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, in order to help humanitarian agencies and governments optimise cyclone preparedness and resource allocation. 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.

Contents

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 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 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 provide 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 a cyclone makes landfall, unexpectedly changes status, 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 as well as under reported and under analysed measurements. Moreover, international landfalls are only marked from 1951 to 1970 and 1991 onward, leading to a 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 combined our landfall coordinates with geographical data, containing country outlines and coordinates, to determine where each landfalling 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. Model A, included both a monthly effect and a non-linear yearly effect, due to the high fluctuations in cyclone landfall numbers per year. Model B, 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.

Model C then investigated where cyclones were most likely to land, amongst the top 10 countries and territories 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 time-based and geographical models, 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 support preparedness, planning and resource allocation in high risk areas.

For more details on data processing, methodology and our Bayesian analysis, please refer to the Appendix.

3 Results

3.1 Cyclone trends over time

To better understand cyclone landfall patterns in the last few decades, we first explored the frequency and distribution of landfalls from 2000 to 2024.

Figure 1 summarises key trends in landfall occurrences over time, including the total number of landfalls per year, the number of landfalls per cyclone, and the average number of landfalls per month. These plots provide insight into both seasonal patterns and the highly variable nature of cyclone landfalls, as well as the typical landfall frequency for each individual tropical cyclone from 2000 to 2024. Annual landfall counts show significant year-to-year fluctuations, reflecting the highly variable nature of cyclone landfalls, whereas the monthly landfall averages show a clear seasonal pattern, with most cyclones making landfall between June and November, consistent with the Atlantic hurricane season (Truchelut et al., 2022). While most tropical cyclones never make landfall, it is important to remember that a single landfall can result in immense destruction of both human life and infrastructure, reinforcing the importance of monitoring these rare yet costly events.

To explore yearly and monthly patterns in cyclone landfalls, we fitted two Bayesian Poisson regression models to our landfall count data, stratified by year and month. These models allowed us to assess both long-term variability and within-year seasonality. The first model (model A) included both a monthly seasonal effect and an annual effect, while the second (model B) focused only on monthly seasonality. Having compared model performance using Leave-One-Out Cross Validation (LOO), we found that both models performed very similarly, although model B provided a slightly better fit to our data, suggesting that there is no consistent long-term trend across the studied 25-year period.

Model B was then used to simulate landfall occurrences in 2025, on a monthly basis, showing an expected peak in landfall activity between August and October.

Figure 2 shows a clear seasonal pattern in our landfall forecasts, aligning with the typical Atlantic and Pacific hurricane seasons. This figure also emphasises the rarity of landfall occurrences outside of these peak months, with a very low probability of seeing any landfalls at all.
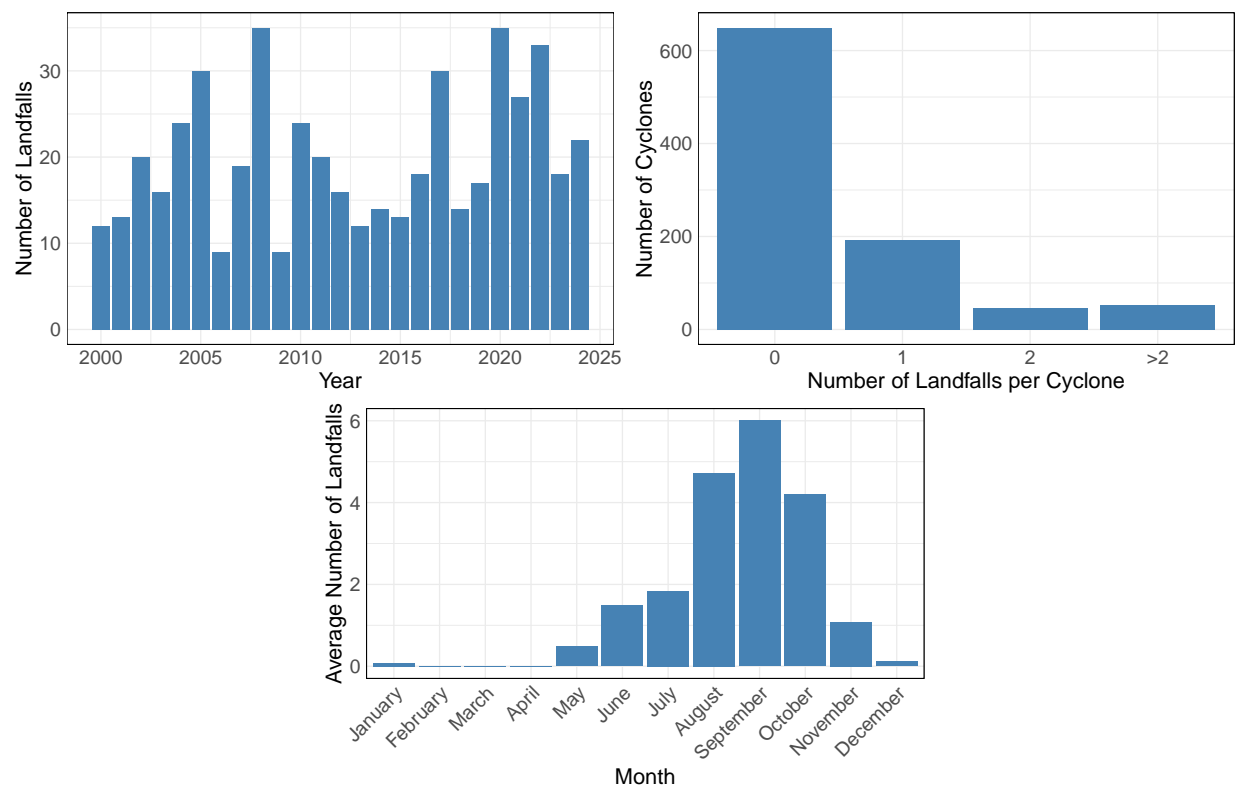
Figure 1: (a) Number of cyclone landfalls per year (2000–2024). (b) Number of landfalls per cyclone (2000-2024). (c) Average number of landfalls per month (averaging across 2000-2024).
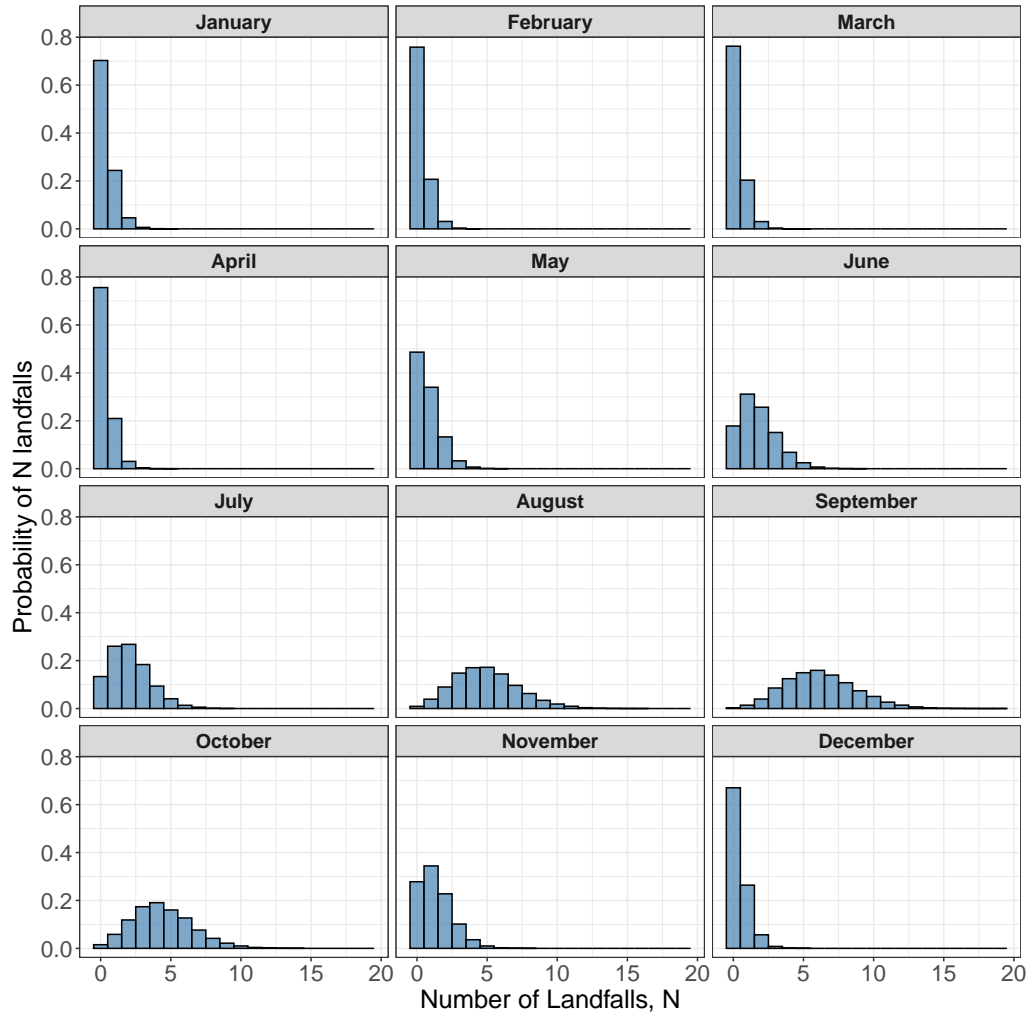
Figure 2: Estimated probability distribution of cyclone landfalls per month in 2025, based on the seasonal model.

