Cyclone Landfall Frequency Analysis: Identifying High-Risk Countries and Territories on the American continent

Abstract

Using Bayesian methods and cylone data from 2000 to 2024, this report tracks and forecasts tropical cyclone landfall occurrences on the American continent. Seasonality patterns and yearly trends are investigated as well as regions most vulnerable to tropical cyclone impacts. The analysis showed a clear seasonal pattern, with most cyclone landfalls expected between June and November, yet no significant yearly trends, suggesting there is no strong long-term change in landfall occurrences over the studied period. The United States of America and Mexico stand out as the most high risk countries, with 30 and 28 predicted landfalls, respectively, over the next five years.

1. Introduction

Tropical cyclones are among the most destructive natural disasters, causing the highest number of casualties compared to all other natural disasters (Li & Li, 2013) as well as severe economic losses globally. These storms cause the most damage as they move toward land, bringing along extreme winds, heavy rainfall, storm surges and floods (Baradaranshoraka et al., 2017). Even when coastal communities are evacuated, cyclones can still cause extreme damage to properties and infrastructure as well as a significant loss of agricultural production. In addition to these, cyclone landfalls have long-term impacts, including economic slowdown, loss of coastal ecosystems and wildlife habitats, loss of businesses (especially local/family-owned enterprises), permanent displacement of coastal communities and significant rebuilding costs that can strain government budgets. More frequent and repeated exposure to these landfalls limits recovery time, leaving areas unprotected or poorly rebuilt, further exacerbating the associated damages.

Governments and disaster relief organisations have put together a variety of measures over the years, intended to mitigate these risks, including early warning systems, evacuation plans and strengthening infrastructure. However, the unpredictability of cyclone landfalls and the extent of regional disparities makes it crucial to understand landfall patterns and trends, in order to optimise response strategies tailored to communities in need.

Historically, landfall locations and frequency have been quite variable, but certain countries, such as the United States of America and Mexico, experience higher incidences than others. As such, understanding and accurately forecasting when and where cyclones are most likely to strike is essential for assessing the vulnerability of different areas, resource allocation and disaster planning. By identifying the areas at greatest risk as well as the landfall seasonal peaks, governments and organisations can ensure that resources such as emergency supplies, medical teams, and infrastructure support are on hand.

In this report, we focus on the evolution of hurricane landfall frequency since 2000 as well as analysing any monthly trends, by using Bayesian Poisson regression models and cyclone track data. We restrict our analysis to the American continent, investigating landfalls originating in the Atlantic basin (Atlantic Ocean, Caribbean Sea, and Gulf of America) and the Eastern Pacific basin (which extends from the western coast of Mexico and Central America to 140°W) (Ocegueda Sanchez, Chavas & Jones, 2025). Our aim is to identify monthly patterns and long-term trends in landfall frequency, as well as to highlight the most vulnerable countries and territories, in order to guide resource allocation and optimise disaster preparedness and response.

2. Data and Methodology

We based our analysis on the Atlantic hurricane database (HURDAT2), 1851-2024, more recently updated on April 4th, 2025 to include in the 2024 hurricane season, as well as the Northeast and North Central Pacific hurricane database (HURDAT2), 1949-2024, most recently updated on March 17, 2025 to include the 2024 hurricane season. These data sets are provided by the National Hurricane Center (NHC), as part of the National Oceanic and Atmospheric Administration (NOAA) and provides records of cyclone coordinates, maximum winds, central pressure, system status at 6 hour intervals. The HURDAT2 databases also include additional records outside of the standard intervals if the cyclone makes landfall, unexpectedly changes status or intensity or reaches a peak in terms of wind speed or pressure.

Although the Atlantic HURDAT2 database goes back to 1851, the NHC indicates that track data from the nineteenth century is often incomplete and inaccurate due to less advanced technology and therefore under reported and under analysed measurements. Moreover, international landfalls are only marked from 1951 to 1970 and 1991 onward, leading to significant gap in historical data.

Our analysis therefore focuses on landfall trends from 2000 to 2024, thus ensuring consistency and making use of improved landfall measurements. However, it is important to note that this introduces some limitations to our models, such as evaluating long term trends using a relatively short time period and sparse data, making our model predictions more susceptible to our initial assumptions.

Once we collected our data, we used geographical data, containing country outlines and coordinates, and combined these with our landfall coordinates to determine where the cyclone crossed a coastline. We say a cyclone makes landfall when the center of the system crosses a coastline. This center, however, usually ranges from 32 to 64km and therefore, certain landfall coordinates did not point to any land. In these cases, we mapped the landfall data points to the nearest country. We also note that certain cyclones make more than one landfall, however for the purpose of this report, our analysis focuses solely on the number of landfalls.

We first modeled the number of cyclone landfalls per month and year using Bayesian Poisson regression models (Elsner, Bossak & Niu, 2001), in order to explore any seasonality effects and yearly trends. Our first model included both a monthly effect and a non-linear yearly effect, due to the high fluctuations in cyclone landfall numbers per year. Our second model only included the monthly effect, ignoring any long-term trends. We then used a statistical method called Leave-One-Out Cross Validation (LOO) to select the model with the best performance, in this case, our second model.

Our third model then investigated where cyclones were most likely to land, amongst the top 10 countries and commonwealths which have experienced the most cyclone landfalls since 2000. These include Antigua and Barbuda, the Bahamas, Belize, Canada, Cuba, the Dominican Republic, Mexico, Nicaragua, Puerto Rico and the United States of America.

Using our best performing seasonal model and our geographical model, we then simulated landfall counts for each month as well as for each of the ten most affected countries and territories over the next five years (2025 to 2029). These forecasts thus support preparedness, planning and resource allocation in high risk areas.

