold.

3.3.1 Threshold Selection

Selecting an appropriate threshold is critical in POT modelling. Threshold selection was guided both by domain knowledge where storms exceeding 64 knots are classified as hurricanes, and by the following diagnostic methods:

- Mean Residual Life (MRL) Plot,
- Parameter Stability Plot,

- Stability of implied upper bound[2] (thresholds surrounding candidate produce relatively constant].

The implied upper bound for a bounded tail GPD is

$$x_{max} = u - \frac{\sigma}{\xi}.$$

If the threshold is appropriate, this implied upper bound should remain relatively stable as the threshold changes. Large fluctuations in the implied upper bound suggests that the threshold selected is too low.

A threshold was selected where:

- The MRL plot appeared approximately linear,
- Parameter estimates were relatively stable,
- The implied upper bound was stable.

3.3.2 Stationary GPD Model

Under the stationary model, exceedances were assumed to follow a GPD with constant parameters:

$$Y \sim \text{GPD}(\sigma, \xi).$$

Parameters were estimated using:

- Maximum Likelihood Estimation (MLE).
- Method of Moments (MoM).

3.3.3 Goodness-Of-Fit Assessment

Several diagnostic methods were used to evaluate the adequacy of the GPD model Graphical diagnostics included:

Model fit was assessed using both graphical and numerical models:

Graphical diagnostics:

- Quantile-quantile (QQ) plots,
- Probability-probability (PP) plots,
- Histogram of Probability Integral Transform (PIT) values.

[2] candidate threshold and values surrounding

Numerical goodness-of-fit:

- Anderson-Darling (AD) statistic,
- Bootstrap p-values for AD statistic.

Because wind speeds were recorded in discretized units, small random noise (jitter) was added before the goodness-of-fit testing to better approximate the continuous distribution assumption required for EVT modelling.

Bootstrap p-values were calculated by simulating exceedances from the fitted GPD and comparing simulated AD statistics to the observed AD statistic.

3.3.4 Early vs Late Comparison

To test for structural differences in storm severity, the dataset was divided into early and late periods based on the median year. Separate GPD models were fitted to each period, and parameter estimates were compared. Goodness-of-fit tests were performed for each period to verify that the GPD model remained appropriate for each period.

3.3.5 Continuous Non-Stationary GPD Models

To test whether storm severity changed over time, non-stationarity GPD models were fitted where the parameters depend on time. Three models were considered:

Model 1: Scale varies with time

$$\log(\sigma_t) = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi m}{12}\right) + \beta_3 \cos\left(\frac{2\pi m}{12}\right),$$

Model 2: Shape varies with time[3]

$$\xi_t = \beta_0 + \beta_1 t,$$

Model 3: Both scale and shape vary with time:

$$\log(\sigma_t) = \beta_0 + \beta_1 t,$$

[3]Modelling a time-varying shape parameter can be unstable, as the shape parameter governs the distribution's tail behaviour and is highly sensitive to a limited number of extreme observations. Small changes in ξ, particularly when negative, can lead to large fluctuations in the implied upper bound and reduce interpretability.

$$\xi_t = \beta_0 + \beta_1 t.$$

The models were compared to the stationary model using:

- LRT,
- AIC,
- BIC.

3.3.6 Changepoint Model

To determine whether storm severity experienced a structural shift rather than a gradual change, a changepoint model was fitted.

The scale parameter was allowed to change after a changepoint year τ:

$$\log(\sigma_t) = \beta_0 + \beta_1 I(t > \tau),$$

where $I(t > \tau)$ is an indicator function equal to 1 if the year is after the changepoint and 0 otherwise.

Models with different candidate changepoints were fitted and compared using likelihood, AIC and BIC to determine whether a structural shift occurred.

4. Results

4.1 Exploratory Results

4.1.1 Storm Tracks and Spatial Distribution

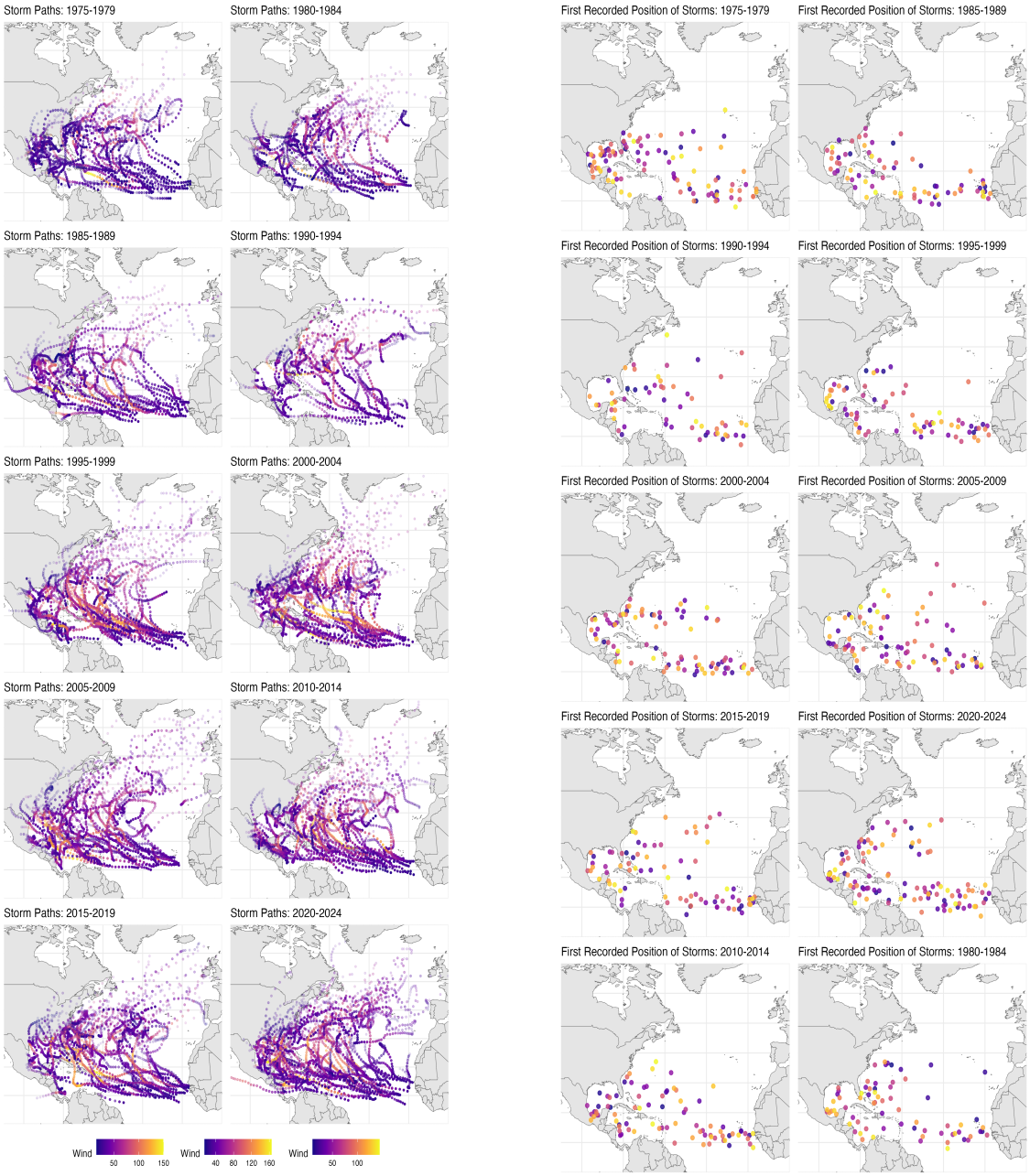

Figure 1: Spatial distribution of Atlantic tropical cyclones. The left panel shows storm tracks faceted by five-year periods with colour indicating wind speed at each observation, while the right panel displays the first recorded position of each storm, representing its origin.

Figure 1 illustrates the spatial distribution of Atlantic tropical cyclones. The left panel presents storm tracks faceted by five-year periods, with colour indicating wind speed at each observation. The right panel displays the origin of each storm, defined as its first recorded position.

Storms predominantly form in the tropical Atlantic and generally track westward before recurving northward. Visual inspection across time does not reveal any substantial spatial shifts in storm origins or trajectories. The overall spatial pattern is further confirmed in *Figure A1*, which displays all storm tracks without faceting.

4.1.2 Storm Frequency Exploration

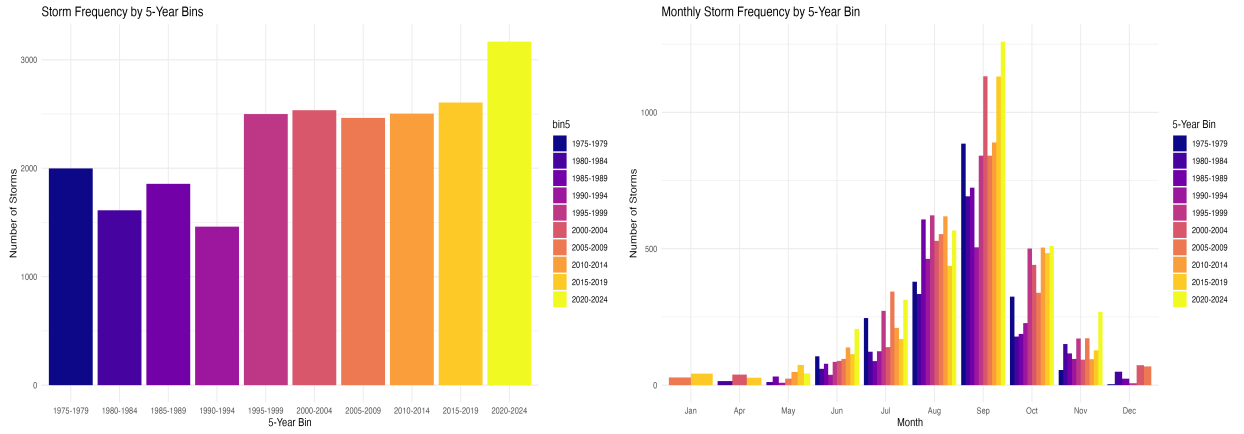

Figure 2 {left}: Total number of storms within each five-year period, illustrating interannual variability in storm frequency across the study period.