We then took our analysis one step further and examined expected landfall forecasts per month over the next five years (2025-2029).

MONTH	2.5% quantile	97.5% quantile	median
January	0	5	2
February	0	4	1
March	0	4	1
April	0	4	1
May	0	8	3
June	3	16	8
July	4	18	10
August	15	36	24
September	20	44	31
October	13	33	22
November	2	13	6
December	0	5	2

Figure 3: Forecasted number of landfalls per month, 2025-2029, based on the seasonal model.

Figure 3 displays the total forecasted number of cyclone landfalls per month from 2025 to 2029, showing the 2.5% and 97.5% quantiles and the median value of total landfalls for each month. These values serve as a measure of statistical uncertainty associated with our model. We expect the actual number of total landfalls over this five year period to fall between 2.5% and 97.5% quantile values with 95% certainty.

The forecast predicts the highest number of total landfalls in September, with a median value of 31 landfalls, followed closely by August and October, with median values of 24 and 22 landfalls, respectively. The months of January to April and December are forecasted to experience significantly fewer landfalls, with the median forecast for total landfalls being close to zero in these months.

Our forecasts also highlight the unpredictability of cyclone landfalls, with considerable spread between the 2.5% quantile and 97.5% quantile in the June to November months. The wide range of uncertainty in our forecasts for these months underlines the importance of continued research into cyclone activity as to refine and improve model predictions.

3.2 High-risk countries and territories

Before developing our geographical model, we first investigated landfalling cyclone paths from 2000 to 2024. This exploratory analysis plays a crucial role in highlighting cyclone patterns such as intensity, frequency and geographical distribution of landfalls. By visualising cyclone paths and location of landfalls, we can determine which countries and territories are consistently impacted, helping to prioritise areas most in need for our forecasting model.

Figure 4 shows the full path of tropical cyclones that made landfall between 2000 and 2024 across the East Pacific and Atlantic basins. Each path represents an individual cyclone, as well as its wind speed as the cyclone advances. This map highlights how far cyclones can travel before making landfall, emphasising the importance of early detection and consistent monitoring for effective cyclone preparedness.

