12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12
Vancouver, Canada, July 12-15, 2015

Using Random Simulation of Hurricane Tracks for Risk Analysis

Luca Garrè
Senior Researcher, Strategic Research and Innovation, DNV GL, Høvik, Norway

Agnes Dudek
Principal Researcher, Strategic Research and Innovation, DNV GL, Høvik, Norway

ABSTRACT: This paper illustrates a simulation approach to hurricane tracks. The procedure implemented was taken from existing literature and applied to storms recorded in the IBTrACS dataset. The results obtained are encouraging: the variables currently simulated, primarily locations of hurricane genesis, track propagation and locations of decay, are in general agreement with results available in the literature and with the historical data. Further improvements of the present work are however needed with respect to a number of metrics particularly relevant for risk assessment.

North Atlantic tropical cyclones are among the deadliest and most destroying weather hazards in coastal regions of Central and Northern America. The devastating impacts that hurricane-originated high winds and storm surges have on both offshore and onshore installations, as well as cities and communities, have gained this weather phenomenon ample coverage in a number of different scientific fora, including the weather and climatic, environmental, engineering and actuarial communities. Owing to this, the availability of approaches to the modelling, and inclusion in risk analysis, of hurricanes effects is plentiful and hard to cover in full. Representative work showcasing hurricane modelling within a risk analytical context can be found, among many others, in Lin et al. (2010).

Records of past hurricanes and of the corresponding damages demonstrate that impacts are highly dependent on general features such as velocity and radius of maximum wind speed, pressure drop, speed of advancement, location of landfall and the overall geographical features at that location. Hence, from a risk analytical perspective, this entire set of parameters needs to be considered in the impact assessment, Vickery et al. (2000).

Due to the nonlinearities characterizing the genesis and development of hurricanes, along with the complex interplay taking place with the surrounding climatic boundary conditions, hurricanes are highly unpredictable; because of this, the abovementioned parameters are affected by a number of uncertainties which need to be accounted for in the risk assessment. To this aim, a number of probabilistic models for hurricane simulation have been introduced in the dedicated literature, see for instance Vickery et al (2007) among others. Whilst these various probabilistic models possess different features, they all share the same common goal: characterizing the overall uncertainty, with reference to the statistical uncertainty of genesis, propagation, intensity, landfall location and lysis (decay) of the storm.

Besides the suitability of a random simulation approach for modelling the inherent randomness of hurricanes' behavior, a further advantage of this setup is the ability to estimate the tail behavior of variables of interest. One exemplification can be made looking at extreme wind speeds over a hurricane prone area. Repetitive simulations of hurricanes can be used to generate large samples of hurricane wind fields, upon which one can perform extreme value analysis. The caveat in this respect is that any bias in the parameters' simulation would translate in possibly larger biases in the tails.

This paper addresses the probabilistic modeling of main features of tracks of hurricanes, i.e., the location of genesis, the development and the final lysis location. Procedurally, the work presented here follows somewhat slavishly the simulation procedure developed and published earlier by Hall and Jewson (2007). Even with minor differences in some aspects, the analysis does not represent an improvement of the referenced work. Owing to this, this paper does not advance random simulation of hurricanes; rather, it motivates such approach in a risk-analytical perspective.

Future efforts will look at the inclusion of wind field simulation into the present modelling framework, along with coupling to storm surge models for impact assessment on coastal and near-coastal installations.

1. THE DATASET

Records of past hurricane tracks in the North Atlantic were used. These were obtained from the IBTrACS dataset, see Knapp et al. (2010), for the years 1950 to 2003. Of these, tropical storms (TS) in the North Atlantic basin were selected, resulting in a sample of 693 hurricanes. Observations of hurricanes are encumbered with uncertainty associated with the accuracy and coverage of the tracking systems employed. However, IBTrACS summarizes estimates of tracks from 12 different agencies worldwide and ensures rigorous quality checks of the data provided.

