Stochastic and Statistical Modelling of Extreme Meteorological Events: Tropical Cyclones

MRes Thesis

CID: 01022787

UoR ID: 23022033

Thomas Patrick Leahy

Prof. Ralf Toumi, Dr Colin Cotter and Dr Axel Gandy.

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2BZ

Thesis submitted as part of the requirements for the award of the

MRes in Mathematics of Planet Earth

Imperial College London and University of Reading

AUGUST 27, 2015

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: 

If we knew what it was we were doing, it would not be called research, would it?

— Albert Einstein

Abstract

Tropical cyclones are devastating destructive forces of nature that can cause loss of life and catastrophic damage. This project implements a stochastic tropical cyclone model for the simulation of tracks in Southeast Asia developed in Rumph et al. (2007, Math Meth Oper Res), evaluates the model by using key statistical figures and applies the model to hazard assessment. A large part of this project is the implementation of this model and the interpretation of the paper. Where the paper seemed somewhat ambiguous, sensible and justified decisions were made. Landfall rates, clash probabilities and return periods are all considered and calculated. The model is run and comparisons are drawn under ENSO conditions. There are several conclusions and suggestions from this work. One such conclusion is that return periods are sensitive to the wind profiles used in their calculation, this may have an impact on how return periods are calculated in future. Short falls and assumptions of the model are criticised and suggestions of further work is given in the summary.

Acknowledgements

I would like to thank my supervisors, Ralf Toumi, Colin Cotter and Axel Gandy who's help and support have guided me through the process. Funding from the Engineering and Physical Sciences Research Council (EPSRC) and Climate KIC. I would also like to sincerely thank the staff of MPE CDT, the stochastic group in Physics whose help has been invaluable, and the MPE CDT cohort 2014 for their help and support.
Finally I would like to thank my friends and family for their support and listening to my complaints!

Thomas Patrick Leahy

Table of contents

Chapter 1

Introduction

Catastrophe models are black boxes. Very few people know how the calculations are performed and how the results are determined. These models attempt to estimate the losses to an insurer that could be sustained as a result of a catastrophic meteorological or seismic event such as a tropical cyclone. However, more generally these models can analyse both natural catastrophes (hurricanes, earthquakes etc.) and human catastrophes (terrorist attacks, warfare, nuclear, etc.). Catastrophe models are invaluable tools for use in actuarial science when conducting risk analysis for insurance purposes. The output from these models are used by insurers and re-insurers to make premium and reserve calculations. The reserves set aside are crucial for the reconstruction of housing and infrastructure damaged and destroyed by such extreme events. Due to the power and sophisticated nature of these models, they are very expensive for insurers to lease.

Catastrophe models (cat models) generally consist of four major components, with additional input considerations coming from insurance specific data. The four main components are the event generation module, the local intensity module, the damage module and the insured loss module. The event generation module attempts to determine the frequency and magnitude of potential catastrophic events by geographical location [1]. Detailed data analysis on available data and also a deep understanding of the physical characteristics for natural catastrophe models is required to generate accurate probability distributions for the modelled variables. The local intensity module stochastically propagates the generated events across the simulated potential hazard zone. This module takes into account several factors including historical data, topography, theoretical and physical relationships between modelled variables, etc.. The damage module uses the information from the local intensity module and exposure data as well as local planning and building information to calculate the potential impact to infrastructure. Mathematical functions called 'damageability relationships', describe the interaction between buildings, both their structural and non-structural components as well as their contents, and the local intensity to which they are exposed [1]. The final module of a cat model calculates insured losses. The insured losses are calculated by considering the policy con-

ditions with the total damage estimates calculated from the damage module. Policy conditions may include deductibles by coverage, site-specific or blanket deductibles, coverage limits and sublimits, loss triggers, among others [1].