The erratic nature of cyclone paths, seen by the sudden changes in direction and intensity can leave communities with limited time to evacuate, reinforcing the critical need for raising awareness of regions most vulnerable so governments can enhance response systems, invest in resilient infrastructure and access needed support.
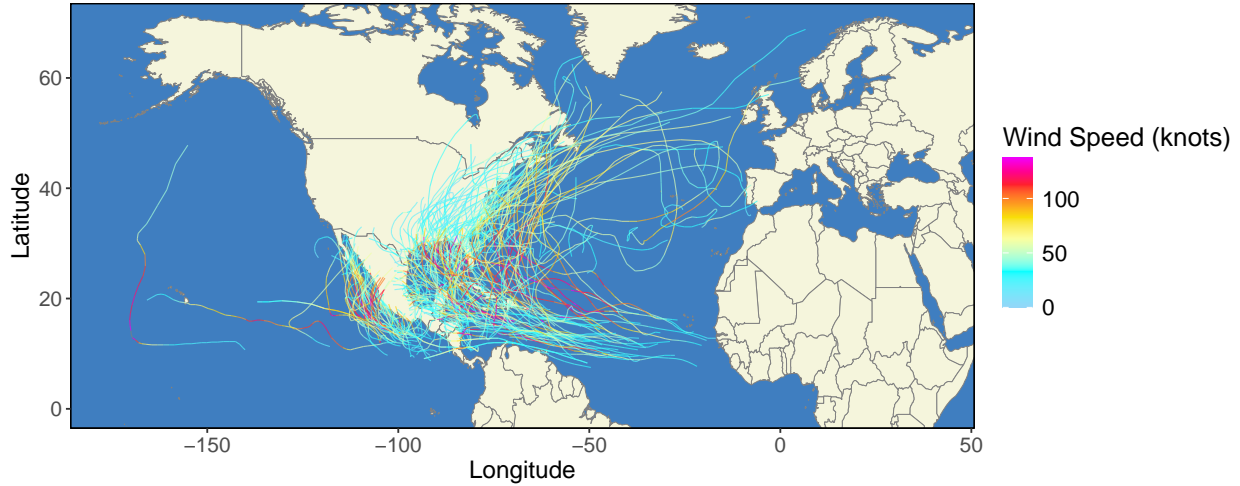
Figure 4: Landfalling cyclone paths in the East Pacific and Atlantic basins, 2000-2024.

We next took a closer look at landfall locations from 2000 to 2024 in order to identify regions most affected by landfalling cyclones.

Figure 5: Number of tropical cyclone landfalls per country, 2000-2024.

Figure 5 provides a national-level overview of landfall frequency in the last 25 years, with the United States, Mexico, Nicaragua, the Bahamas, Cuba and Canada standing out as the most impacted by cyclone activity, thus identifying countries with the highest exposure to cyclone landfalls. This is crucial for establishing which governments need the most support.

Figure 6, in contrast, provides a geographical overview of landfall locations since 2000, allowing us to identify which coastal regions are must vulnerable to cyclone activity. We see that the Gulf Coast and East Coast of the United States experience a high number of landfalls, whereas the Pacific coast of the United States hasn't seen any since 2000. On the other hand, both Mexico's East and West coasts have high landfall exposure. This map is particularly important for local disaster preparedness, showing which areas are most

Figure 6: Landfall geographical locations, 2000-2024.

susceptible to damages caused by cyclone landfalls.

To explore geographical patterns in cyclone landfalls, we fitted a Bayesian Poisson regression model to our landfall count data, now stratified by year and location. This model (model C) allowed us to assess country and territory-level vulnerability related to cyclone landfalls. Model C ignores yearly time trends based on the results of our previous time-based models (where we saw a better performance from our model without a year effect). Instead, model C incorporates country and territory-level effects for the top ten countries with the most landfalls since 2000: the United States of America, Mexico, Cuba, the Bahamas, Nicaragua, Canada, Belize, Dominican Republic, Antigua and Barbuda and Puerto Rico, as seen in Figure 5.

We then used model C to simulate landfall occurrences in 2025, for each of the countries and territories in our top ten list. Figure 7 clearly isolated the United States of America and Mexico as the most vulnerable regions in 2025, even amongst the top ten landfall-exposed countries. This figure reinforces the rarity of landfall occurrences outside of these two countries, with other countries having a very low probability of seeing any landfalls.
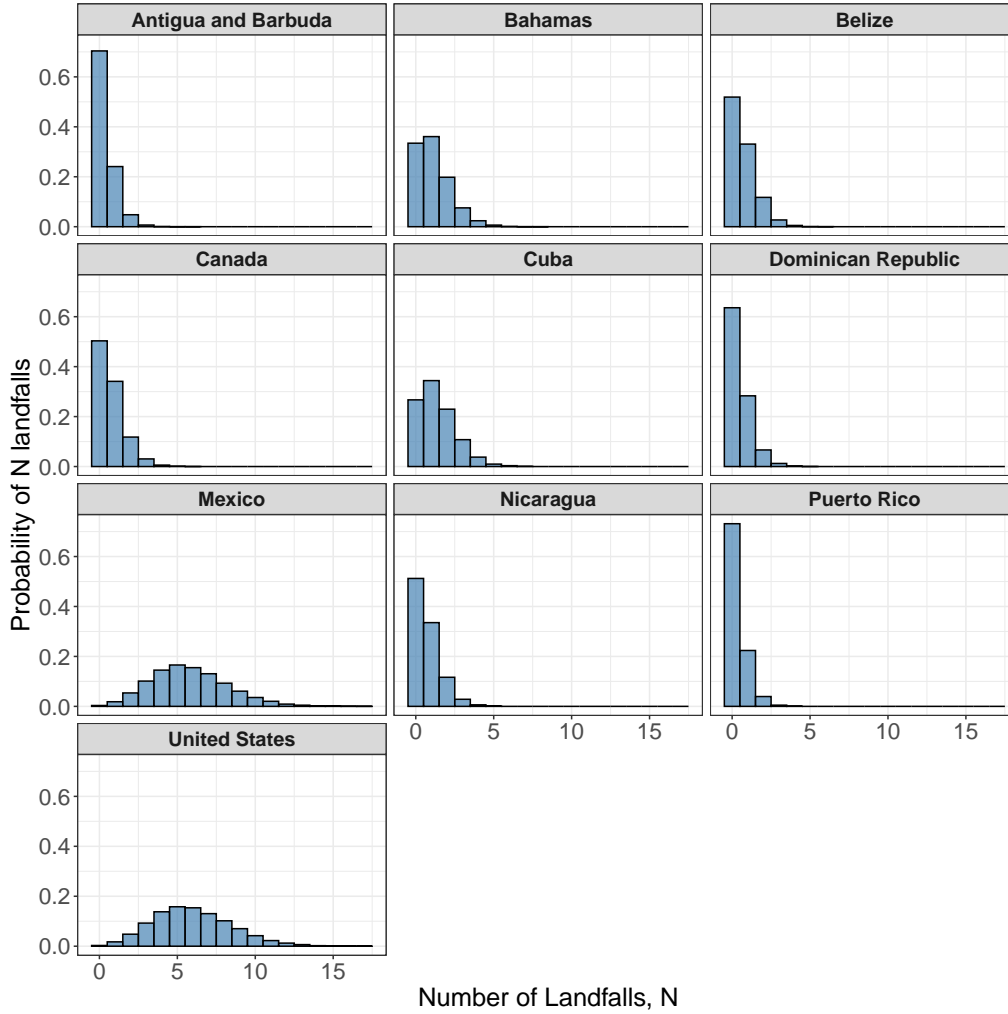
Figure 7: Estimated probability distribution of cyclone landfalls per country and territory in 2025.