2. THE MODEL

As anticipated, the simulation approach is based on the steps and procedures developed and illustrated in Hall and Jewson (2007). The routines are implemented in Python, van Rossum (1995), and are organized according to three basic modules:

1. random sampling of the genesis location across the North Atlantic basin;
2. random propagation of the hurricane from the genesis location and from successive points;
3. random lysis of the hurricane;

2.1. Genesis location

The location of genesis identifies the geographical site where the hurricane is originated. This location is sampled randomly through sum of bivariate normal kernels using the sklearn Python implementation, Pedregosa et al. (2011). The optimal bandwidth is derived from the recorded historical genesis points via 20-fold cross validation provided in the sklearn implementation. The so-obtained genesis probability density is plotted in Figure 1, together with the observed genesis locations as recorded in the dataset, represented by black dots. As can be seen from the figure, the spatial features of the genesis probability are remarkable and highlight the existence of regions of the basin which, due to the local climatic features, are more prone to the generation of hurricanes. These regions are located off the West African coast and in the Gulf Of Mexico, plus areas off Florida and Nicaragua. It is noted that these estimates are well aligned with similar results available in the literature.

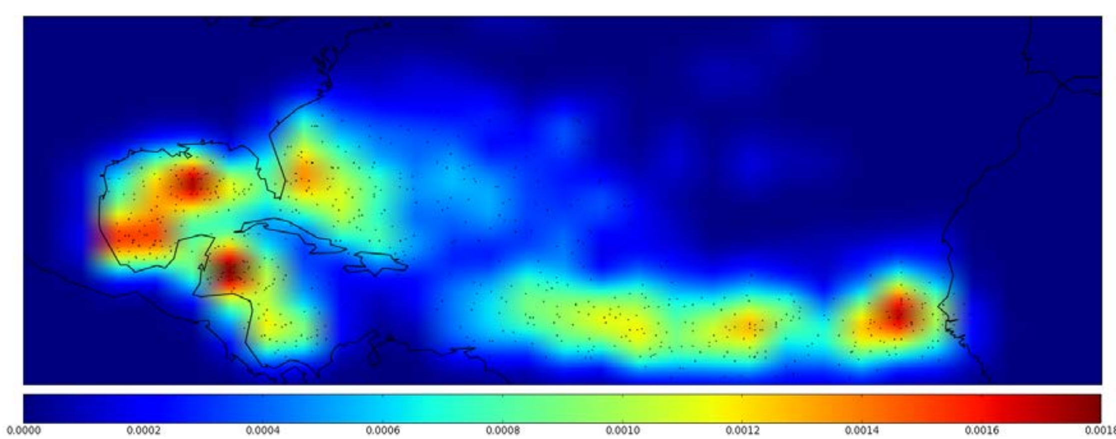

Figure 1: Genesis pdf along with historical genesis points

2.2. Propagation

The track can be thought of as a series of consecutive track points, each defined as a pair of latitude and longitude, which originates from the genesis point and progress until the hurricane dies out. Propagation is obtained via simulation of six-hourly displacements of the hurricane, $\boldsymbol{d} = \{d_{\mathrm{z}}, d_{\mathrm{m}}\}$, from one point to the next, where the scalars d_{z} and d_{m} are the zonal (east-west direction) and meridional (south-north direction) displacements, respectively.

As for genesis, statistics of six-hourly displacements show relevant spatial features, meaning that this variable depends to a large degree on the location of the storm. This spatiality is embedded in the simulation procedure in the following manner. First, the hurricane is found at a given location $t = \{lon_t, lat_t\}$. This could be either the genesis location or an intermediate track point. For this location, the average displacement vector $\bar{\boldsymbol{d}}^t = \{\bar{d}_z^t, \bar{d}_m^t\}$ is computed as (same for $\bar{d}_m^t$):

$$\bar{d}_z^t = \left(\sum_j d_{zj} e^{-l_j^2/2b}\right) \Big/ \left(\sum_j e^{-l_j^2/2b}\right) \tag{1}$$