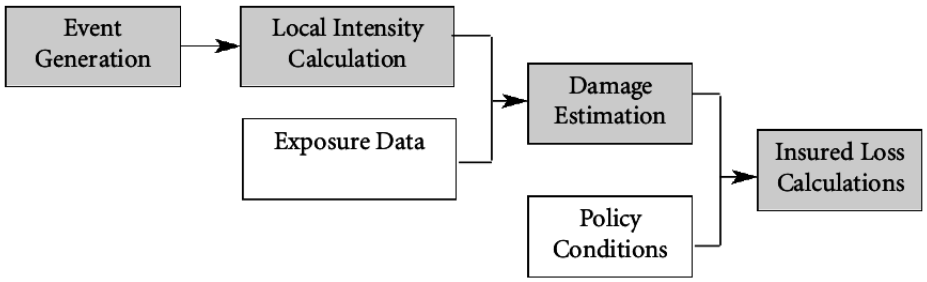


Figure 1.1: Catastrophe Model Components (in grey)[1]

Tropical cyclones are extreme forces of nature that have the potential to cause vast amounts of destruction and devastation. One of the most costly tropical cyclones was the category five* Typhoon Haiyan that caused major destruction and loss of life in the five days that it swept across four different countries. Typhoon Hayian affected the Philippines, Vietnam, China and Taiwan causing 6334 fatalities and overall financial losses of $\$10.5$bn of which $\$700$m was insured†. This was a catastrophic event with the majority of the damage and loss of life coming from the severe winds and storm surge. Aside from the lamentable loss of life, a striking financial figure emerges, approximately 93.33% of overall losses were uninsured. When a catastrophic event occurs, the extent of physical damage to private and commercial property and national infrastructure determines the total economical losses, a large share of which is typically uninsured, as shown by Typhoon Hayian. The insured losses, must be honoured by the global insurance market. This is one reason why cat models are vital. By simulating potential events and the various effects that they have on risk levels that are underwritten by insurance companies, the insurer is provided with loss statistics and a level of confidence of exceeding a given threshold of loss levels. The insurance company can then take appropriate action, for example purchasing reinsurance cover and/or mitigating risk by reducing its exposure to certain classes in its portfolio or introduce catastrophe peril deductibles on insurance contracts to minimise the Probable Maximum Loss (PML) potential.

This project implements a stochastic model, developed by [20] to generate and simulate synthetic tropical cyclone tracks over the western North Pacific and South-East Asia. Where the paper describing the model seemed somewhat ambiguous, sensible and justified decisions were made in the implementation. There are several stages of the modelling process in the tropical cyclone track model that are considered, such as modelling the genesis locations, intensity estimation, propaga-

*Saffir-Simpson Hurricane Wind Scale

†Munich Re, NatCatSERVICE, 2015

tion of the cyclone tracks and a termination model. Individual functions in the code are defined to implement each part of the model and details of the implementation are given in Chapter 4. The historical observations of each track are classified depending on their track shapes, this is discussed in detail in Chapter 3 along with the source of the data and the definition of the observation window. The output of this model will be a large number of synthetic tropical cyclone tracks, which will provide a large dataset for the assessment of risk. Once the model has been implemented, the project then evaluates the implemented model by assessing mean landfalls rates and their variability at specific locations of interest in the observation window. The project then applies the synthetic track model to hazard assessment. In this part of the project, quantities of financial interest are calculated. Clash probabilities, i.e. the probability of a tropical cyclone affecting two locations along its path are calculated as well as return periods for specific locations. In order to calculate return periods, a model of the wind profile of tropical cyclones is required. An investigation into wind profile models is conducted and the wind profile models are applied and comparisons are drawn. Further, the hazard assessment analysis is conducted again when the model is run under ENSO conditions and again differences are highlighted.

The model is implemented using the Python programming language with some use of Excel and R for data analysis. Several Python packages are used including `Numpy` and `SciPy`.

There are many tools used to enhance and elucidate the understanding and management of risk of natural and human catastrophes. Newly available loss and meteorological data, the continuously expanding knowledge of the science of natural hazards, and improvements in computing speeds and capacity all contribute to the improvement and sophistication of catastrophe models and the development of catastrophe modelling. Note that ‘cat’ may be used as a replacement for catastrophic in this thesis.

Chapter 2

Literature Survey

There has been a long tradition of applying statistics of extremes to geoscience with a particular emphasis to meteorology and climate. The German mathematician, Emil Julius Gumbel was one of the first to investigate this connection in his paper "On the frequency distribution of extreme values in meteorological data" [9]. Due to the vast amounts of data available that is becoming available in the field of meteorology and climate, particularly since the introduction of satellites, the field has become a very fertile area of application of statistics and extreme value theory.