We then took our analysis one step further and examined expected landfall forecasts for each of our top ten countries and territories over the next five years (2025-2029).

Figure 8 displays the total forecasted number of cyclone landfalls per country and territory from 2025 to 2029, showing the 2.5% and 97.5% quantiles and the median value of total landfalls for each month.

The forecast predicts the highest number of total landfalls in the United States of America, followed closely by Mexico , with median values of 30 and 29 landfalls, respectively. Puerto Rico, the Dominican Republic and Antigua and Barbuda have the lowest forecasted landfalls, with a median value of total landfalls over the next 5 years close to zero, despite being in the top ten most exposed countries and territories.

Our forecasts once again highlight the unpredictability of cyclone landfalls, with considerable spread between the 2.5% and 97.5% quantiles for some of our countries and territories such as the United States of America, Mexico, the Bahamas and Cuba. The wide range of uncertainty in our forecasts for these countries once again underlines the importance of continued research into cyclone activity as to refine and improve model predictions.

LOCATION	2.5% quantile	97.5% quantile	median
Antigua and Barbuda	0	5	2
Bahamas	1	11	5
Belize	0	8	3
Canada	0	8	3
Cuba	2	13	7
Dominican Republic	0	6	2
Mexico	18	41	28
Nicaragua	0	8	3
Puerto Rico	0	5	1
United States	19	42	30

Figure 8: Forecasted number of landfalls per country and territory, 2025-2029.

4 Recommendations

Based on our analysis of historical landfall patterns and probabilistic forecasts, we recommend the following actions to reinforce cyclone preparedness and resource allocation:

- Allocate emergency supplies and necessary resources ahead of the peak season (June to November) with response units in place and ready to act, especially in August to October, when the probability of landfalls reaches a peak.
- Prioritise the United States of America and Mexico, followed by the Bahamas and Cuba, for storing medical kits, food, water, and to put in place temporary shelters.
- Support local authorities in vulnerable areas through evacuation drills and strengthening infrastructure (especially hospitals, schools, and emergency shelters).

By aligning these targets with the seasonal and geographic insights from our models, organisations can reduce disaster response times, minimise casualties, and accelerate recovery.

5 Limitations of models and data biases

While our Bayesian Poisson regression models provide valuable insights, several limitations should be considered:

- Landfall coordinates which fell over water were mapped to the nearest country, introducing geographical uncertainty.
- Our models treated all landfalls equally, without distinguishing storm intensity or amount of damage caused.
- The data used in our model was sparse, and the resulting choice to ignore yearly time trends in cyclone behaviour over 2000–2024, may lead to inaccuracies due to accelerating climate change.
- Wide credible intervals in our five-year forecasts reflect both variability and uncertainty in our forecasts.
- Our models do not account for population density, infrastructure quality, or socioeconomic resilience, which influence actual humanitarian need. While countries like the United States may experience the highest number of landfalls, their relatively stronger economic resilience allows for stronger infrastructure and faster recovery. In contrast, economically weaker countries in the region may face catastrophic consequences from even a single cyclone landfall. These regions might rely more heavily on humanitarian aid to reduce the impact of cyclone landfalls on vulnerable populations and infrastructure.

6   Conclusion

This report demonstrates a clear seasonal cycle in tropical cyclone landfalls, with most occurring between June and November. Our results indicate that there is no significant trend in annual landfall counts from 2000 to 2024. Moreover, our forecasts identify the United States and Mexico as the highest-risk countries over the next five years, with a substantial probability of multiple landfalls in each hurricane season.

Through our use of Bayesian models, we hope to provide actionable insights to help humanitarian agencies and governments optimise cyclone preparedness and resource allocation. Despite some uncertainties and data limitations, these findings offer a foundation for region-specific disaster risk reduction efforts.

7   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.

Truchelut, R.E., Klotzbach, P.J., Staehling, E.M., Wood, K.M., Halperin, D.J., Schreck, C.J. & Blake, E.S. (2022) Earlier onset of North Atlantic hurricane season with warming oceans. *Nature Communications.* 13 (1), 4646. doi:10.1038/s41467-022-31821-3.

8 Appendix - Technical description of analysis and data processing

8.1 Data processing

Both the Atlantic and Northeast and North Central Pacific hurricane databases (HURDAT2) contain cyclone information recorded at six hour intervals as well as when a cyclone changes status unexpectedly. The information contained in each row of these databases includes:

- storm ID (contains Basin, year, and cyclone number for that year);
- name, if available;
- number of recorded track data for each cyclone;
- date;
- time;
- record identifier (identifies landfalls or indicates a reason for the inclusion of the record outside of the regular six hour intervals);
- cyclone position (measured in latitude and longitude);
- system status;
- maximum sustained wind (in knots);
- minimum pressure (in millibars);
- radius of maximum wind (in nautical miles).