where j indexes all historical six-hourly displacements recorded for all selected storms. Owing to this indexing, d_{zj} (d_{mj}) is the j-th zonal (meridional) displacement in the records, and l_j is the great circle distance between the averaging location t and the record location of d_{zj} (d_{mj}). Following this, the above formula returns the spatial average of historical displacements: each observation is weighted inversely to the distance from the averaging location through the exponential; the bandwidth b quantifies the distance after which observations have limited influence on the average. By virtue of this, the average vector inherits the spatial features of the records, as closer observations are weighted more than distant ones. Implementing the jackknife out-of-sample approach of Hall and Jewson (2007), in the present work the optimal bandwidth has been estimated around 270kms. Figure 2 shows values of $\bar{\mathrm{d}}_z^t$ computed across the modelled basin. The basin is basically divided into two main regions: a southern region (below latitude 25°-27°) characterized by the east-to-west direction of the average zonal displacement (negative values), in contrast to a northern region (below latitude 25°-27°) where the west-to-east direction is predominant. These estimates reproduce well the zonal behavior of hurricanes.

Having computed $\bar{d}^t$, simulation of $\boldsymbol{d}^t$ is carried out as $\boldsymbol{d}^t = \bar{\boldsymbol{d}}^t + \boldsymbol{u}^t$, where $\boldsymbol{u}^t = \{\sigma_u^t \tilde{u}^t, \sigma_v^t \tilde{v}^t\}$ are displacement anomalies parallel and perpendicular to $\bar{\boldsymbol{d}}^t$, respectively. These represent random deviations of each simulated displacement from the average vector. Following Hall and Jewson (2007), these can be characterized as independent, zero-mean, normally distributed, and lag-one autocorrelated variables, with $\tilde{u}$ and $\tilde{v}$ being standard normal variables. The mutual independence allows the simulation of the two anomalies separately.

The standard deviations σ_u^t and σ_v^t are computed through an averaging functional which resembles that of Eq. (1). In practical terms, the terms d_{zj} of Eq. (1) are substituted by u_j^2 and v_j^2, which are the squares of the historical anomalies (parallel and perpendicular) of each j-th recorded displacement from the average displacement vectors $\bar{\boldsymbol{d}}^t$ computed at each j-th record location. Square root returns the standard deviations. The bandwidth value, estimated employing an out-of-sample jackknife approach which follows that of the original work, is again equal to 270kms.

Normality and lag-one autocorrelation motivate the following lag-one autoregressive model (same structure for $\tilde{v}^t$):

$$\tilde{u}^t = \phi_u^t \tilde{u}^{t-1} + \sqrt{1 - {\phi_u^t}^2}\, \epsilon_u^t \tag{2}$$

where ϕ_u^t is the lag-one autocorrelation coefficient at location t, $\tilde{u}^{t-1}$ is the anomaly of the previous displacement, ϵ_u^t is a standard normal innovation and the square root returns the magnitude of the innovation, Hall and Jewson (2007).

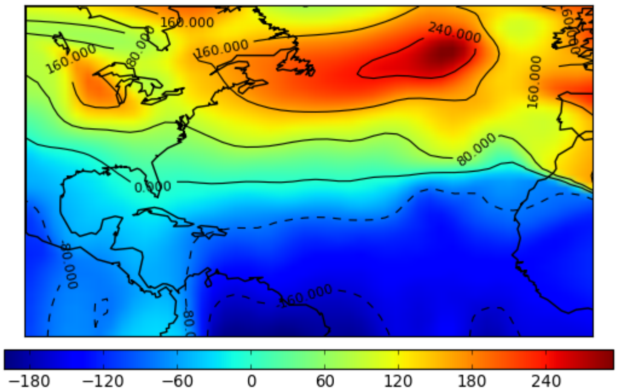

Figure 2: Zonal component of the average displacement vector

Further to this, the autocorrelation coefficients ϕ_u^t are computed as (same for ϕ_v^t):

$$\phi_u^t = \frac{\sum_j (u/\sigma_u)_j (u/\sigma_u)_{j-1} e^{-l_j^2/2b}}{\sum_j e^{-l_j^2/2b}} \tag{3}$$