This short literature survey will review and analyse the various approaches of the use of statistical and stochastic models applied to the generation of synthetic tropical cyclone tracks. It will discuss the different approaches that have been put forward, from the modelling of the initial genesis points through to termination of the synthetic tracks. This review will highlight and build upon a literature review previously conducted, see [17].

In general, most models of this type are statistically driven as they are much less computationally expensive and make use of available data, whereas General Circulation Models are based on the laws of physics. GCMs often have difficulty capturing the complexity of certain atmospheric processes on sub-grid scales and often these processes are coarsely parametrised. There is a significant amount of research in understanding the effects of these sub-grid processes and a realistic method of representing them in GCMs, see [18], [4]. Stochastic cyclone models have become a widely accepted approach for the estimation of risk. Since the first application of stochastic cyclone models to analyse cyclone wind there have been many further models put forward, ([8], [29], [15], [6], [10], [20], [21]).

Tropical Cyclone Track models

Initialisation

There are is wide variety of approaches to modelling the cyclogenesis. In [29], they sample directly for the historical genesis sites in their data set. An interpolation of historical genesis points is applied in [15]. The simulated tropical cyclogenesis in [6] construct a spatial-temporal dependent probability density function (pdf) of genesis locations based upon historical genesis events. In this approach the pdfs are constructed by binning historical genesis events and applying variable-Gaussian smoothing, which is described in detail in the supplement to the paper. Alternatively [20], [21] use a nearest neighbour and approach and develop a genesis kernel pdf based on the underlying data. A similar approach is taken in [10], to make maximum use out of the available historical data.

Track Propagation

Track propagation is the next step in the modelling process. This step generates polygonal segments that compose the tropical cyclone tracks. Again there are a variety of different approaches. Many of the models put forward make the same basic assumption that cyclones located in 'similar' areas of the observation window behave similarly, see [20], [21],[10],[6],[29]. In both models put forward in [6], although different variables are considered and in the models described in [20], [21], the stochastic cyclone models develop the tracks via a Markov chain* by sampling from local pdfs generated from historical data, for both models presented by [20], [21], the variable considered as a Markov chain is the change in wind speed. Alternatively [10], [29] and [15] opt for autoregressive (AR) models to represent the propagation of the synthetic tropical cyclone tracks. For example in [15], the AR model consists of a time series of latitude, longitude and of central pressure. In [10] means and variances of latitudinal and longitudinal displacements and the remaining anomalies are modelled as autoregressive.

Termination

An integral part of models generating synthetic tropical cyclone tracks is the lysis of the tracks. In [29], once a simulated storm makes landfall, the reduction in central pressure with time is modelled using the "filling models" described in [28]. If the storm proceeds to return back over open water where there is potential for the storm to regain energy, the model returns to the propagation method. In the first model proposed in [6], the tracks are terminated using two criteria; the first is a termination pdf constructed in a similar manner to the genesis pdf from the historical data, and the second criteria is when searches at multiple space-time resolutions fail to provide evidence for a transition [6]. The second model proposed does not allow transitions into regions where generating statistics become inferior, so a simpler termination procedure is followed. The tracks are terminated if the cyclone's maximum wind speed falls below 13ms^{-1}, if the cyclone travels outside a predefined

*A Markov chain is a random process on a state space characterised by its memoryless property. The distribution of the next state depends only on the current state.

latitude-longitude box, or after 30 days duration, whichever happens first[6]. The model for termination in [10] is decided probabilistically and is constructed using historical termination rates from the data. Similarly, in [20], [21], the termination decision at each step is decided stochastically via a Bernoulli experiment with a probability of success $p(\mathbf{t}, Z)$. The probability of success depends both on the storms current location and wind speed, and the termination probability is determined as the maximum of the two probabilities p_Z and $p_{\mathbf{t}}$ [20].

Outlook

Developing and applying stochastic models for the generation of synthetic tropical cyclone tracks is by no means a complete field. There are many gaps in knowledge and understanding both in the statistical methods framework and in their application to weather and climate. The physics of tropical cyclones is not fully understood, for example interaction with ocean eddies and there is a large amount of uncertainty on how the tropical cyclone tracks and locations of tropical cyclones will vary in a changing climate [17]. In future work it is essential to evaluate the skill of the stochastic simulation models and to quantify their uncertainty. In a recent (2012) IPCC report "Managing The Risks Of Extreme Events And Disasters To Advance Climate Change Adaptation"[7], the many references to tropical cyclones highlight the importance of new and improved, accurate and comprehensive models for analysing these extreme phenomena [17]. Direct quotes from the report say, "average tropical cyclone maximum wind speed is likely to increase", "heavy rainfalls associated with tropical cyclones are likely to increase with continued warming", and there is also high confidence that "increases in exposure will result in higher direct economic losses from tropical cyclones" [7].