For the purpose of this report, the two databases were combined and the following features were extracted from cyclones in 2000 to 2024: - storm ID; - name, if available; - day; - month; - year; - time; - record identifier; - cyclone position (measured in latitude and longitude, converted to positive and negative decimals); - system status; - maximum sustained wind (in knots, converted to integers); - minimum pressure (in millibars, converted to integers).

This data set was then used to identify all landfall events in the studied period, using the record identifier "L". We then matched our landfall data to countries and territories in the North American continent, using vector maps of country and territory boundaries. If the latitude and longitude coordinates of a landfall did not fall within a country or territory boundary, we matched it to the closest country or territory boundary point. This finalised landfall data set, containing geographical information was then used for all the models in this study.

8.1.1 Monthly and yearly landfall frequency (Models A and B)

For both seasonal models A and B, landfalls were grouped by month and year, from 2000 to 2024. We then counted the number of landfalls in each group and matched this to all month and year combinations, in order to count which groups (month/year combinations) had zero landfalls.

We gave each month a numerical ID number, ranging from 1 to 12 (January to December) and we included our prediction data for 2025 to 2029, with unknown counts, at the end of our seasonal landfall frequency data set.

Finally, years were standardised to the range [-1,1] using the following formula:

$$\text{standardised year} = \frac{\text{year} - 2000}{2029 - 2000}$$

.

8.1.2 Geographical landfall frequency (Model C)

The top 10 most exposed countries and territories were identified based on total number of landfalls in the studied period.

We then filtered our landfalls data set to only include these 10 countries and territories, and grouped landfalls by year and country or territory (location), from 2000 to 2024. We then counted the number of landfalls in each group and matched this to all location and year combinations, in order to count keep track of the groups with zero landfalls.

Similarly to our previous frequency data set, we gave each location a numerical ID number, ranging from 1 to 10 and we included our prediction data for 2025 to 2029, with unknown counts, at the end of our data set.

Finally, we encoded the location information in our geographical count data set using a one-hot design matrix, where each row matches a year-location combination and each column represents a location and contains a 1 if the row matches the location, or a 0 otherwise.

8.2 Model A: Monthly seasonal model with yearly effect

Our first model represents a Bayesian Poisson Regression model, with a monthly seasonality effect, and a non-linear yearly effect (modeled using a Gaussian Process). This model was fitted using Markov Chain Monte Carlo (MCMC) methods.

Mathematically, our model is denoted:

$$Y_i \sim \text{Poisson}(\lambda_i)$$
$$\log \lambda_i = \beta_{m(i)} + f(T(i))$$
$$\beta \sim \text{N}(0, \frac{M}{M-1}(I - \frac{1}{M}\mathbf{1}\mathbf{1}^T) \cdot \sigma_\beta^2), \text{ (sum-to-zero constraint)}$$
$$f \sim \text{HSGP}(\alpha, \rho)$$
$$\sigma_\beta \sim \text{Half-Cauchy}(0, 1)$$
$$\alpha \sim \text{Half-Cauchy}(0, 1; b, m)$$
$$\rho \sim \text{Inv-Gamma}(5, 1)$$

where

- i indexes observations, $i = 1, ..., 300$;
- M is the number of months, $M = 12$;
- Y_i is the observed count of landfalls for observation i;
- $m(i)$ is a map from i (the observations) to corresponding month ID (integers between 1 and 12);
- $T(i)$ is a map from i (the observations) to the corresponding standardised year;
- $\beta \in \mathbb{R}^{12}$ is the monthly seasonal effect, where β_1 is the effect for month 1 (January), β_2 is the effect for month 2 (February) and so on;
- f is given a zero-mean HSGP prior with squared exponential kernel with GP variance α and length scale ρ.

- σ_β scales the variance of β to implement non-centered parameterisation on top of the sum-to-zero constraint, where both constraints were implemented to improve parameter identifiability, sampling efficiency and convergence of our model parameters;
- λ_i is the expected number of landfalls per month and year

We ran our MCMC algorithm using 3 parallel Markov chains, 4000 iterations, discarding the first 500 as "warm-up" iterations. We verified model convergence by investigating key model parameter convergence diagnostics, as well as checking trace plots for our worst performing parameter and looking at pairwise posterior geometries (thus allowing us to verify that all our parameters are identifiable).

variable	rhat	ess_bulk	ess_tail
beta_sd	1.002096	1767.604	3521.748
gp_sigma	1.000996	2914.562	2902.679
gp_lengthscale	1.000500	3606.458	5243.286
z[1]	1.000687	4547.580	3199.904
z[3]	1.000585	5483.388	4442.956

Figure 9: Model convergence diagnostics for our 5 worst performing parameters.

Figure 9 displays 3 statistical efficiency and convergence measures for our worst performing parameters (in terms of the "ess_bulk" measure, where a lower "ess_bulk" value represents lower efficiency). In general, for convergence, we expect an "rhat" value close to 1, and for efficiency we expect "ess_bulk" and "ess_tail" values greater than 1500 (500 for each parallel chain). As we can see, all "rhat" convergence diagnostic values are less than 1.01, which suggests good mixing and convergence. Moreover both "ess_tail" and "ess_bulk" suggest good sampling efficiency (as they are greater than 500 per Markov chain, a common baseline for efficiency in Bayesian MCMC models).

We then checked model fit by performing a posterior predictive check (we simulate data from our model's posterior distribution and compare these to the actual observed data to see if our model explains the data well). To verify model fit, we calculated the probability that an observed data point lies in the 95% credible interval of the corresponding posterior distribution.