Figure 3 {right}: Monthly storm counts by five-year period, highlighting the strong seasonal concentration of activity during the peak hurricane season (August–October).

Exploratory analysis of storm counts reveals pronounced storm behaviour. *Figure 2* shows the number of storms within each five-year period, indicating interannual variability, but no clear monotonic trend over time. *Figure 3* displays monthly storm counts by five-year bin, highlighting a strong concentration of activity between August and October, and minimal activity outside the hurricane season.

These findings suggest that while annual storm frequency may be approximately stationary, storm occurrence is strongly seasonal.

4.1.3 Storm Intensity Exploration

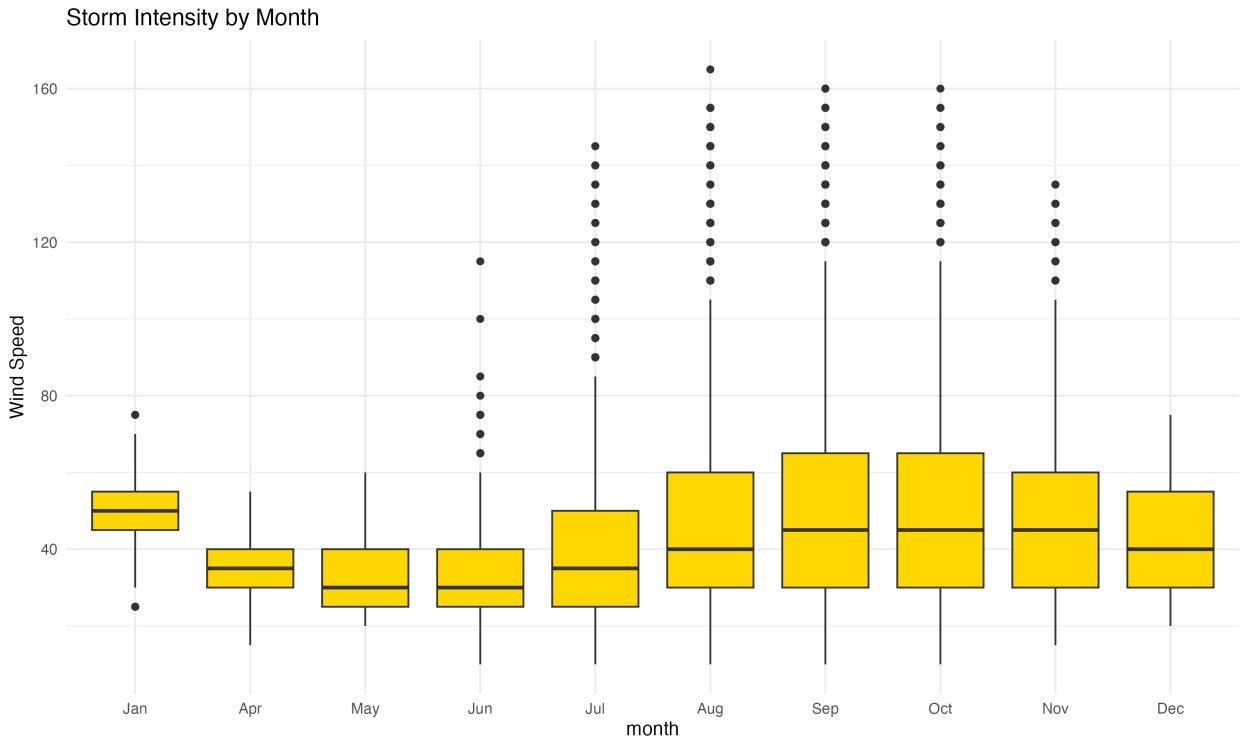

Figure 4: Boxplots of maximum wind speeds by month, demonstrating seasonal variation in storm intensity with higher median and extreme values during peak hurricane season months.

Seasonal variability in storm intensity is evident in *Figure 4*, which presents monthly distributions of maximum wind speeds. Higher median and upper quantile wind speeds occur during peak hurricane season months.

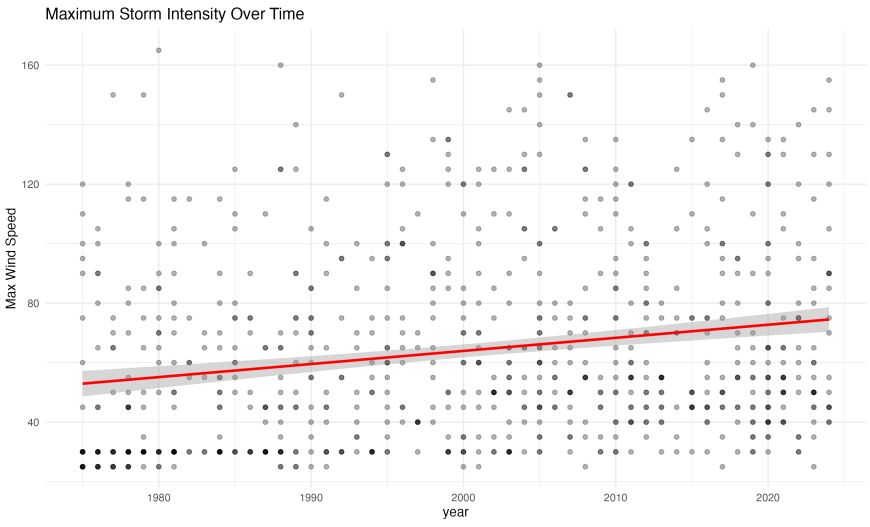

Figure 5: Maximum wind speed per storm over time with a fitted linear trend, providing a preliminary visual assessment of potential changes in storm intensity.

A preliminary assessment of long-term changes in intensity is provided in *Figure 5*, which plots maximum wind speed per storm over time with a fitted linear trend. Although variability is substantial, the trend suggests a modest increase in intensity.

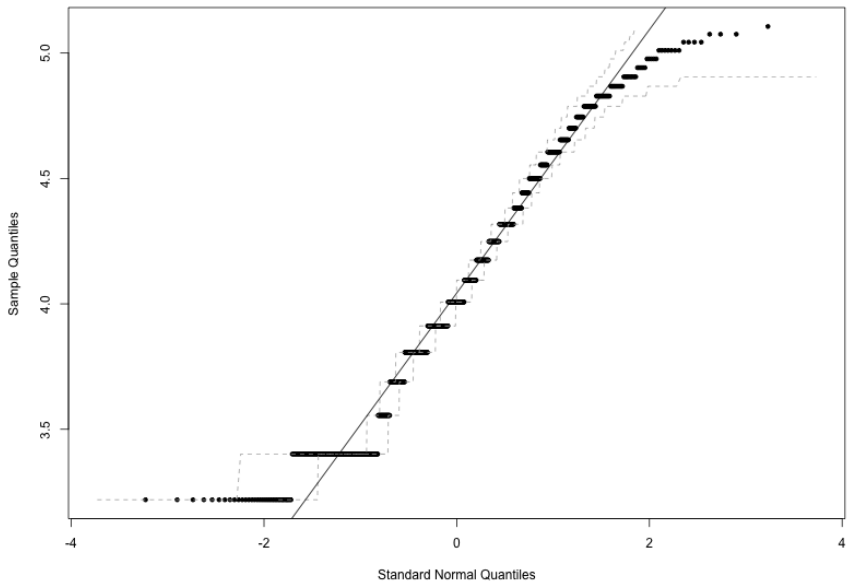

Figure 6: Normal QQ plot of log-transformed maximum wind speeds, showing deviations in the upper tail and supporting the use of Extreme Value Theory for modelling extremes.

The deviation from normality in the upper tail of the log-transformed, discretized wind speeds, shown in *Figure 6*, supports the use of EVT for modelling extreme storm intensities. Additional seasonal variability in intensity across five-year bins is illustrated in *Figure A3*.