Hazard Assessment

Financial Aspects

There is a wide literature on the financial aspects of hazard assessment in relation to natural catastrophes. In [3], there is a section dedicated to so-called "Mega-Catastrophe and Clash Loss". This discusses, for example how tropical cyclones or hurricanes can cause large losses and devastation by impacting several locations simultaneously. The landfall rates and return periods of these types of storms are of great importance to insurance companies and governmental agencies due to their potentially devastating effects. Also [3] contains an in-depth discussion regarding vulnerability identification, modelling of cat risk and the management and mitigation of risk through cat bonds and cat derivatives.

In [21], clash probabilities are investigated for the Atlantic basin for joint regions of interest. A method for calculating return periods is also presented in [21].

Wind Profiles

Wind profile models for tropical cyclones is important not only for physical understanding but also for the calculation of return periods. There are a wide variety of approaches, [12] [13] [21] [31]. A wind profile has increasing wind speeds from the eye of the storm to the radius of maximum

winds and then decreasing wind speeds thereafter. Some models, for example [31] model the wind profile using one continuous function. However, there are other approaches that have a split model between the profile from the eye of the storm to the radius of maximum wind speed and then from the maximum wind speed to the radius of gale force winds.

An important aspect and practical aspect of reconstructing the wind profiles is calculating the minimum sea level pressure (MSLP) given the maximum wind velocity as it is required as input for some wind profile models. One of the most popular empirical relationships used to calculate the MSLP is given in [2]. In recent literature, more advanced relationships have been developed for example in [16], that takes into account the effects of a storms translation speed and latitude.

Chapter 3

Data

In this chapter, the data that is used for this project is discussed in detail including the source of the data, the method of classification and other data analysis techniques that were employed.

3.1 Data Source

The data for this project was obtained via the Joint Typhoon Warning Centre (JTWC) of the US Navy, namely the Best Track Data for the western North Pacific. The data used ranges from 1945 to 2004 inclusive. The data varies in the available information. For example, in the later years of the data set more varied set of measurements are recorded.

3.2 Bad Data points

In the data, there are two storms one from September 1958 and another from July 1959 where no wind speeds were recorded for the entirety of these storms, they lasted for two days and three days respectively. The failed recording of the wind speeds for these storms could be due to faulty equipment or that their wind speeds were relatively negligible. These two storms were removed from the data set.

In the rest of the data, there were other "non-recordings" for particular tropical cyclones, here linear interpolation was used to estimate the wind speed between two recorded speeds.

3.3 Data analysis

Analysis of the data and its formatting to get it into a usable shape was completed using both Python, Excel and R. In the model which is being implement, the observation window is bounded by the equator in the south, a latitude of $60°$N in the north, $100°$E in the west and the date line in the

east. The observation window is the divided into four zones. These zones roughly correspond to the major land and sea masses in the observation window.