where the terms $(u/\sigma_u)_j$ are the ratio between the historical anomaly of the recorded j-th displacement – computed from the average displacement vector $\overline{\boldsymbol{d}}^t$ at the j-th record location – and the standard deviation of the anomaly, computed at the same location. The optimal bandwidth was estimated to be approximately 800kms. A final remark goes to the innovation terms ϵ_u^t , as these should in principle be normally distributed. Nevertheless, Hall and Jewson (2007) found that residuals r obtained using the model in Eq. (2) show slightly fatter tails than the normal distribution would feature. Hence, tables of residuals are generated from the historical data and random draws from these tables provide the innovations. The present work follows the same strategy, with residuals defined as $r^t = (u^t/\sigma_u^t - \phi_u^t u^{t-1}/\sigma_u^{t-1})/\sqrt{1-{\phi_u^t}^2}$.

By means of Eq. (2) the parallel and perpendicular random anomalies are computed at each progressive track point. These are then scaled with the corresponding standard deviations and finally added to the mean vector for that point. The hurricane is then moved by the obtained displacement and the procedure repeated. At the genesis point there is no previous anomaly against which Eq. (2) can be regressed and propagation is initiated taking $\tilde{u}^t = \epsilon_u^t$.

2.3. Lysis

Lysis defines the decay of the storm. This occurs when the storm's wind speeds decrease to 63 km/hour or lower. The probability of hurricane lysis p_l at a given location t is given as:

$$p_l^t = \left(\sum_j \delta_j e^{-l_j^2/2b}\right) \Big/ \left(\sum_j e^{-l_j^2/2b}\right)$$

where the delta-kronecker δ is equal to one for reported lysis sites and zero everywhere else. Again, the index j steps through all historical points. Figure 3 shows p_l computed for the whole basin, along with the historical lysis sites. The spatial structure of p_l follows the coastal lines especially for the Northern American continent and Europe, meaning that as hurricanes go ashore lysis is more likely. The same structure is seen for the Northern Atlantic ocean. These well-known trends are due to the unfavorable conditions hurricanes meet over land and cold waters which lead to their decay. The lysis module is applied repetitively at each track point: lysis occurs if a random draw from a standard uniform distribution is lower than p_l^t.

3. RESULTS

This section presents some metrics for quality evaluation of the model which address a number of variables of interest within hurricane impact assessment.

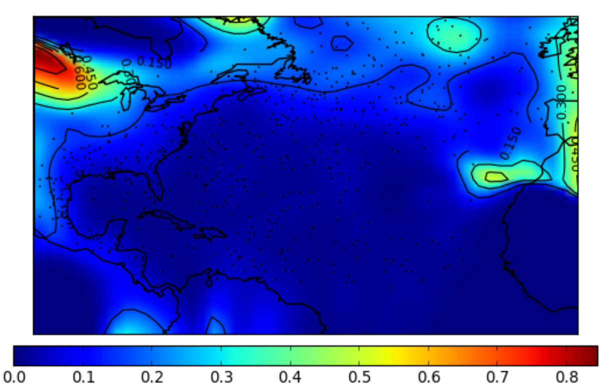

Figure 3: The lysis probability

The results were obtained from an ensemble of 20 runs; for each run the same number of storms from the IBTrACS dataset was simulated, leading to nearly 14000 simulated storms in total. Furthermore, the genesis module was turned off for these simulations and tracks were simply started at the historical genesis locations.

3.1. Track crossings at different longitudes

The first metric evaluated is the number of crossings of the tracks at different longitudes. In essence, nine north-to-south lines across the North Atlantic were selected, located 10° longitude apart, between 260° and 340° longitude. For each line, the tracks crossing that line were stored, together with the latitudes at which the crossings occurred. Hurricanes in the North Atlantic typically cross such north-to-south lines in both directions (east-to-west and west-to-east). Hence, for each line, the crossings ought to be differentiated according to the direction and evaluated separately.