4.2 Storm Frequency Analysis

4.2.1 Annual Storm Counts – Stationarity Test

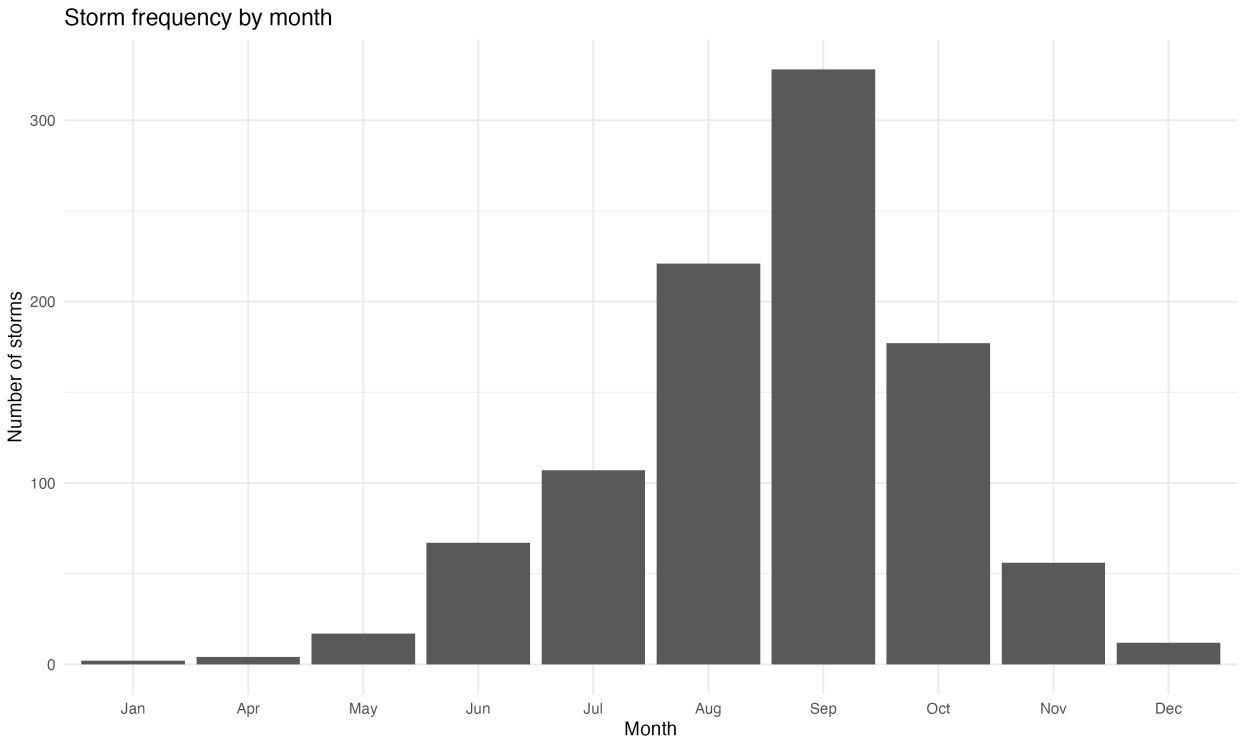

Figure 7: Aggregate monthly storm frequency across the entire study period, confirming the pronounced seasonal cycle in storm occurrence.

Storm counts vary strongly by month but do not show a clear increasing trend over time. To formally test whether annual storm counts changed over time, a Poisson regression model with a linear time trend was compared to a stationary Poisson model.

```
Analysis of Deviance Table

Model 1: N ~ 1
Model 2: N ~ year
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        49     76.224
2        48     74.406  1   1.8178   0.1776
```

Figure 8: Analysis of deviance comparing the stationary Poisson model with a Poisson model including a linear time trend for annual storm counts.

The results of the likelihood ratio test shown in *Figure 8* show a p-value of approximately 0.1776, indicating that including year as a predictor does not significantly improve model fit. Therefore, there is no statistical evidence of a long-term trend in annual storm counts, and storm frequency can be considered stationary on an annual scale.

This result suggests that, on an annual scale, storm frequency can reasonably be approximated as stationary.

4.2.2 Seasonal Variation in Storm Frequency

Strong seasonal variation in storm occurrences was observed in the exploratory analysis, evidenced by *Figure 3*. To quantify this seasonal effect, a Poisson regression model incorporating sinusoidal terms was fitted:

$$\log(\lambda_{t,m}) = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi m}{12}\right) + \beta_3 \cos\left(\frac{2\pi m}{12}\right).$$

The seasonal model produced a highly significant improvement in model fit compared to the non-seasonal model (p-value $< 2.2 \times 10^{-16}$), providing strong evidence of seasonal variation in storm frequency. This confirms that storm occurrence is strongly concentrated during the Hurricane Season and minimal during off-season months.

4.2.3 Stability of Seasonal Patterns Over Time

To determine whether the seasonal distribution of storms has changed over time, interaction terms between time and seasonal sine and cosine terms were added.

The interaction terms were not statistically significant (p-value = 0.8916), indicating that the shape of the seasonal cycle has remained stable over time.

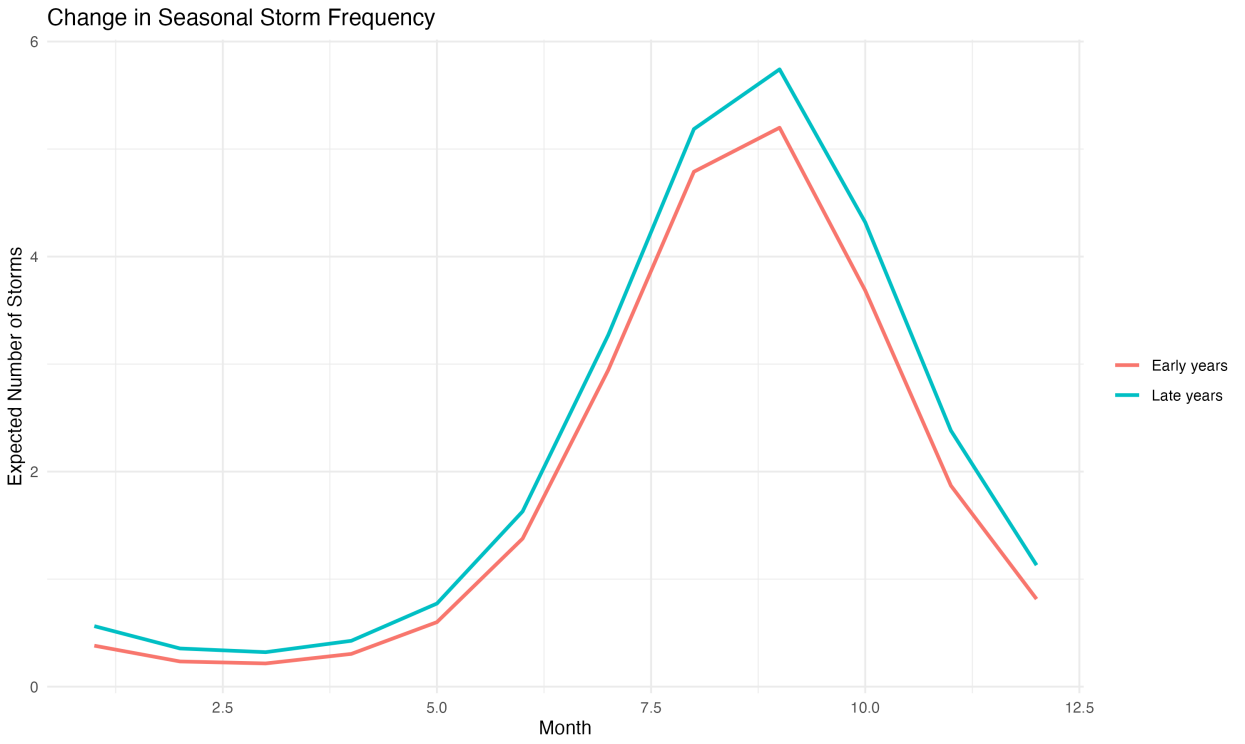

Figure 9: Estimated seasonal storm frequency for early and late periods based on the interaction model, demonstrating the stability of the seasonal pattern over time.

Figure 9 shows that the seasonal pattern is very similar between early and late periods with storm activity remaining concentrated in the same months. This supports the conclusion that while storms are strongly seasonal, the timing and peak activity has not shifted significantly. This is further corroborated by *Figure A3* which shows the monthly distribution of storm occurrences by five year periods.

4.2.4 Analysis of Hurricane Shoulder Season

To investigate whether the hurricane season expanded into "shoulder months", months were classified as:

- Peak season: August - October,
- Shoulder season: June and November,

- Off-season: Remaining months outside of the Hurricane Season.

While the overall seasonal structure remains stable, a key question is whether the relative contributions of "shoulder months", the months at the very beginning and end of the season which experience few storms, has changed significantly over time, which would indicate a potential expansion of the Hurricane Season.

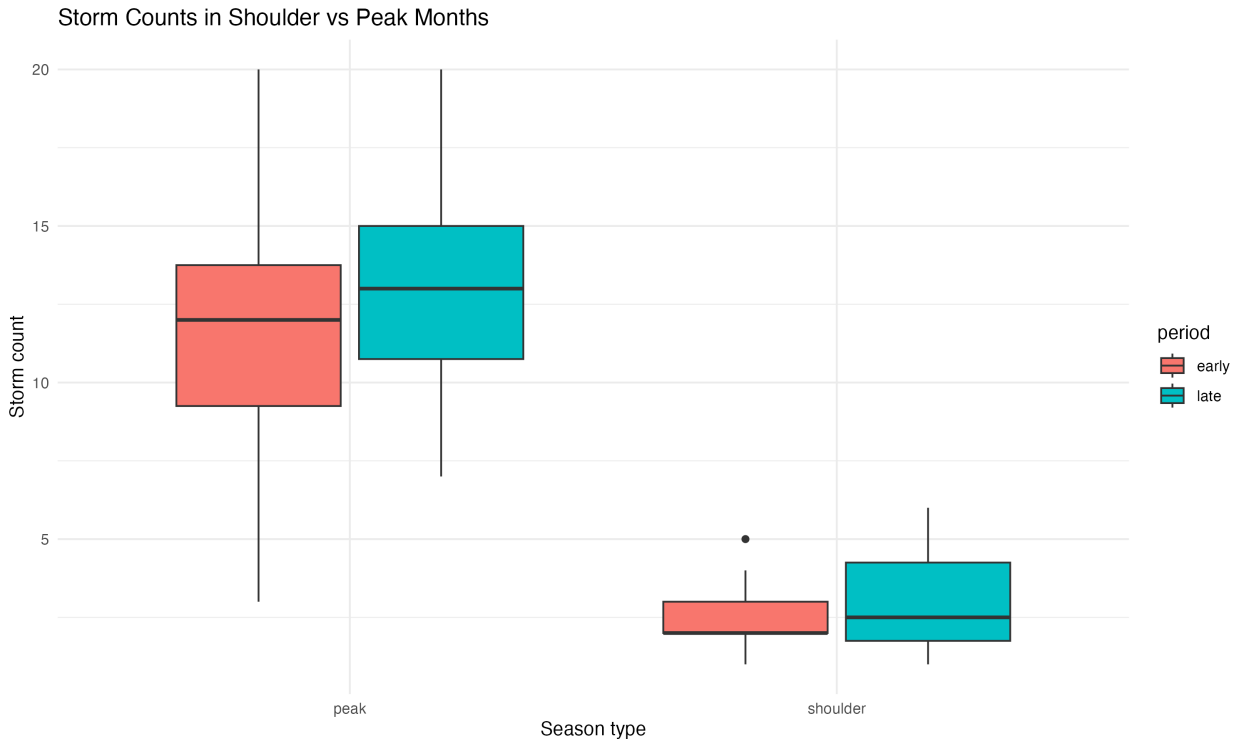

Figure 10: Distribution of storm counts in peak (August–October) and shoulder (June and November) months across early and late periods, used to assess potential expansion of the hurricane season.