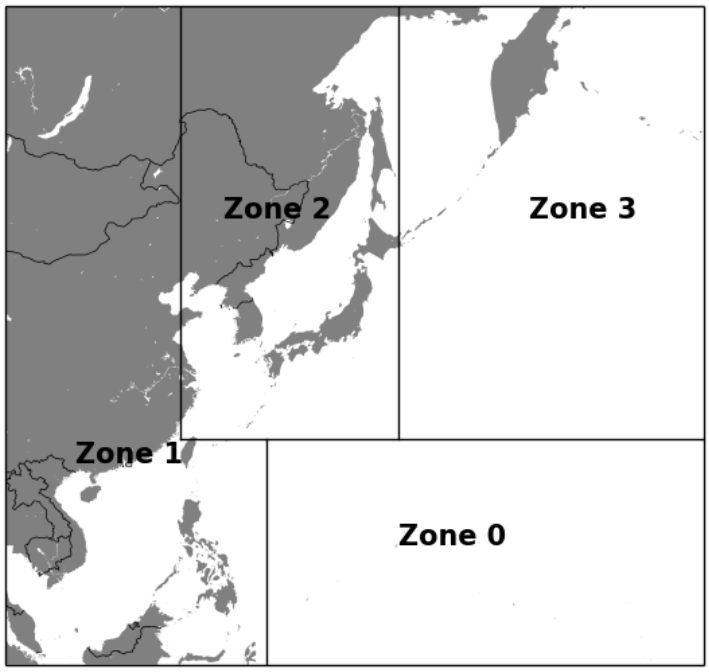


Figure 3.1: Observation window

The observation window is divided as follows (Figure 3.1):

- Zone 0 – is bounded below by the equator and above by $25°$N, from the left by $130°$E and to the right by the date line.
- Zone 1 – is L-shaped and is bounded below by the equator to the left by $100°$E above by $25°$N and $60°$N and to the right by $120°$E and $130°$E.
- Zone 2 – is bounded below by $25°$N and above by $60°$N, from the left by $120°$E and to the right by $145°$E.
- Zone 3 – is bounded below by $25°$N and above by $60°$N, from the left by $145°$E and the date line to the right.

Each tropical cyclone from the data set is classified into six different classes depending on which zone the cyclone starts in, which zones it transitions through and in which zone the tropical cyclone terminates. The algorithm for the generation of the synthetic tropical cyclone tracks is run individually for each class. The idea behind splitting the data into classes is to create a decent amount of homogeneity among the track shapes without having too many classes [20].Class 4 is created with the storms in class 2 that have their staring point in zone 1, and class 5 is created by the

Table 3.1: Criteria for the classification of the cyclone tracks.

Start in zone	Touched zones	End in zone	Class
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
0 or 1	0 and 1	0 or 1	1
0 or 2	0 and 2	0 or 2	2
0 or 3	0 and 3	0 or 3	3
1 or 2	1 and 2	1	1
1 or 2	1 and 2	2	2
2 or 3	2 and 3	2 or 3	2
0, 1 or 2	0, 1 and 2	0	0
0, 1 or 2	0, 1 and 2	1	1
0, 1 or 2	0, 1 and 2	2	2
0, 1 or 3	0, 1 and 3	0	0
0, 1 or 3	0, 1 and 3	1	1
0, 1 or 3	0, 1 and 3	2	2
0, 2 or 3	0, 2 and 3	0, 2 or 3	2
1, 2 or 3	1, 2 and 3	1	1
1, 2 or 3	1 , 2 and 3	2 or 3	2
0, 1, 2 or 3	0, 1, 2 and 3	0	0
0, 1, 2 or 3	0, 1, 2 and 3	1	1
0, 1, 2 or 3	0, 1, 2 and 3	2 or 3	2

storms in class 1 which have their starting point east of 122°E. The process of splitting the tropical cyclones into classes induces a certain amount of arbitrariness, though this is not considered to be detrimental to the model, since no information is lost and a helpful tool for simulation is gained [20]. The algorithm for the generation of the synthetic tropical cyclone tracks and all of the algorithms steps are completed separately for each class. Note the total values are the sum of classes 0-3 inclusive, since classes 4 and 5 are subclasses.

Table 3.2: Historical Track data

Class	Number of tracks	Number of data points
0	179	3863
1	803	20368
2	317	10572
3	411	14084
4	65	1443
5	559	16395
Total	1710	48887

Chapter 4

The Model

In this chapter, the model is outlined and the first aim of this project i.e. the implementation in Python of the model is discussed. Heuristic comparisons between the simulations and the data are also drawn.