For more details on methodology and technical analysis, please refer to the Appendix.

3. Results

3.1 Cyclone trends over time

- plots of yearly number of cyclones and monthly cyclone numbers/ landfall frequency
- talk about models (again not very technical)
- landfall probabiltiy plots
- landfall predictions

- what this means

3.2 areas at high risk

- map plots
- talk about model
- landfall probabiltiy plots
- landfall predictions
- what this means

4. Recommendations

5. Limitations of models and data biases

6. Conclusion

References

Baradaranshoraka, M., Pinelli, J.-P., Gurley, K., Peng, X. & Zhao, M. (2017) Hurricane wind versus storm surge damage in the context of a risk prediction model. *Journal of Structural Engineering*. 143 (9), 04017103. doi:10.1061/(ASCE)ST.1943-541X.0001824.

Elsner, J.B., Bossak, B.H. & Niu, X.-F. (2001) Secular changes to the ENSO-u.s. Hurricane relationship. *Geophysical Research Letters*. 28 (21), 4123–4126. doi:https://doi.org/10.1029/2001GL013669.

Li, K. & Li, G. (2013) Risk assessment on storm surges in the coastal area of Guangdong Province. *Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards*. 68 (2), 1129–1139. doi:10.1007/s11069-013-0682-2.

Ocegueda Sanchez, J.A., Chavas, D.R. & Jones, J.J. (2025) Interannual variability of tropical cyclone landfalls in the eastern north pacific: Environmental drivers and implications. *Geophysical Research Letters*. 52 (8), e2024GL113807. doi:https://doi.org/10.1029/2024GL113807.

Appendix - Technical description of analysis and data processing

1. Dataset Limitations and Assumptions

Source of the Data: The dataset used in this analysis was sourced from [name of the dataset/source]. It contains historical hurricane data, including landfall locations, intensities, and dates for each hurricane event. Missing Data: While the dataset covers a wide range of hurricanes, there may be gaps in certain years or geographical regions. For example, data for some lesser-known areas or smaller storms might be underrepresented. Data Quality: The dataset was cleaned and preprocessed to remove outliers, but it's important to note that the data relies on historical records, which may contain inaccuracies or changes in reporting standards over time. Geographical Considerations: The analysis focuses on hurricanes that made landfall in the following regions: [list countries or regions]. This exclusion of other types of data (such as storms in open oceans or storms that only impact coastal areas) may limit the scope of the analysis for some applications. Temporal Scope: The dataset spans from [start year] to [end year]. Any shifts in hurricane frequency or intensity over time are accounted for in this period, but trends beyond this range were not considered. 2. Data Preprocessing and Cleaning

Data Import: The dataset was loaded using the following code: import pandas as pd data = pd.read_csv('hurricane_data.csv') Missing Values: Any missing values in the dataset were handled by [imputation method, dropping rows, etc.]. In some cases, the [specific data column(s)] had to be excluded due to a high percentage of missing values. Geographical Filtering: The analysis only included data where hurricanes made landfall in the [target countries/regions]. The filtering process was executed with the following code: landfall_data = data[data['Country'].isin(['Country1', 'Country2', 'Country3'])] 3. Statistical Model Details

Poisson Regression Model To model the frequency of hurricane landfalls across different countries, a Poisson regression model was employed, as the number of landfalls is a count variable.

Model Overview: The Poisson regression model is used because the outcome (number of landfalls) follows a count distribution. This model assumes that the mean of the count variable is equal to its variance. Model Equation:

Assumptions: The count of hurricane landfalls in each country is independent. There is no overdispersion (variance equals the mean), which was checked using a goodness-of-fit test. Model Code Example: from statsmodels.api import GLM from statsmodels.genmod.families import Poisson

Define the model

model = GLM(count_data, X, family=Poisson())

Fit the model

result = model.fit()

Print summary of results

print(result.summary()) Model Checking Goodness of Fit: The model was evaluated for its fit using Pearson's chi-squared test. This test checks whether the model adequately fits the data, considering overdispersion. Overdispersion Test: We tested for overdispersion, which could indicate that the Poisson assumption of mean equals variance does not hold. If overdispersion was found, alternative models such as negative binomial regression could be considered. 4. Reproducibility

Code and Environment Setup: The analysis was performed in a Python environment with the following package dependencies: pandas (version x.x.x) numpy (version x.x.x) statsmodels (version x.x.x) matplotlib (version x.x.x) To ensure reproducibility, the environment can be recreated by installing the required dependencies through the requirements.txt file:

pip install -r requirements.txt Code Execution: All code in the Jupyter notebooks can be executed sequentially, starting from the first cell. This includes data preprocessing, model fitting, and visualization steps. The output for the plots and model summary can be reproduced by running the respective code blocks. Randomness in Analysis: To ensure reproducibility when using randomized algorithms, we have fixed random seeds in the analysis: import numpy as np np.random.seed(42) 5. Potential Extensions and Future Work

Further Regional Breakdown: The analysis can be extended by investigating landfall frequency in more granular regions within each country (e.g., coastal vs. inland areas). Model Enhancements: Future work could explore more sophisticated models such as negative binomial regression if overdispersion is found, or even spatial models that account for geographical proximity and hurricane paths. Longer Time Horizons: Extending the dataset to include more recent hurricane data, or projecting landfall frequency into the future based on climate change scenarios, could provide valuable insights. Conclusion This appendix provided detailed information on the dataset used, model details, assumptions, and steps for ensuring the reproducibility of this analysis. All code is available for execution in the provided Jupyter notebooks, and a clear step-by-step process for recreating the analysis has been outlined.