Through visual inspection of *Figures 10*, the distribution of storms throughout the year appears to remain roughly consistent.

A Poisson regression model including an interaction between period (early vs late) and season type was fitted to test whether shoulder-season storm activity increased disproportionately over time.

The interaction term was not statistically significant (p-value = 0.842), indicating no evidence that shoulder-season storm activity increased relative to peak-season activity. While peak months consistently have significantly higher storm counts, the relative distribution between peak and shoulder months has remained stable over time.

This suggests that there is no statistical evidence that the hurricane season has expanded into shoulder months.

4.2.5 Structural Change in Storm Frequency via Changepoint

To investigate whether Atlantic storm frequency experienced a structural shift, a Poisson changepoint model of the form

$$\log(\lambda_t) = \beta_0 + \beta_1 I(t > \tau).$$

was fitted, where τ represents an unknown changepoint year, and fit was evaluated by comparing log-likelihood, AIC and BIC.

The optimal changepoint was identified to be the year 2018, corresponding to the minimum AIC and BIC values shown in *Table A1.* A likelihood ratio test comparing this model with a homogeneous Poisson model indicated that the inclusion of the changepoint significantly improved model fit (p-value = 0.008792.

These results provide evidence of a structural shift late in the 50 year period of study, rather than a gradual change over time.

4.3 Storm Severity Analysis

4.3.1 Threshold Selection and Stability

Storm severity was modelled using the POT approach from EVT, in which observations exceeding a sufficiently high threshold are modelled using the GPD. The threshold was initially selected as 64 knots, corresponding to the minimum wind speed required for a storm to be classified as a hurricane under the Saffir-Simpson Hurricane Wind Scale, thereby defining hurricane-strength storms as extreme events.

To validate this threshold, candidate thresholds between 60 and 70 knots were evaluated using parameter stability plots and Mean Residual Life plots. Across this range, the estimated scale parameter remained approximately between 47.0 and 48.6, and the shape parameter remained between approximately -0.42 and -0.48. The implied upper bounds are also relatively consistent across thresholds.

Table 1: Estimated GPD scale and shape parameters and corresponding implied upper bounds for candidate thresholds between 60 and 70 knots, used to assess threshold stability.

	threshold	scale	shape	upper_bound
1	60	47.53986	-0.4188319	173.5058
2	62	47.61170	-0.4316394	172.3043
3	64	47.89285	-0.4476366	170.9905
4	66	48.60103	-0.4696357	169.4867
5	68	47.03460	-0.4609579	170.0367
6	70	47.52648	-0.4807445	168.8602

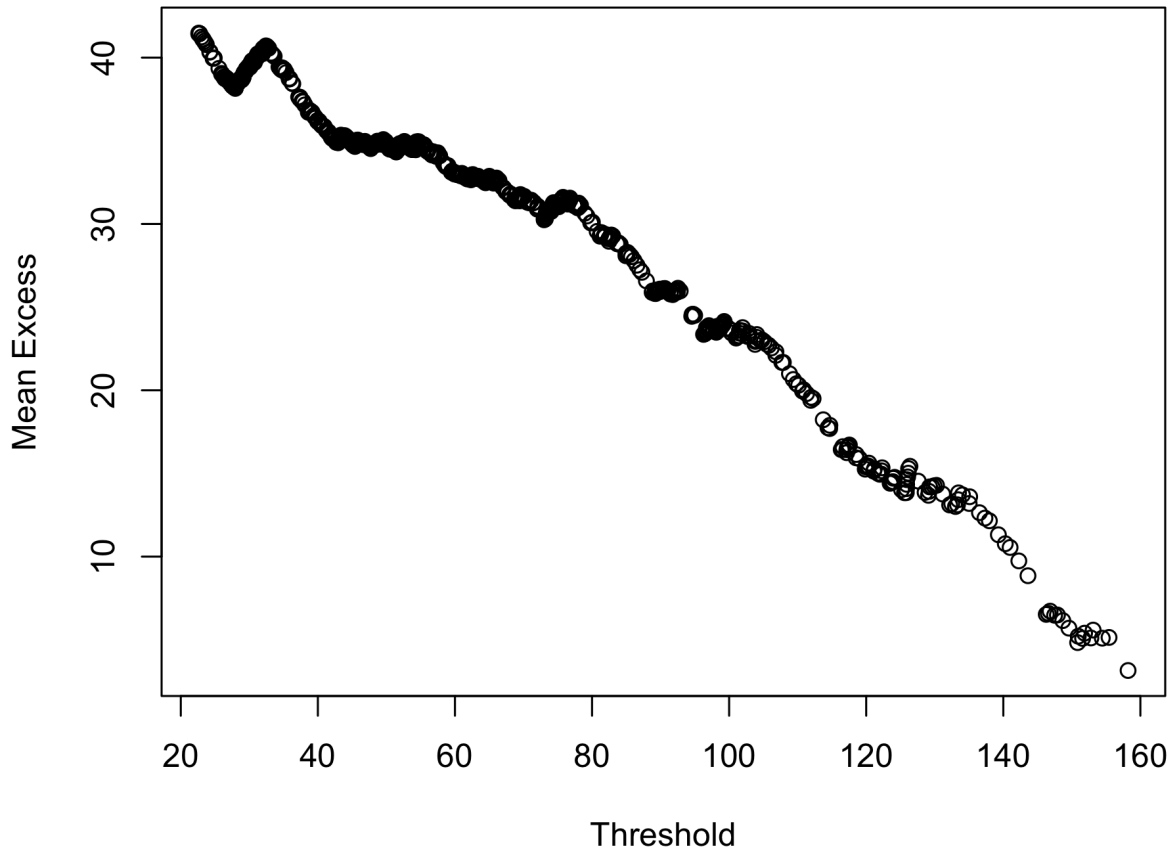

Figure 11: Mean residual life plot for wind speed exceedances, with approximate linearity above the selected threshold of 64 knots supporting the suitability of the threshold for GPD modelling.

The Mean Residual Life plot exhibited an approximately linear trend for values above 64 knots, supporting the assumptions that exceedances above this threshold follow a GPD. The stability of parameter estimates across candidate thresholds indicate that the selected threshold lies within a region where the GPD approximation is valid. Therefore, a threshold of 64 knots was selected for all subsequent analyses.

4.3.2 Stationary GPD Model

Using the selected threshold, a stationary GPD model was fitted to all exceedances of the selected threshold of 64 knots. Parameter estimation was performed using both Maximum Likelihood Estimation (MLE) and Method of Moments (MoM) for comparison.

Maximum Likelihood Estimation yielded:

- Scale parameter: $\sigma \approx 47.89$.
- Shape parameter: $\xi \approx -0.45$.

Method of Moments Estimation yielded:

- Scale parameter: $\sigma \approx 43.48$.
- Shape parameter: $\xi \approx -0.34$.

The shape parameter estimate from MLE is close to -0.5, which is near the boundary where maximum likelihood estimation can become unstable. This is one motivation for also using Method of Moments estimation as a comparison.

The negative shape parameter indicates a bounded upper tail, implying that the distribution of extreme wind speeds has a finite upper bound under the fitted model. This result is consistent with the physical constraints on storm intensity such as ocean temperature limits, and atmospheric conditions.

The bounded tail does not imply that extremely intense storms cannot occur, but that the probability of arbitrarily large wind speeds approaches zero beyond a theoretical maximum value.