4.1 Initialisation

The first step in the modelling process is to generate the genesis points of the simulated tropical cyclone tracks. On observation, by plotting the historical genesis points in the observation window (Figure 4.1), it is clear that the genesis points are inhomogeneously distributed. Also, there is no meteorological justification of any interaction between cyclogenesis points, this is reinforced by the length of the data set considered. As a result, an inhomogeneous Possion process is chosen as a model for the cyclogenesis. The selection of an inhomogeneous spatial Poisson process is further justified by its two fundamental properties. The first of said properties is independence. More precisely, if k is any integer and if $B_1, \ldots, B_k$ are any disjoint (mutually non-intersecting) Borel sets, then the random variables $N(B_1), \ldots, N(B_k)$ are stochastically independent [25]. This property is also common to homogeneous Poisson processes, it is the second property that distinguishes the imhomogeneous case from the homogeneous. The second property of an inhomogeneous Poisson process is the number $N(B)$ of points in a bounded Borel set B has a Poisson distribution of the parameter $\Lambda(B)$. Here Λ is a measure, the so-called *intensity measure* [25]. It is important to note that there is no location $\mathbf{t}$ such that $\Lambda(\{\mathbf{t}\}) > 0$, or more precisely,

$$\frac{\mathbb{P}\left[N\left(t, t+h\right]=1\right]}{h} \rightarrow \lambda(\mathbf{t}) \quad \text{as} \quad h \downarrow 0, \tag{4.1}$$

$$\frac{\mathbb{P}\left[N\left(t, t+h\right]>1\right]}{h} \rightarrow 0 \quad \text{as} \quad h \downarrow 0, \tag{4.2}$$

where $\lambda(\mathbf{t})$ is the density function (with respect to the Lebesgue measure) of Λ, assuming the distribution function Λ is absolutely continuous. The function $\lambda(\mathbf{t})$ is called the *intensity function*. The intensity function of the inhomogeneous Poisson process has an intuitive infinitesimal interpretation; $\lambda(\mathbf{t})d(\mathbf{t})$ is the infinitesimal probability that there is a point of the inhomogeneous process in a disk of area $d(\mathbf{t})$ centred at $(\mathbf{t})$ [26].

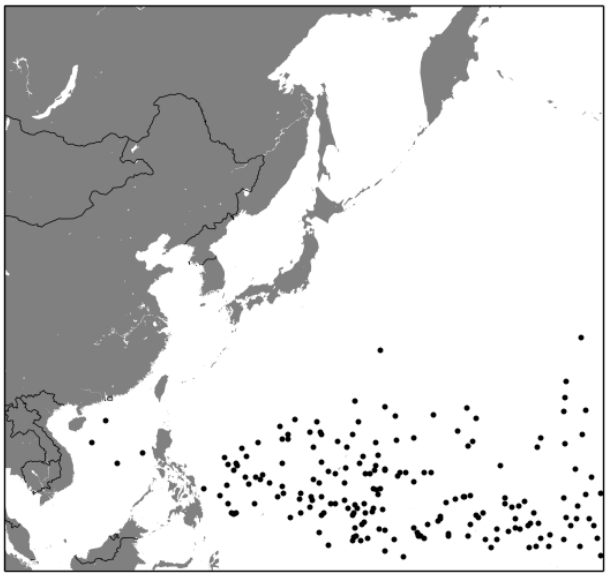


Figure 4.1: Class 0 historical genesis points.

Inhomogeneous Poisson processes have an important thinning property, this is discussed in the implementation in Section 4.1.1.

A detailed hypothesis test is conducted in [20], the details of which are not presented here. However, the hypothesis test that is carried out is that the number of storms in the data is Poisson distributed, i.e

$$H_0 : P \in \{\text{Poi}(\lambda), \lambda > 0\} \qquad \text{vs.} \qquad H_\text{a} : P \notin \{\text{Poi}(\lambda), \lambda > 0\}, \tag{4.3}$$

where P is the distribution of the number of tropical cyclones per year within the observation window [20]. The result is that the hypothesis H_0 is not rejected.

For the selected genesis model, it is necessary to determine the intensity function. The intensity function $\lambda(\mathbf{t})$ determines the distribution of the points within the observation window. The intensity is estimated using a non-parametric estimation technique namely the generalised nearest neighbour estimator as shown in equation 4.4 (described in detail in [23]),

$$\hat{\lambda}(\mathbf{t}) = r_k(\mathbf{t})^{-2} \sum_{i=1}^{m} K_e\left\{r_k(\mathbf{t})^{-1}(\mathbf{t} - \mathbf{T}_i)\right\} \tag{4.4}$$

where $r_k(\mathbf{t})$ is the distance to the k-th nearest point of genesis from the location $\mathbf{t}$, $\mathbf{T}_i$ is the location of the i-th historical point of genesis and K_e the Epanechnikov kernel [20]:

$$K_e(\mathbf{t}) = \begin{cases} \frac{2}{\pi}\left(1 - \mathbf{t}^\top \mathbf{t}\right) & \text{if } \mathbf{t}^\top \mathbf{t} < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

The parameter k is chosen such that $k = \lfloor \sqrt{m} \rfloor$, where m is the number of historical points of genesis for each class [20].