Figure 5 – end of paper – shows the results. For each longitude considered two plots are provided: the left one collects the east-to-west crossings, whereas the one to the right collects the west-to-east crossings. Each plot has latitudes on the x-axis (0° to 90°, marked every 10°), and number of crossing on the y-axis (0 to 100, marked every 20). The red line represents the number of crossings featured in the historical dataset, whereas the grey shaded area encompasses the central 95% confidence region from the simulations. The red line is contained within the grey shaded regions across most of the latitudes for all plots. Some discernible deviations are seen in the central part of the basin (280° and 290°), for some latitudes. This overall agreement is comforting, as it implies that, for this metric, the historical observations can be regarded as one realization of the simulation process for most of the basin.

3.2. Track crossings at different latitudes

A similar metric as for the longitude crossings can be made looking at the latitude crossings instead. The same principle applies throughout. The east-to-west lines are placed 10° latitude apart, from latitude 10° to 50°. For each latitude two plots are generated, see Figure 6 at the end of the paper, one for south-to-north crossings (to the left) and one for north-to-south crossings (to the right). Longitudes of crossing are given on the x-axis, (from 200° to 360°, marked every 20°) while crossing counts are given on the y-axis (0 to 60, marked every 10). Again, a good agreement is found between the historical records and the model. Further, the majority of tracks present only a south-to-north component, whereas only spurious observations and simulations feature the opposite direction.

3.3. Track point density

Track point density measures the number of track points within a given region of the basin. This metric reflects the average length of observed and simulated six-hourly displacements and, by virtue of this, the average historical and simulated speed of the hurricanes. This is a critical aspect as wind-induced loads are scaled with the square of wind speed.

Following Hall and Jewson (2007), Figure 4 presents the so-called Z statistics of the track points. This statistics is obtained as the difference between the number of track points per area in the dataset minus the average number of track points per area from the 20 ensemble runs, divided by the standard deviation of the number of track points per area, again computed from the 20 ensemble runs. For the generations of the Z statistics the basin was divided into 3° by 3° squares and track points counted for each square. Regions where the Z statistics is lower (higher) than –(+)1.96 represent deviations which are statistically relevant at the 5% confidence level. These are highlighted in the figure in red and blue, respectively. The red regions identify excess of simulated track points compared to historical, and viceversa for the blue regions. As can be seen, besides some spurious regions, there is one remarkably large region, off Cape Hatteras, where the model underestimates the records. For the rest of the basin the model shows no statistically relevant deviations.

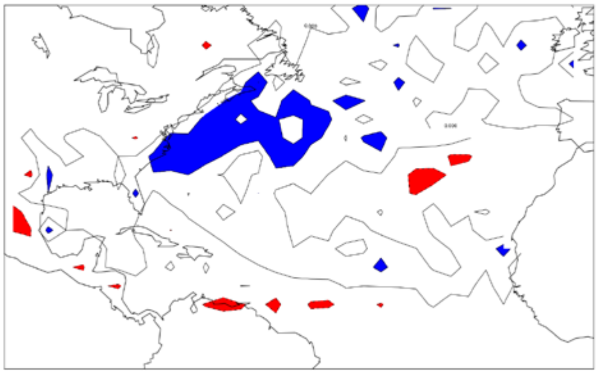

Figure 4: Z statistics for track point density

3.4. Location of landfall

Another obvious variable for hurricane impact assessment is location of landfall, which usually corresponds to the site of highest damages. Figure 7 presents a comparison similar to that carried out for track crossings, but now for the number of landfalls across the coastline of Northern and Central America. Again, the red line represents the historical statistics, whereas the grey shaded area encompasses the 95% central region of the ensemble runs. The x-axis represents distances along the coastline, whereas the y-axis contains the number of landfall (0 to 40, marked every 10). Locations of known cities and sites are reported for reference. A good accuracy is seen across the entire shoreline, with a few, notable deviations in correspondence of Cape Hatteras, Miami and New Orleans.