4.3.3 Goodness of Fit Assessment

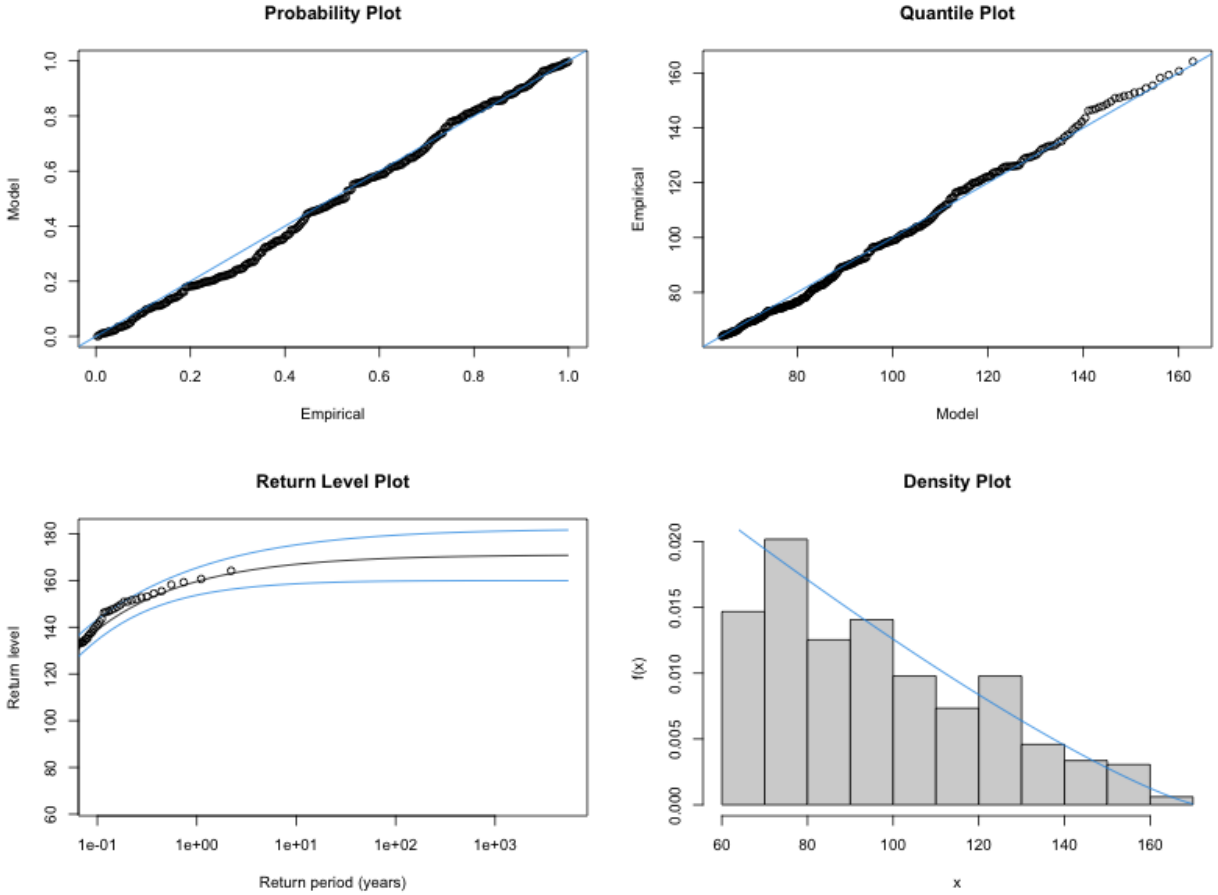

Figure 12: Diagnostic plots for the fitted stationary GPD model, including QQ and return level plots, indicating close agreement between empirical and theoretical distributions.

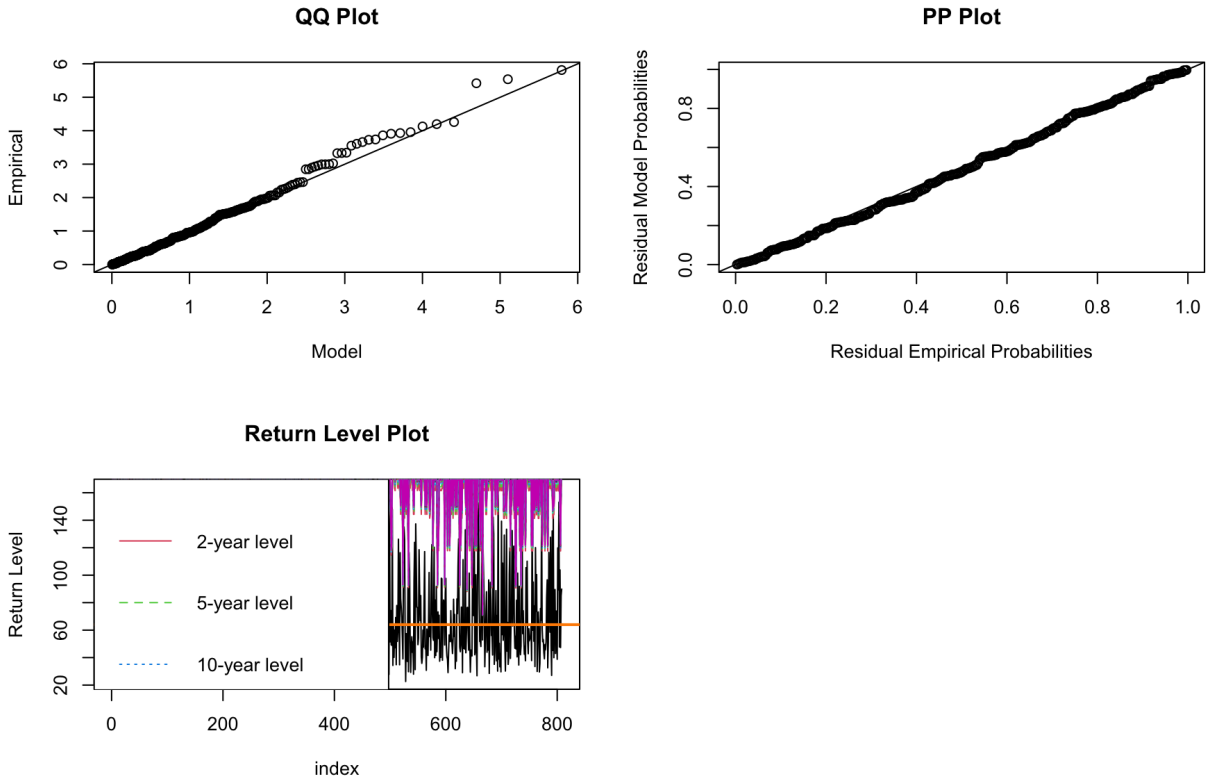

Figure 13: QQ plot, PP plot, and return level plot for the fitted GPD model, further supporting the adequacy of the model fit to exceedance data.

Goodness-of-Fit for the GPD model was evaluated using graphical diagnostics including QQ plots, PP plots, and Probability Integral Transform (PIT) histograms, as well as bootstrap p-values calculated using the Anderson-Darling statistic shown in *Table A2*, which is particularly sensitive to discrepancies in the tail of the distribution.

Wind speeds were discretized in increments of five knots in the dataset; therefore, small random noise was added to the exceedances before goodness-of-fit testing to satisfy the continuous distribution assumption required for EVT diagnostics and the Anderson-Darling test.

Table 2: Bootstrap Anderson–Darling p-values comparing the goodness-of-fit of GPD models estimated using Maximum Likelihood and Method of Moments.

Anderson-Darling Bootstrap p-values

	Period	MLE	MoM
1	Full	0.248	0.600
2	Early	0.477	0.617
3	Late	0.695	0.858

The bootstrap p-values in *Table 2* are all greater than 0.05, providing no evidence against the GPD model and indicating that the GPD provides an adequate representation of the exceedance distribution. The Method of Moments model produced larger p-values in every instance, suggesting a better fit. Overall, these diagnostics support the use of the GPD model for storm severity exceedance.

Table 3: Estimated GPD parameters and 95% confidence intervals for early and late periods, used to assess structural differences in storm severity.

Early/Late Parameter Confidence Intervals

	lower	upper
Early_Shape	31.574	47.261
Early_Scale	-0.472	-0.195
Late_Shape	47.688	69.188
Late_Scale	-0.723	-0.453

	MLE	MoM
Scale	-0.4475272	-0.3353933
Shape	47.8854600	43.4752483

Figure 14: MLE and MoM parameter estimates

The visual diagnostics for the Goodness-of-Fit of the MLE and MoM estimates in *Figure A5* indicate that GPD is an appropriate model.

4.3.4 Parameter Comparison Between Early and Late Periods

To investigate potential changes in storm severity over time, the dataset was divided into early and late periods based on the median year. Separate stationary GPD models were fitted to each period.

Estimated parameters were approximately:

- Early period:
  - Scale (σ) ≈ 41.27.
  - Shape (ξ) ≈ -0.24.
- Late period:
  - Scale (σ) ≈ 57.19.
  - Shape (ξ) ≈ -0.57.

The later period exhibits a substantially larger scale parameter, indicating greater variability and spread in extreme wind speeds. The strongly negative shape parameter in the later period indicated a more strongly bounded upper tail, and possible instability in the estimate. Confidence intervals for the scale parameters showed slight overlap, and a formal hypothesis test comparing parameter differences produced a borderline p-value (≈ 0.067), suggesting moderate but not strong evidence of a change between periods.

The visual diagnostics for the Goodness-of-Fit of the early and late MLE and MoM estimates in *Figure A6 and A7* indicate that GPD is an appropriate model.

4.3.5 Continuous Non-Stationary GPD Models

To formally test whether storm severity exhibits gradual changes over time, non-stationary GPD models were fitted in which the scale or shape parameters alone, or both parameters together were allowed to depend on time and seasonal covariates.

Three models were compared to the stationary model:

1 - Scale parameter varies with time 2 - Shape parameter varies with time 3 - Both scale and shape parameters vary with time

The time-varying scale model showed a significantly improved fit with a p-value ≈ 0.008. The time-varying shape model resulted in a not significant p-value (≈ 0.60). The both-varying model also exhibited a slight improvement in model fit (p-value = ≈ 0.028)

Model comparison using AIC favoured the model with a time-varying scale parameter, while BIC slightly favoured the stationary model due to its lower complexity.

These results indicate that storm severity non-stationarity in storm severity is primarily driven by changes in the scale parameter rather than the shape parameter.

Because the upper bound depends on both σ and ξ, changes in scale do not necessarily imply changes in the theoretical maximum intensity, but rather changes in the spread and variability of extreme wind speeds.

4.3.6 Changepoint Models

4.3.6.1 Early vs Late Period A second set of non-stationary models was fitted in which the scale and shape parameters were allowed to differ between early and late periods using a binary period indicator.

Likelihood ratio tests showed:

- The model with scale varying by period: p-value ≈ 0.26.
- The model with shape varying by period: p-value ≈ 0.81.
- The model with both varying by period: p-value ≈ 0.067.

Allowing both parameters to vary between periods produced the lowest AIC value among candidate models, suggesting that the early and late period distributions differ structurally, although the hypothesis test was only marginally significant.

These results suggest some structural differences between early and late storm severity distributions, but the evidence is weaker than in the continuous time-trend models.