4.1.1 Implementation

In this section, the implementation of the initialisation stage of the model is discussed. In order to implement the inhomogeneous Poisson process to simulate the genesis points, the property of thinning associated with Poisson fields is employed. This mechanism called location-dependent independent thinning, involves 'thinning' a homogeneous process into a inhomogeneous process. It is assumed that $\lambda(\mathbf{t})$ is bounded, i.e. there is a λ^* such that

$$\lambda(\mathbf{t}) \leq \lambda^*, \tag{4.6}$$

for all $\mathbf{t}$. The function,

$$p(\mathbf{t}) = \frac{\lambda(\mathbf{t})}{\lambda^*}, \tag{4.7}$$

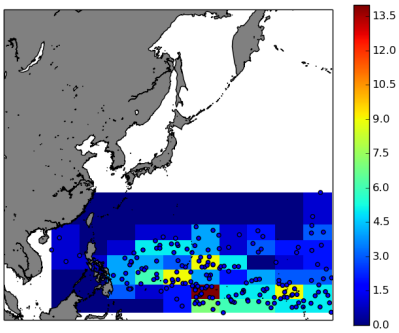
is used as the location dependent thinning function [25]. Then the values of the function are $0 < p(\mathbf{t}) \leq 1$.

The first step of the simulation is to initiate a homogeneous Possion process with intensity λ^*, for the model developed here, λ^* is the maximum intensity of all genesis points in the data. Recall that this step is executed individually for each class, ie. there will be six individual maximum intensities $\lambda_0^*, \ldots, \lambda_5^*$ corresponding to each class of tropical cyclones. For each point generated by the homogeneous process, it is independently decided whether or not each generated point is eliminated or not. This decision is made stochastically by a random draw from a standard uniform distribution as follows, if

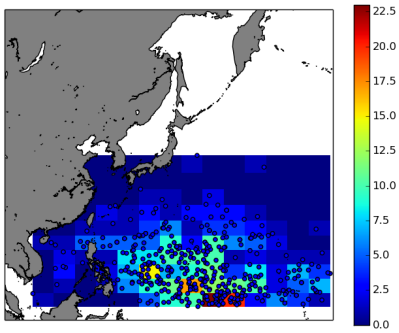
$$u > p(\mathbf{t}), \tag{4.8}$$

then that point is rejected, $u \sim Unif\,[0, 1]$. If the point is not rejected, it is retained as a sample of the inhomogeneous Poisson process. It is also necessary to remove genesis points that occur over land, as this is not physically possible. This is achieved by making use of the `is_land` function as

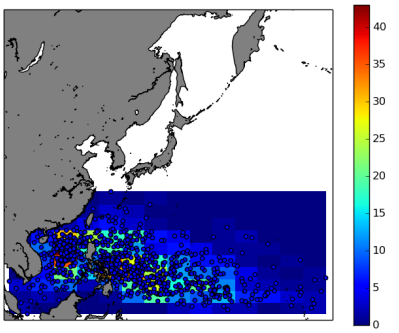
part of the `Basemap` package. Also, the intensity is set to zero in the region within 3° of latitude of the equator due to the negligible Coriolis force.