4. CONCLUSIONS

This paper illustrated the main results obtained with a simulation procedure for hurricane tracks which was developed in Hall and Jewson (2007). The work reproduced the modules of the procedure therein presented and illustrated some preliminary results. This is the first step towards a more comprehensive model, where additional elements of hurricanes will also be included.

The results obtained are encouraging, as all variables currently simulated, primarily locations of hurricane genesis, track propagation and locations of lysis, are in general agreement with similar results available in the literature and the historical data. Notwithstanding this, further improvements are still needed, especially with respect to the density of track points and landfall location. It is noted that the performance featured in the original work appears higher, as the same statistical tests were passed with stricter specifications (deviations higher than one standard deviation from the ensemble average were deemed significant, whereas here the 2.5% and 97.5% quantiles were used). This could be due to various reasons, including the different dataset or different implementation employed.

The approach is purely statistical: is based on records and their use for the derivation of the required spatial statistics and it lacks physically-based components for the incorporation of the dynamics of hurricanes. Notwithstanding this, the model can be used for fast simulation of thousands of hurricanes tracks whenever their inherent variability needs to be accounted for in the risk assessment.

5. REFERENCES

Hall, T.M., and Jewson, S. (2007). “Statistical modelling of North Atlantic tropical cyclone tracks”. *Tellus*, 59A, 486 – 498.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., and Neumann, C.J. (2010). “The International Best Track Archive for Climate Stewardship (IBTRaCS) – Unifying Tropical Cyclone Data”. *Bull. Amer. Meteor. Soc.*, 91, 363 – 376.

Lin, N., Emanuel, K.A., Smith, J.A., and Vanmarcke, E. (2010). “Risk assessment of hurricane storm surge for New York City”. *Journal of Geophysical Research*, 115, D18121.

Pedregosa, F., et al. (2011). “Scikit-learn: machine learning in Python”. *Journal of Machine Learning Research*, 12, 2825 - 2830.

van Rossum, G. (1995). “Python tutorial”. *Technical Report CS-R9526*. Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Vickery, P.J., Skerlj, P.F., and Twisdale, L.A. (2000). “Simulation of hurricane risk in the U.S. using empirical track model”. *Journal of structural engineering*, 126, 1222 – 1237.

Vickery, P.J., Masters, F.J., Powell, M.D., Wadhera, D. (2007). “Hurricane hazard modeling: the past, present and future”. *Proceedings of the ICWE12 Conference*, Cairns, Australia.