4.3.6.2 Optimal Changepoint Year To assess whether the intensity of extreme storms has structurally changed over time, a changepoint model was incorporated into the GPD framework for threshold exceedances. The scale parameter was allowed to vary according to

$$\log(\sigma_t) = \beta_0 + \beta_1 I(t > \tau),$$

where τ is the year 2002, which, of all candidate changepoint years compared, produced the lowest AIC and BIC values, s shown in *Table A2.*. Although the optimal changepoint year of 2002 minimized AIC and BIC, the likelihood ratio test did not indicate a statistically significant improvement over the stationary model (p = 0.1855). Therefore, there is insufficient evidence to conclude that storm severity experienced a distinct structural shift at a single point in time.

5. Conclusion

This study examined long-term changes in Atlantic tropical cyclone activity using two complementary components: storm frequency and storm severity. Frequency was analyzed using Poisson regression models and seasonal analysis, while severity was modelled using the POT approach with GPD. Together, these analyses provide a comprehensive picture of how storm behaviour has evolved over the period 1975 - 2024.

Frequency

The analysis of storm frequency showed that the total number of storms per year does not exhibit a statistically significant long-term linear trend. A Poisson regression model comparing a constant-rate model with a model including year as a predictor produced a non-significant likelihood ratio test, indicating that annual storm counts can reasonably be modelled as stationary over time.

When storm counts were analyzed at the monthly level, a pronounced and persistent seasonal pattern was detected. The inclusion of sinusoidal seasonal terms significantly improved the Poisson regression model, confirming that storm frequency varies systematically throughout the year. This result is expected and reflects the well-known Atlantic Hurricane Season annual cycle, with the majority of storms occurring between August and October.

To determine whether the timing and relative distribution of storm frequency across months has been changing over time, an interaction model between year and seasonal terms was fitted. The interaction terms were not statistically significant, indicating that the seasonal pattern has remained stable across the testing period. In other words, while storms occur more frequently during peak hurricane season month, the shape and timing of the seasonal distribution have not significantly changed.

To perform a changepoint analysis, the optimal changepoint year was calculated to be 2018, and the subsequent likelihood ratio test confirmed that the shift was significant. This indicates a late stage structural shift in the storm frequency, rather than a gradual shift.

Because the estimated changepoint occurs near the end of the study period, the result should be interpreted cautiously, as limited post-changepoint observations may influence the stability of the estimate.

Overall, the frequency analysis indicates that annual storm counts are approximately stationary, with strong seasonal variation but no significant long-term trend or seasonal expansion. Annual storm counts do not exhibit a significant linear trend, but changepoint analysis suggests a possible late-period structural shift.

Severity

Storm severity, measured by maximum sustained wind speed, was modelled with the POT approach with a threshold of 64 knots, corresponding to the minimum hurricane-strength wind speed. The GPD provided an adequate fit to exceedances, and threshold stability analysis supported the chosen threshold.

The estimated shape parameter was negative, indicating a bounded upper tail for extreme wind speed, suggesting that there is a theoretical upper limit to storm intensity under the fitted model. This is consistent with physical constraints such as ocean temperature limits and atmospheric conditions.

When the data was divided into early and late periods, differences in parameter estimates were observed. Continuous non-stationary models indicated that the scale parameter changes over time, while the shape parameter remained relatively stable. Early versus late period models suggested structural differences between periods, although individual parameters were only moderately significant. Continuous non-stationary models provided the strongest evidence of non-stationarity, particularly through changes in the scale parameter, while changepoint analysis did not reveal a statistically significant single structural shift.

These results suggest that the main change in storm severity is not necessarily an increase in the theoretical maximum intensity, but rather changes in the variability and distribution of extreme wind speeds. Because the upper bound of the GPD depends on both scale and shape parameters, changes in scale can still affect extreme quantiles and return levels even if the shape parameter remains relatively stable.

Overall

The results of this study suggest that Atlantic tropical cyclones are not becoming significantly more frequent, but the statistical distribution of extreme storm intensities show evidence of non-stationarity, particularly in the scale parameter of the extreme value distribution. This indicates that the variability and distribution of extreme storm intensities may be changing over time.

Together, these results suggest that while storms are not necessarily becoming more frequent, the behavior of extreme intensities may be changing over time, with greater variability and evidence of a structural shift in the distribution of extreme wind speeds.

Implications and Limitations

From a statistical perspective, this study demonstrates the importance of separating frequency and severity modelling when analyzing environmental extremes. Even when event frequency appears stationary, the severity distribution may still exhibit non-stationary behaviour, which has implications for risk estimation, return levels and infrastructure planning.

The results also illustrate the usefulness of extreme value methods such as the POT approach and non-

stationary Generalized Pareto models for detecting structural changes in extreme event distributions.

Several limitations of this analysis should be noted. First, storm wind speed data are discretized, which can affect goodness-of-fit tests such as the Anderson-Darling test that assume continuous data. Second, the choice of threshold in POT modelling can influence the parameter estimates, although threshold stability analysis suggested that the selected threshold was appropriate. Third, the early-vs-late period split assumed a single structural change point, whereas changes in climate processes may occur gradually over time.

Additionally, the models assume independence between storm events, which may not fully capture clustering of storms within active seasons.

6. References

Dataset

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS). Bulletin of the American Meteorological Society.

Singh, K. (2025). NOAA Atlantic Hurricane Database (HURDAT2) – Cleaned Dataset. Kaggle. National Oceanic and Atmospheric Administration (NOAA). HURDAT2 Atlantic Hurricane Database.

Extreme Value Theory

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.

Software Packages

Gilleland, E., & Katz, R. W. (2016). extRemes: Extreme Value Analysis in R.

Pfaff, B. (2013). ismev: An Introduction to Statistical Modeling of Extreme Values.

7. Appendix

7.1 Exploratory Analysis

7.1.1 Spatial Exploration

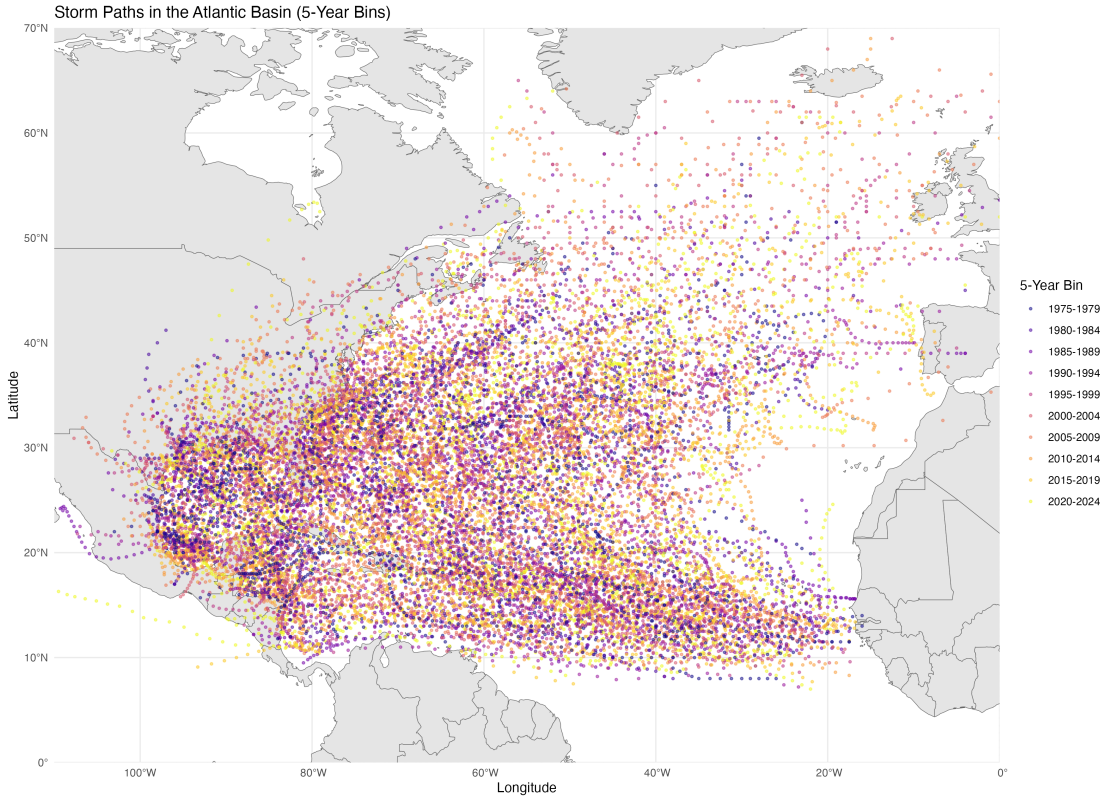

Figure A1: Geographic tracks of all recorded Atlantic tropical cyclones between 1975 and 2024, with colours indicating five-year periods.