We found that 95.7% of our observed landfall counts fell within our model's posterior predictive intervals, suggesting a good model fit. This can be seen in Figure 10.
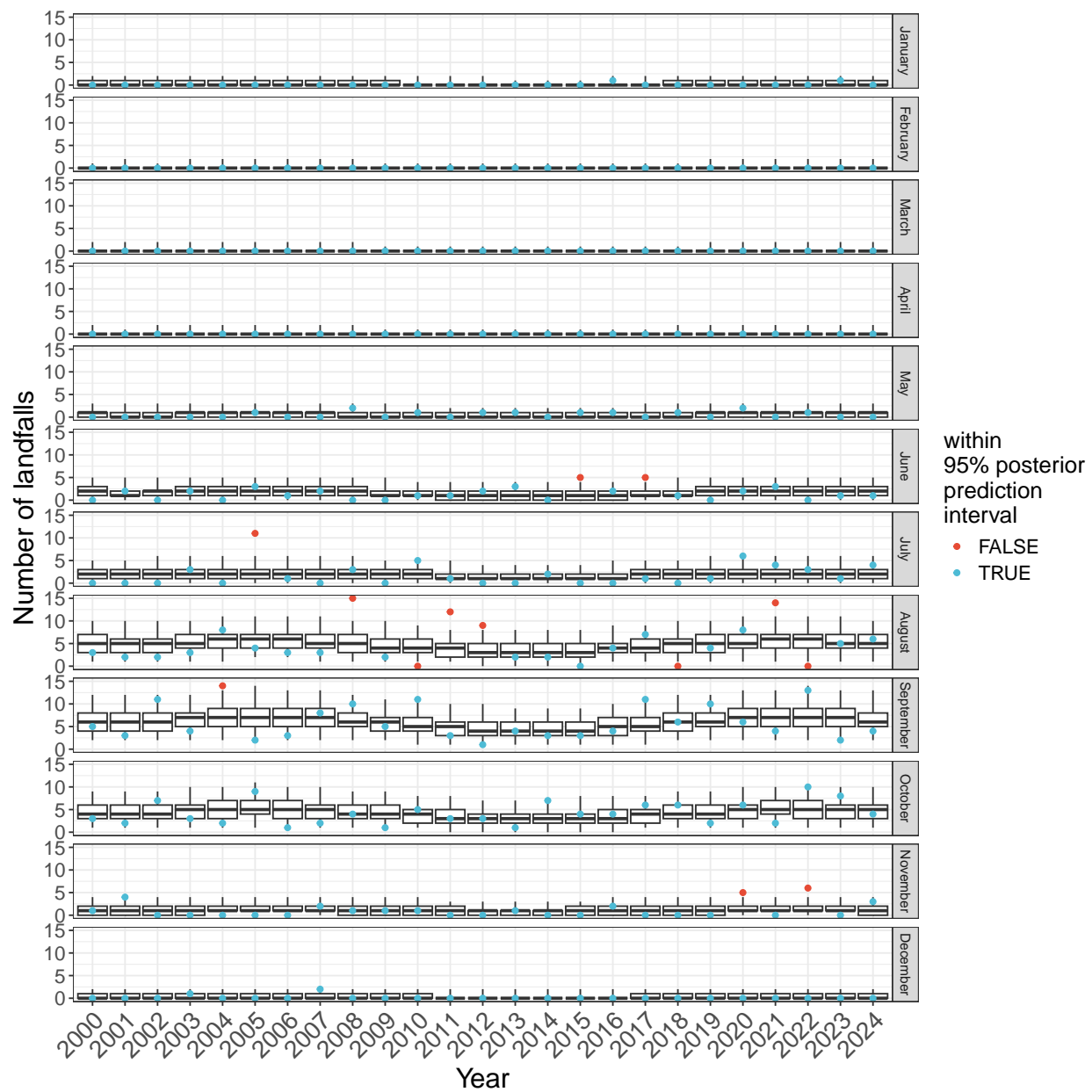
Figure 10: 95% posterior credible intervals and observed data for each month and year, 2000-2024.

8.3   Model B: Monthly seasonal model without yearly effect

Our second model represents a Bayesian Poisson Regression model, with a monthly seasonality effect. This model was also fitted using Markov Chain Monte Carlo (MCMC) methods.

Mathematically, our model is denoted:

$$Y_i \sim \text{Poisson}(\lambda_i)$$
$$\log \lambda_i = \beta_{m(i)}$$
$$\beta \sim \text{MVN}(0, \frac{M}{M-1}(I - \frac{1}{M}\mathbf{1}\mathbf{1}^T) \cdot \sigma_\beta^2), \text{ (sum-to-zero constraint)}$$
$$\sigma_\beta \sim \text{Half-Cauchy}(0, 1)$$

where

- i indexes observations, $i = 1, ..., 300$;
- M is the number of months, $M = 12$;
- Y_i is the observed count of landfalls for observation i;
- $m(i)$ is a map from i (the observations) to corresponding month ID (integers between 1 and 12);
- $\beta \in \mathbb{R}^{12}$ is the monthly seasonal effect, where β_1 is the effect for month 1 (January), β_2 is the effect for month 2 (February) and so on;
- σ_β scales the variance of β to implement non-centered parameterisation on top of the sum-to-zero constraint, where both constraints were implemented to improve parameter identifiability, sampling efficiency and convergence of our model parameters;
- λ_i is the expected number of landfalls per month and year

Similarly to model A, we ran our MCMC algorithm using 3 parallel Markov chains, 4000 iterations, discarding the first 500 as "warm-up" iterations. We verified model convergence by investigating key model parameter convergence diagnostics, as well as checking the trace plot for our worst performing parameter and looking at pairwise posterior geometries (allowing us to verify that all our parameters are identifiable).

variable	rhat	ess_bulk	ess_tail
beta_sd	1.001771	1910.645	2124.029
beta[4]	1.000347	10610.548	7692.970
beta[3]	1.000231	11440.343	8176.608
beta[11]	1.000684	12001.986	8540.237
beta[9]	1.000049	12156.832	9786.602

Figure 11: Model convergence diagnostics for our 5 worst performing parameters.

Figure 11 displays the 3 statistical efficiency and convergence measures for our worst performing parameters (in terms of the "ess_bulk" measure). As we can see, all "rhat" convergence diagnostic values are less than 1.01, which suggests good mixing and convergence. Moreover both "ess_tail" and "ess_bulk" suggest good sampling efficiency (as they are greater than 500 per Markov chain).