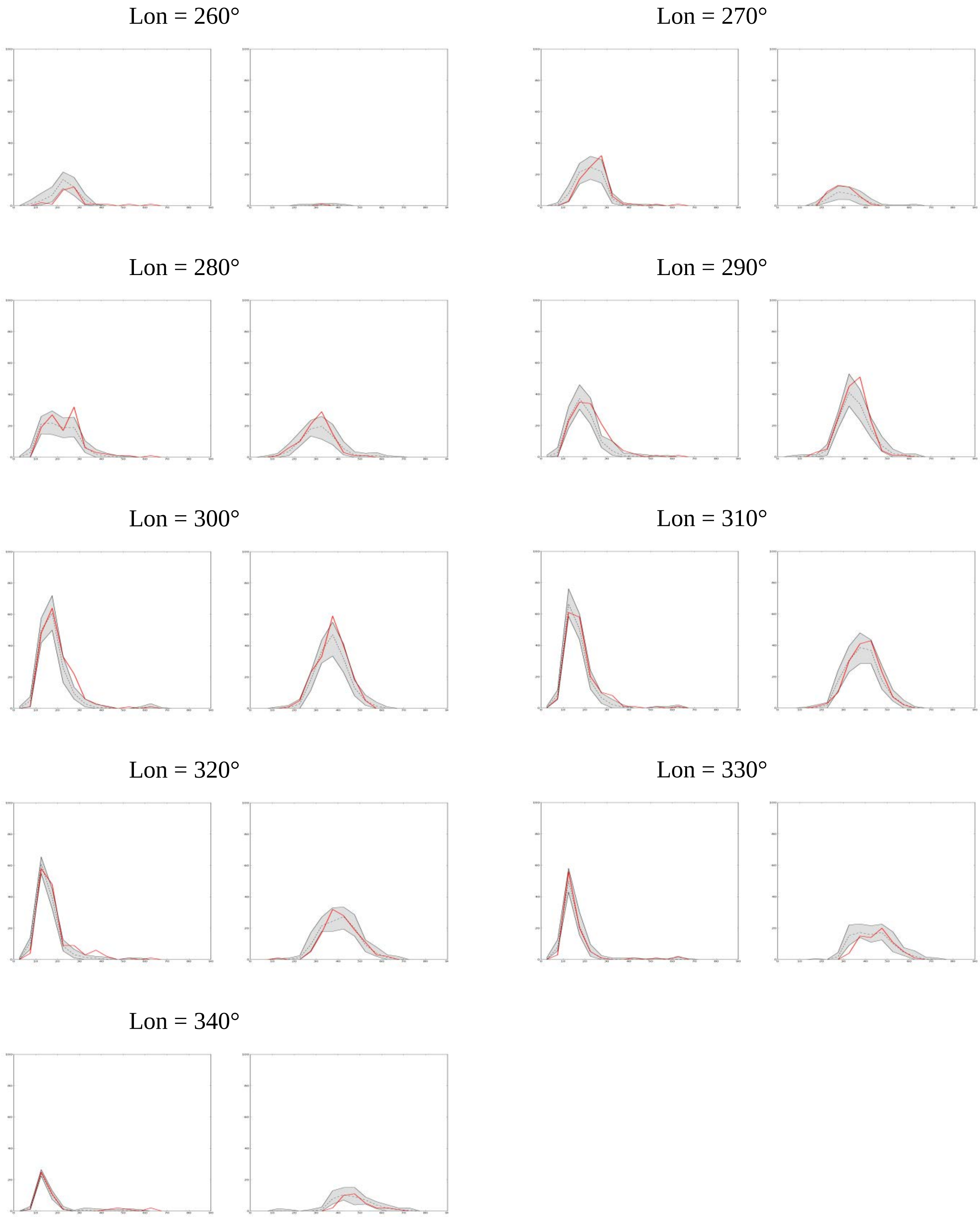

Figure 5: Track crossings at different longitudes

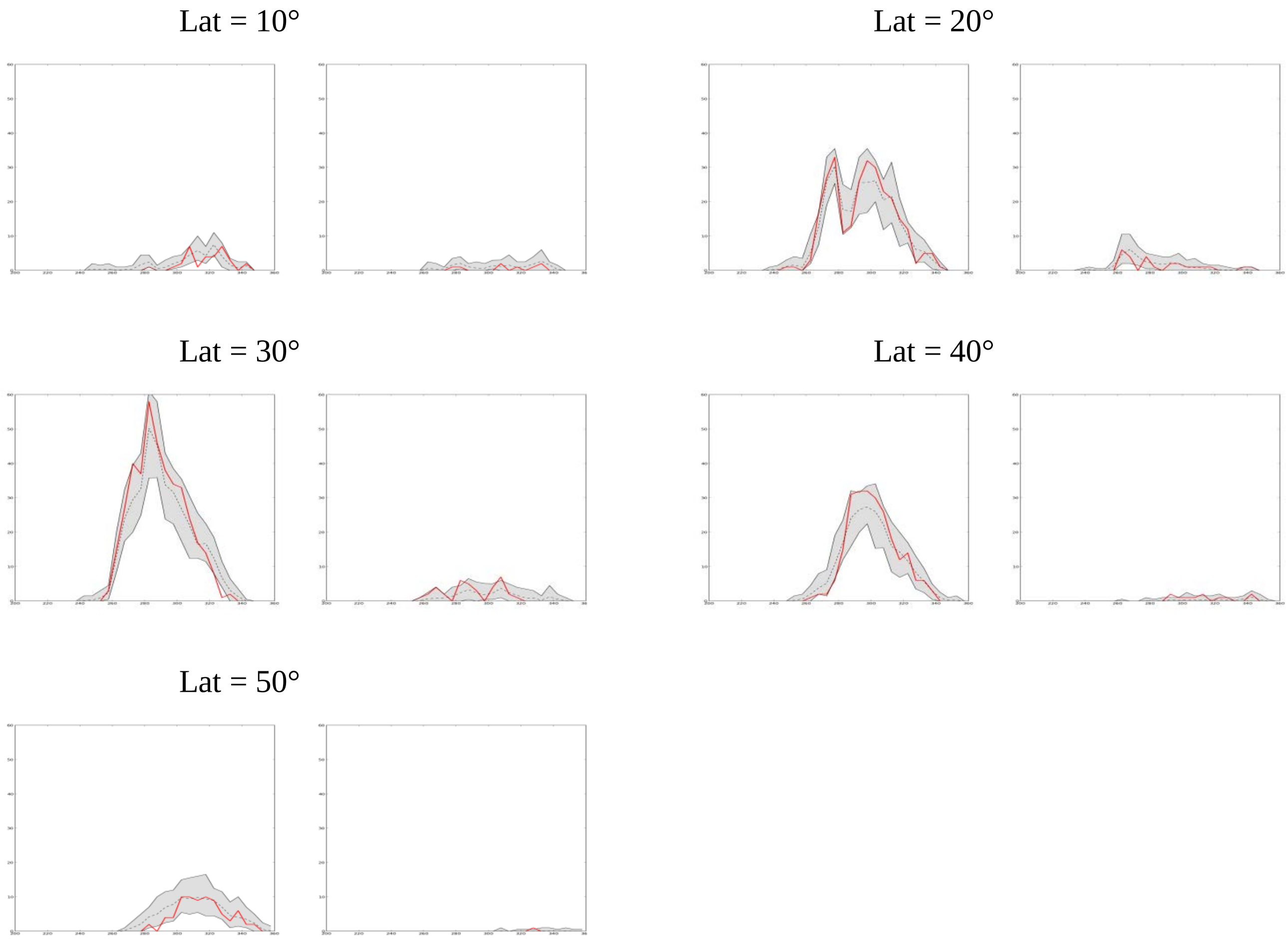