7.1.2 Storm Frequency Exploration

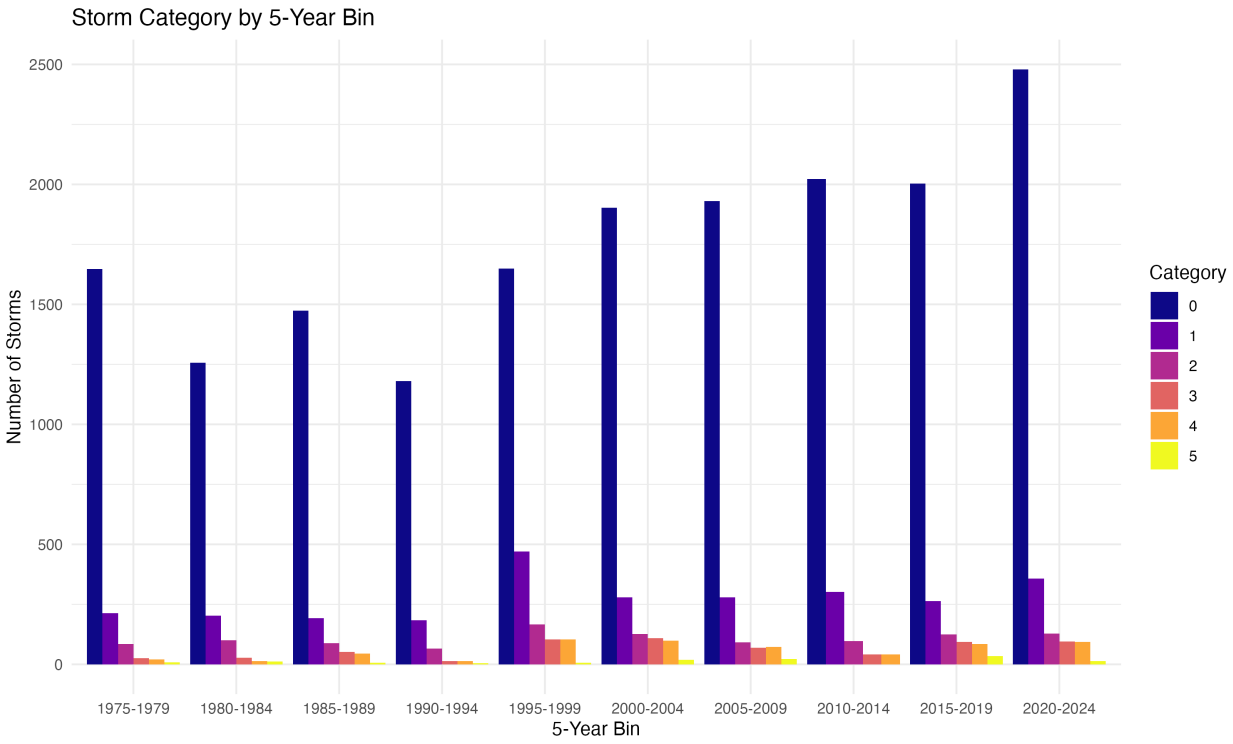

Figure A2: Number of storms in each Saffir–Simpson category per five-year bin, illustrating temporal variability in storm intensity classifications, where 0 is a non-hurricane cyclone and 5 is the strongest classification of a hurricane.

7.1.3 Storm Intensity Exploration

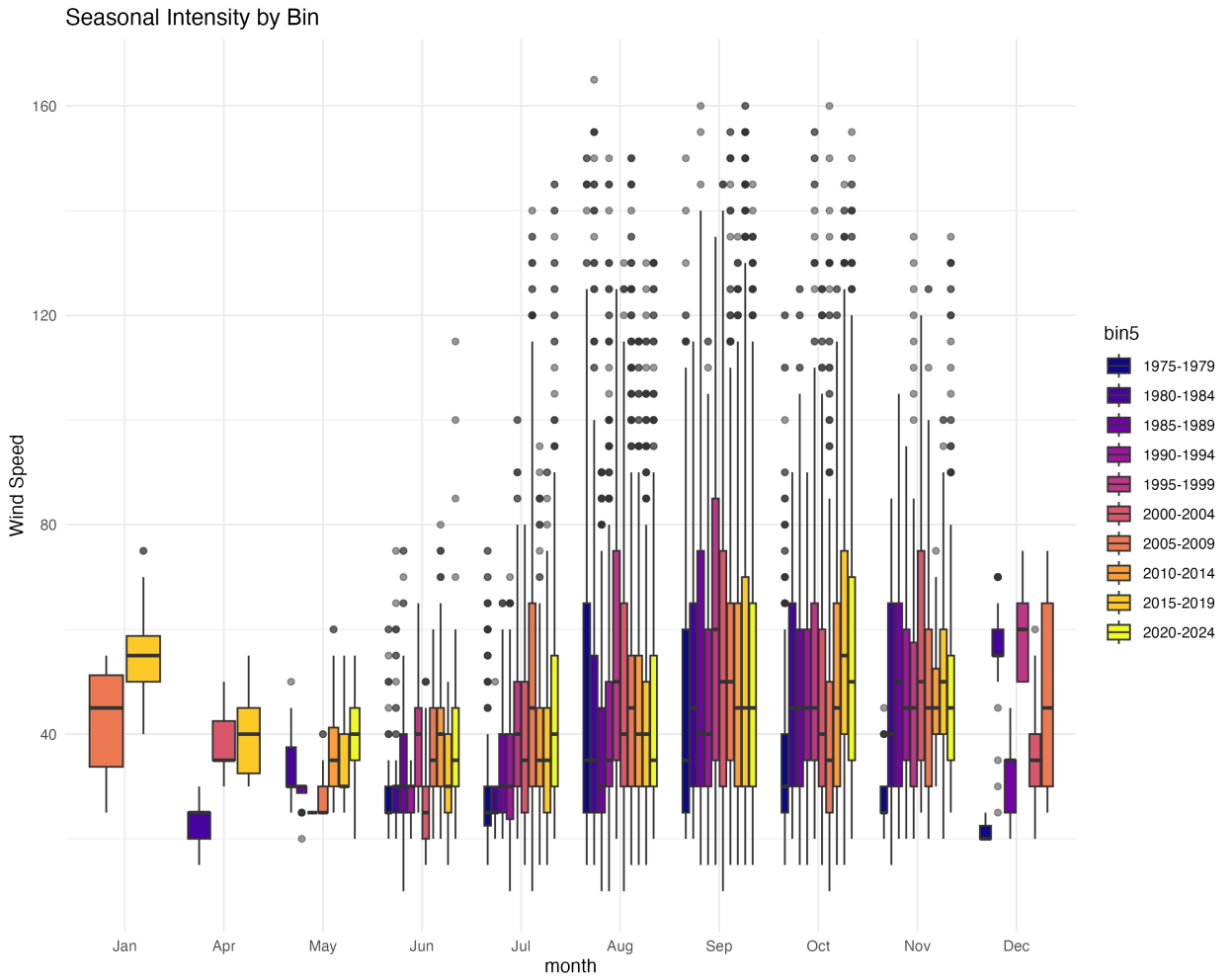

Figure A3: Monthly distribution of maximum wind speeds across five-year bins, highlighting seasonal and temporal variability in storm severity.

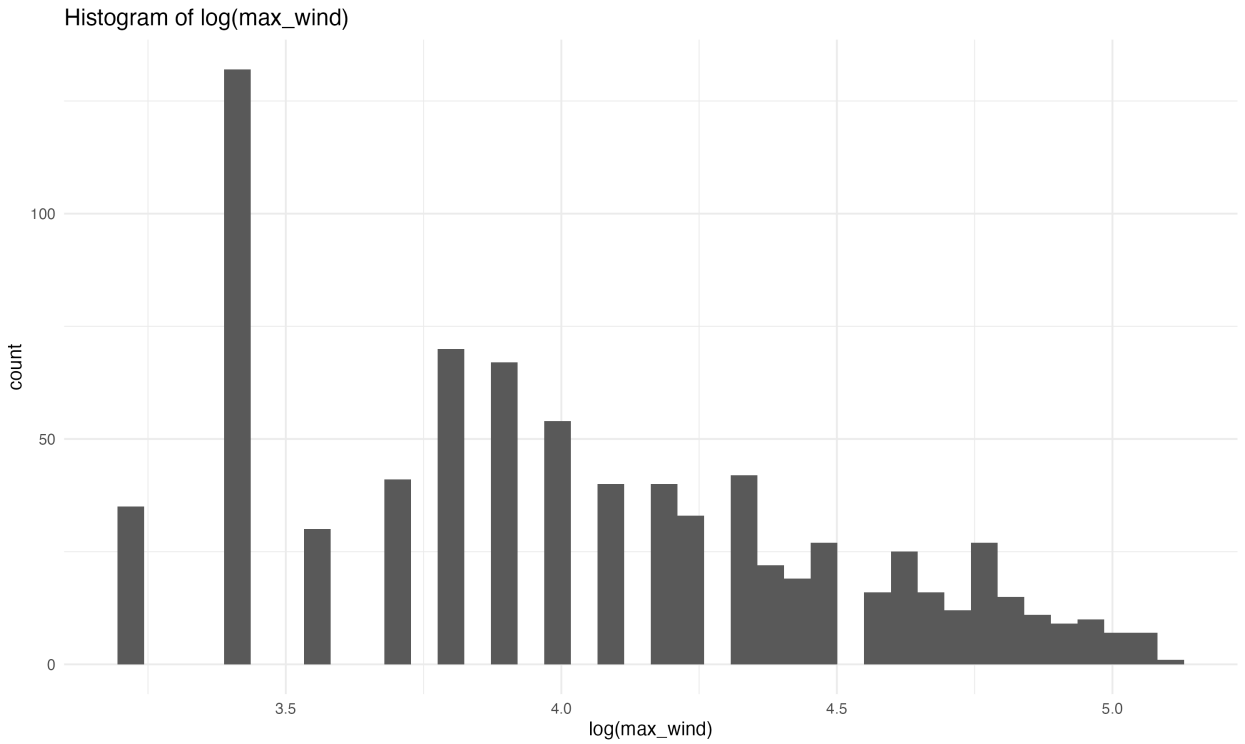

Figure A4: Histogram of log-transformed maximum wind speeds, illustrating the distributional characteristics and tail behaviour of storm intensity.

7.2 Frequency Analysis

7.2.1 Changepoint Model Comparison

Table A1: AIC and BIC comparison of stationary and non-stationary GPD models for storm severity.