We then checked model fit by performing a posterior predictive check, similarly to Model A.

We found that 95% of our observed landfall counts fell within our model's posterior predictive intervals, suggesting a good model fit. This can be seen in Figure 12.
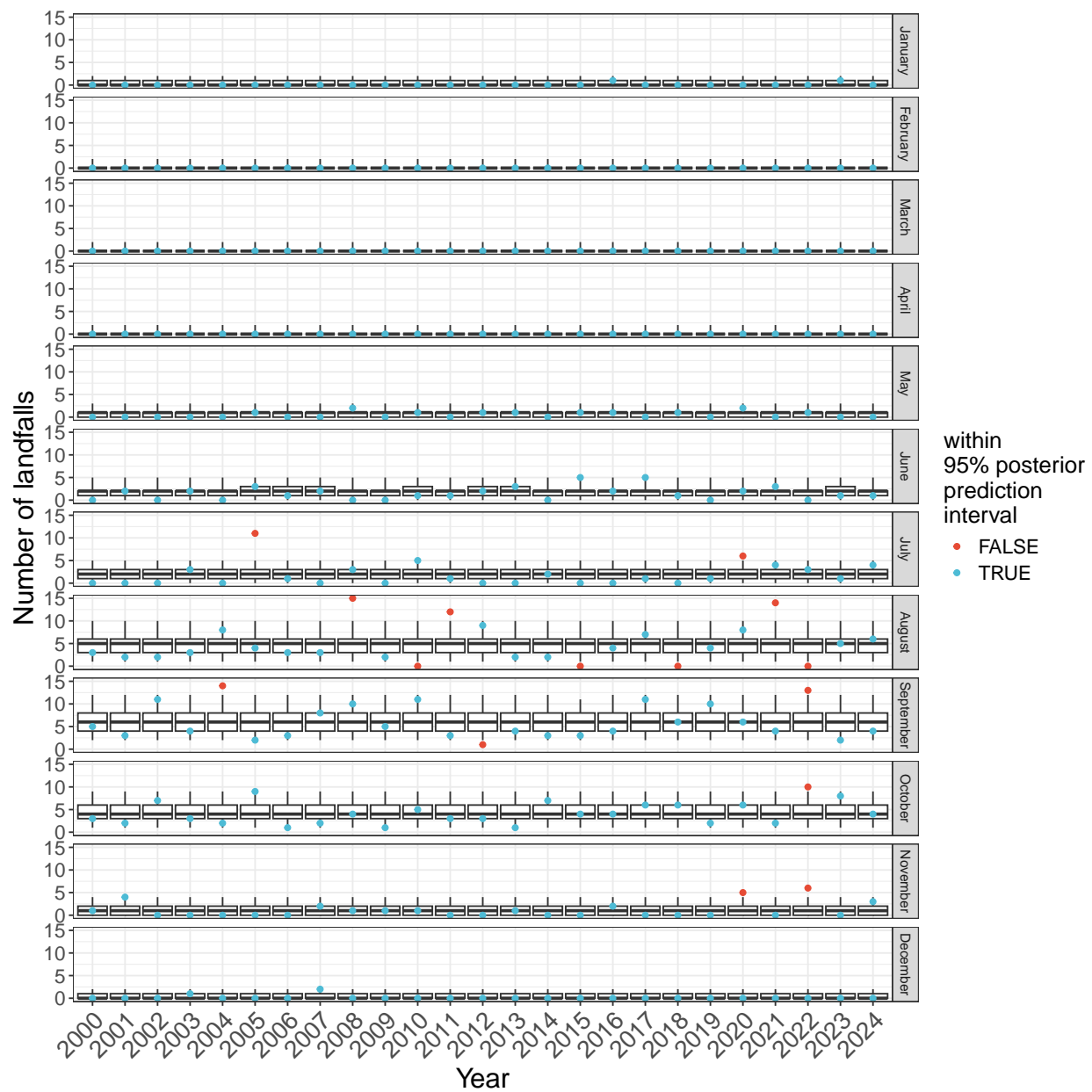
Figure 12: 95% posterior credible intervals and observed data for each month and year, 2000-2024.

8.4 Model A and Model B Comparison

To compare the predictive performance of our models A and B, we computed the expected log posterior predictive density (ELPD) using approximate Leave-One-Out (LOO) Cross Validation.

	elpd_diff	se_diff	elpd_loo	se_elpd_loo
Model_B	0.000000	0.0000	-436.1321	26.24346
Model_A	-3.753091	2.9986	-439.8852	25.76100

Figure 13: LOO Cross Validation measures for comparative model performance.

Figure 13 shows model performance comparison measures such as:

- elpd_diff: Difference in expected log pointwise predictive density (ELPD) between models
- se_diff: Standard error of the difference in ELPD
- elpd_loo: ELPD for leave-one-out cross-validation
- se_elpd_loo: Standard error of the ELPD for leave-one-out cross-validation

The difference in ELPD between Model B and Model A is approximately 3.73, as can be seen in Figure 13, with a standard error of 3.00. This suggests that Model B has a slightly better predictive performance, but overall, both models behave similarly as this difference is marginal. For simplicity, we used Model B to forecast landfalls per month and year.

8.5 Model C: Geographical model

For our final geographical model, our Bayesian Poisson regression model now includes a location (country or territory) effect. This model was then fitted using Markov Chain Monte Carlo (MCMC) methods.