Figure 6: Track crossings at different latitudes

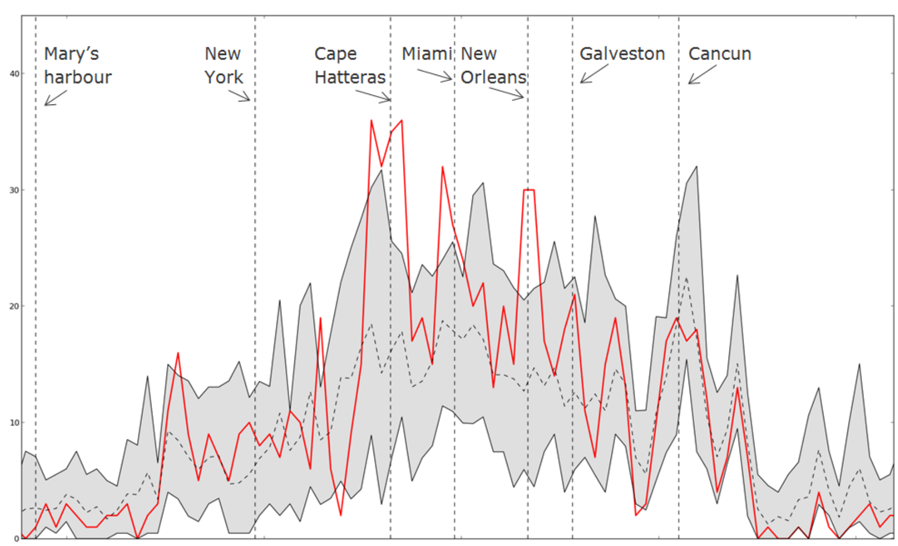

Figure 7: Landfall counts across the Central and North American continental coastline