AIC BIC Model Comparison for Frequency Changepoint

	tau	logLik	AIC	BIC
1	1,977	-152.048	308.096	311.920
2	1,978	-151.013	306.025	309.849
3	1,979	-150.723	305.445	309.269
4	1,980	-150.626	305.251	309.075
5	1,981	-149.866	303.732	307.556
6	1,982	-151.672	307.344	311.168
7	1,983	-152.554	309.109	312.933
8	1,984	-152.431	308.863	312.687
9	1,985	-152.542	309.085	312.909
10	1,986	-152.638	309.276	313.100
11	1,987	-152.610	309.220	313.044
12	1,988	-152.641	309.281	313.105
13	1,989	-152.635	309.270	313.094
14	1,990	-152.634	309.268	313.092
15	1,991	-152.550	309.100	312.925
16	1,992	-152.192	308.384	312.208
17	1,993	-151.681	307.361	311.185
18	1,994	-151.241	306.482	310.306
19	1,995	-151.780	307.560	311.384
20	1,996	-151.472	306.944	310.768
21	1,997	-150.436	304.871	308.695
22	1,998	-150.119	304.239	308.063
23	1,999	-150.103	304.207	308.031
24	2,000	-150.388	304.776	308.600
25	2,001	-150.506	305.011	308.835
26	2,002	-150.165	304.330	308.154
27	2,003	-150.850	305.699	309.523
28	2,004	-150.665	305.331	309.155
29	2,005	-152.056	308.113	311.937
30	2,006	-151.564	307.128	310.952
31	2,007	-151.519	307.039	310.863
32	2,008	-151.580	307.159	310.983
33	2,009	-150.890	305.780	309.604
34	2,010	-151.462	306.925	310.749
35	2,011	-151.838	307.677	311.501
36	2,012	-152.069	308.138	311.962
37	2,013	-151.820	307.639	311.463
38	2,014	-150.763	305.526	309.350
39	2,015	-149.811	303.623	307.447
40	2,016	-149.519	303.038	306.862
41	2,017	-149.635	303.270	307.094
42	2,018	-149.208	302.417	306.241
43	2,019	-149.216	302.431	306.255
44	2,020	-151.934	307.869	311.693
45	2,021	-152.318	308.635	312.459
46	2,022	-152.148	308.295	312.120

7.3 Severity Analysis

7.3.1 GPD Model Fit

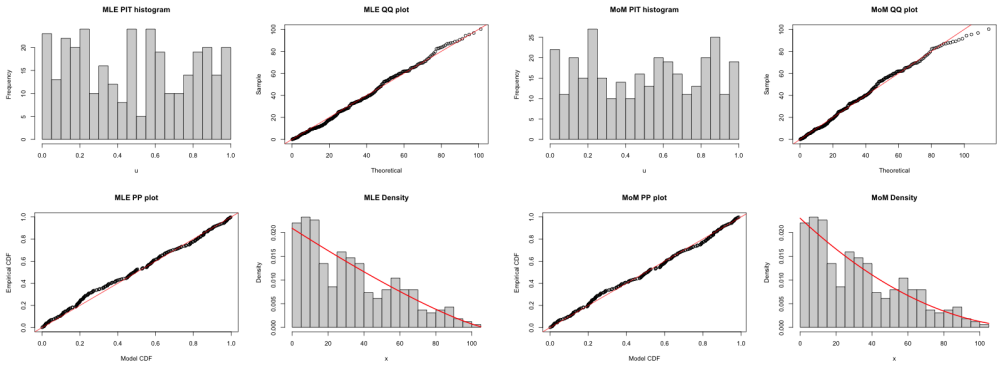

Figure A5: Comprehensive goodness-of-fit diagnostics (PIT, QQ, PP, and density plots) for the stationary GPD model using both Maximum Likelihood {left} and Method-of-Moments parameter estimates {right}.

Table A2: AIC and BIC comparison of stationary and non-stationary GPD models for storm severity.

AIC BIC Model Comparison

	AIC	BIC
Stationary	2,895.531	2,903.111
Scale Trend	2,889.831	2,908.781
Shape Trend	2,899.674	2,918.624
Shape+Scale Trend	2,893.420	2,923.740

7.3.2 Early vs Late Model

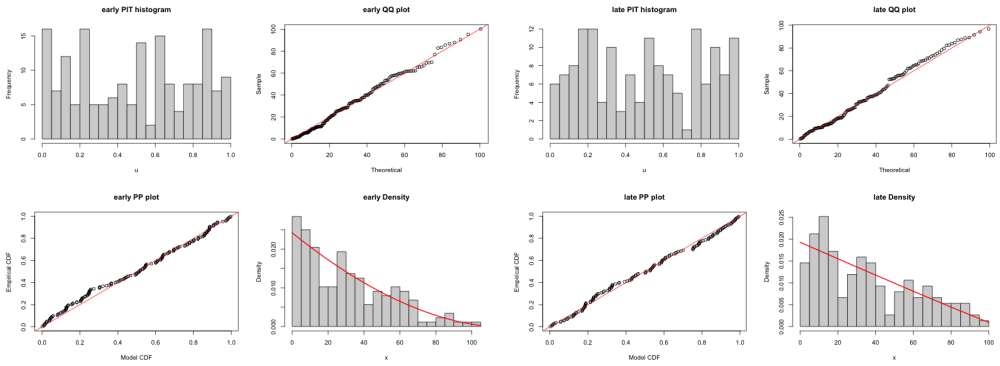

Figure A6: Goodness-of-fit diagnostics for GPD models fitted to early and late periods using Maximum Likelihood estimation.

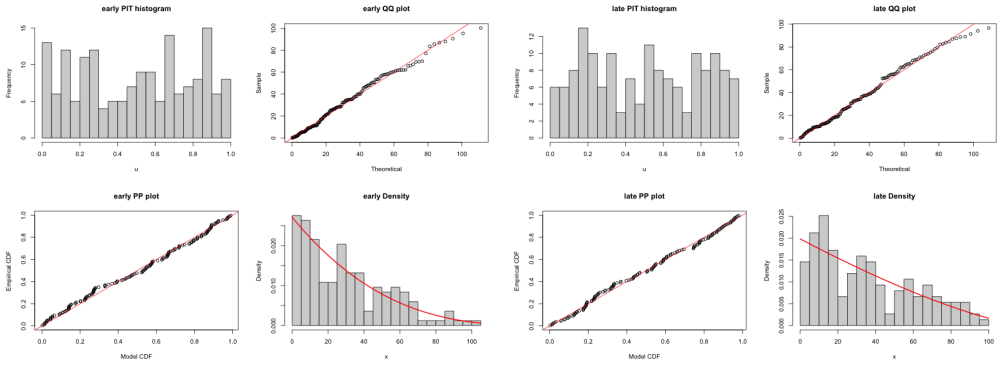

Figure A7: Goodness-of-fit diagnostics for GPD models fitted to early and late periods using Method-of-Moments estimation.

7.3.2 Changepoint Model

Table A3: Log-likelihood, AIC, and BIC values for candidate changepoint years in the GPD storm severity model.

AIC BIC Model Comparison for Severity Changepoint

	tau	logLik	AIC	BIC
1	1,977	-1,445.507	2,899.014	2,914.
2	1,978	-1,445.201	2,898.403	2,913.
3	1,979	-1,445.637	2,899.275	2,914.
4	1,980	-1,444.427	2,896.854	2,912.
5	1,981	-1,444.289	2,896.578	2,911.
6	1,982	-1,444.328	2,896.655	2,911.
7	1,983	-1,444.093	2,896.185	2,911.
8	1,984	-1,443.414	2,894.828	2,909.
9	1,985	-1,443.203	2,894.406	2,909.
10	1,986	-1,442.140	2,892.280	2,907.
11	1,987	-1,441.453	2,890.905	2,906.
12	1,988	-1,442.152	2,892.304	2,907.
13	1,989	-1,442.292	2,892.583	2,907.
14	1,990	-1,440.809	2,889.618	2,904.
15	1,991	-1,441.310	2,890.620	2,905.
16	1,992	-1,441.687	2,891.374	2,906.
17	1,993	-1,440.739	2,889.479	2,904.
18	1,994	-1,439.996	2,887.992	2,903.
19	1,995	-1,440.514	2,889.027	2,904.
20	1,996	-1,440.629	2,889.257	2,904.
21	1,997	-1,440.022	2,888.044	2,903.
22	1,998	-1,440.095	2,888.190	2,903.
23	1,999	-1,442.426	2,892.852	2,908.
24	2,000	-1,441.999	2,891.997	2,907.
25	2,001	-1,441.451	2,890.902	2,906.
26	2,002	-1,441.590	2,891.180	2,906.
27	2,003	-1,441.791	2,891.581	2,906.
28	2,004	-1,443.068	2,894.136	2,909.
29	2,005	-1,443.454	2,894.908	2,910.
30	2,006	-1,443.263	2,894.527	2,909.
31	2,007	-1,443.595	2,895.190	2,910.
32	2,008	-1,444.115	2,896.231	2,911.
33	2,009	-1,444.282	2,896.565	2,911.
34	2,010	-1,444.514	2,897.027	2,912.
35	2,011	-1,444.456	2,896.913	2,912.
36	2,012	-1,443.123	2,894.245	2,909.
37	2,013	-1,442.643	2,893.286	2,908.
38	2,014	-1,441.939	2,891.877	2,907.
39	2,015	-1,441.962	2,891.924	2,907.
40	2,016	-1,442.073	2,892.147	2,907.
41	2,017	-1,443.706	2,895.412	2,910.
42	2,018	-1,443.532	2,895.064	2,910.
43	2,019	-1,443.315	2,894.631	2,909.
44	2,020	-1,444.324	2,896.648	2,911.
45	2,021	-1,445.589	2,899.179	2,914.
46	2,022	-1,444.052	2,896.105	2,911.