Mathematically, our model is denoted:

$$
\begin{aligned}
Y_i &\sim \text{Poisson}(\lambda_i) \\
\log \lambda_i &= \beta_0 + X\beta \\
\beta_0 &\sim \text{Normal}(0, 2^2) \\
\sigma_\beta &\sim \text{Half-Cauchy}(0, 1) \\
\beta &\sim \text{MVN}(0, \frac{L}{L-1}(I - \frac{1}{L}\mathbf{1}\mathbf{1}^T) \cdot \sigma_\beta^2), \text{ (sum-to-zero parameterisation)}
\end{aligned}
$$

where

- i indexes observations, $i = 1, ..., 250$;
- j indexes the countries or territories, $j = 1, ..., 10$;
- Y_i is the observed count of landfalls for observation i;
- X is a 300×10 dimensional design matrix consisting of one-hot columns for each location;
- β_0 is the baseline parameter;
- $\beta \in \mathbb{R}^{10}$ is the monthly seasonal effect, where β_1 is the effect for country/location 1, β_2 is the effect for country/territory 2 and so on;
- σ_β scales the variance of β to implement non-centered parameterisation on top of the sum-to-zero constraint, where both constraints were implemented to improve parameter identifiability, sampling efficiency and convergence of our model parameters;

- λ_i is the expected number of landfalls per month and year

We ran our MCMC algorithm using 3 parallel Markov chains, 6000 iterations, discarding the first 500 as "warm-up" iterations. We verified model convergence by investigating key model parameter convergence diagnostics, as well as checking the trace plot for our worst performing parameter and looking at pairwise posterior geometries.

variable	rhat	ess_bulk	ess_tail
beta_sd	1.0015133	2791.417	4290.671
beta[1]	0.9999457	13655.243	11731.427
beta_0	1.0002044	14036.765	11101.066
beta[7]	0.9999560	15589.426	13743.762
beta[10]	1.0001697	16408.876	14790.375

Figure 14: Model convergence diagnostics for our 5 worst performing parameters.

Figure 14 displays the 3 statistical efficiency and convergence measures for our worst performing parameters (in terms of the "ess_bulk" measure). As we can see, all "rhat" convergence diagnostic values are less than 1.01, which suggests good mixing and convergence. Moreover both "ess_tail" and "ess_bulk" are above 1500, suggesting good sampling efficiency.

We then checked model fit by performing a posterior predictive check, similarly to Models A and B.

We found that 96.8% of our observed landfall counts fell within our model's posterior predictive intervals, suggesting a good model fit. This can be seen in Figure 15.
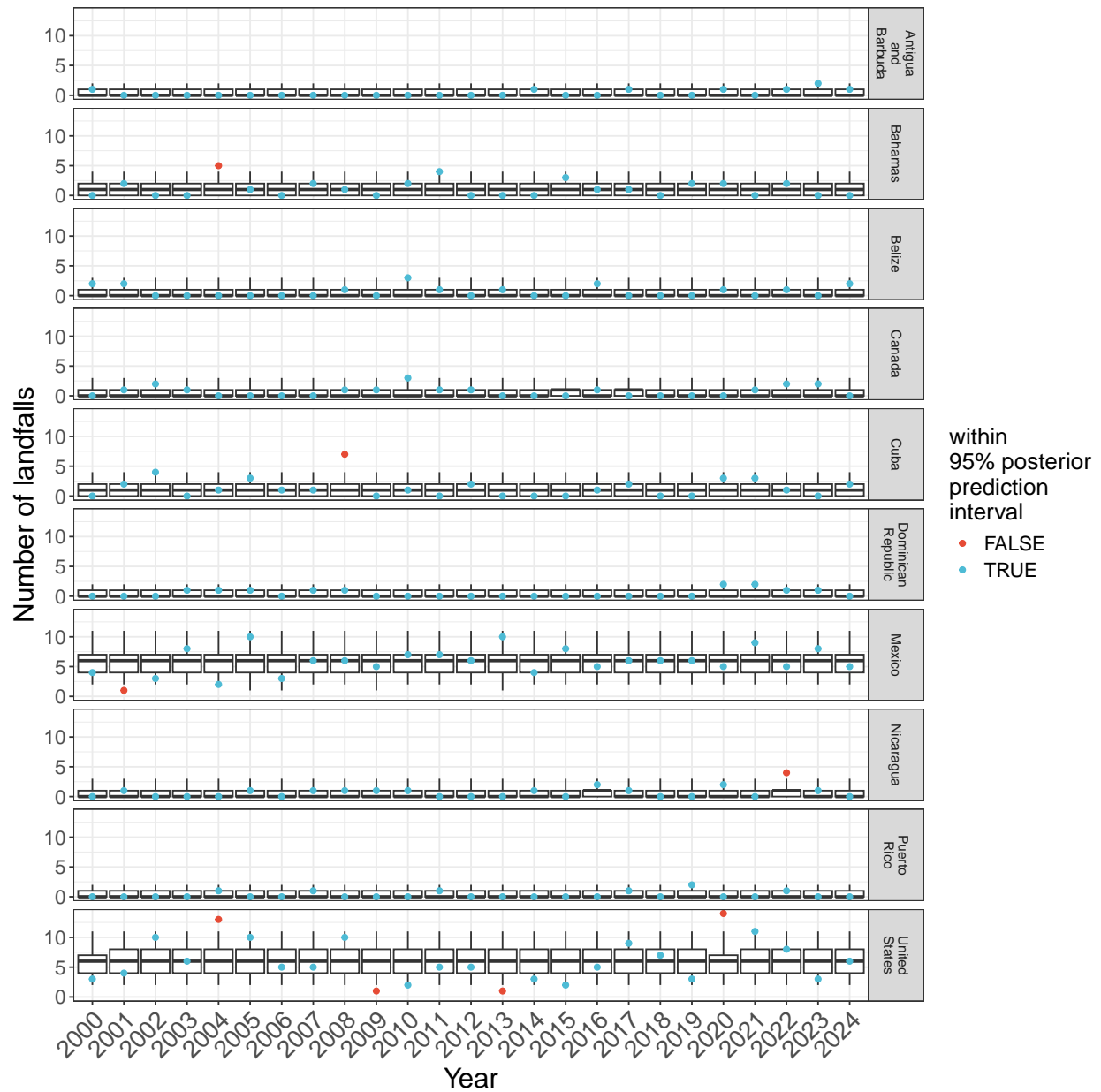
Figure 15: 95% posterior credible intervals and observed data for each month and year, 2000-2024.