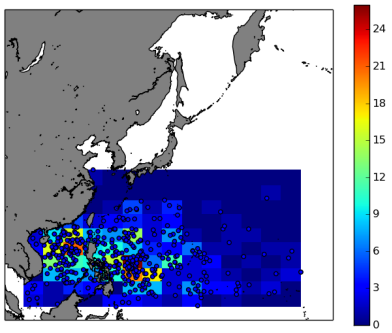
(a) Class 0 Data



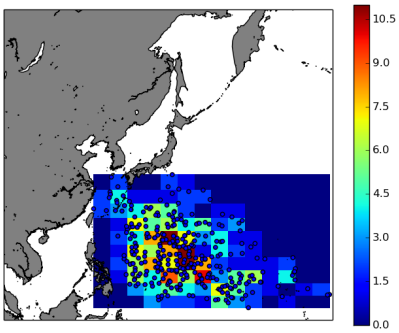
(b) Class 0 Simulated



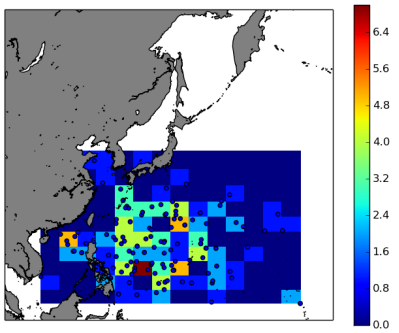
(c) Class 1 Data



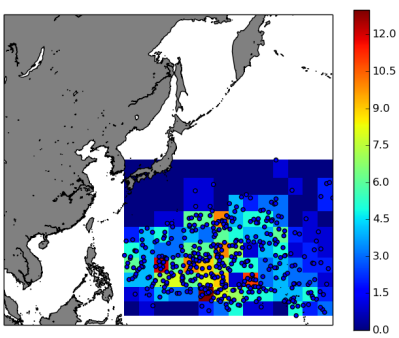
(d) Class 1 Simulated



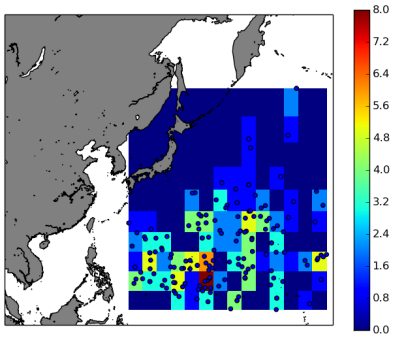
(e) Class 2 Data



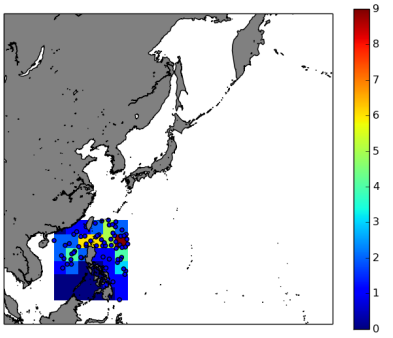
(f) Class 2 Simulated



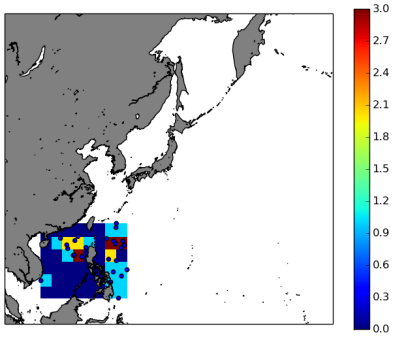
(g) Class 3 Data



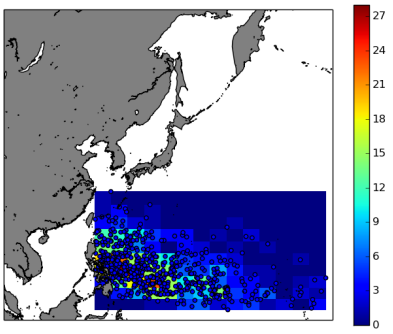
(h) Class 3 Simulated



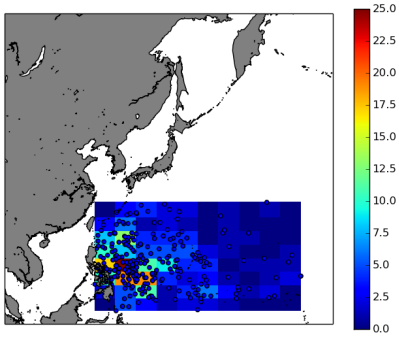
(i) Class 4 Data



(j) Class 4 Simulated



(k) Class 5 Data



(l) Class 5 Simulated

Figure 4.0: Comparison of the data and simulated genesis points for each class.

4.2 Track Propagation

This part of the modelling process is relies on a basic assumption that tropical cyclones located in similar areas behave in a similar manner. This assumption is reinforced heuristically by analysing the track shapes of the data in the observation window and further by splitting the data into classes as described in Chapter 3. The model for the propagation of the synthetic tracks consists three characteristics namely direction of movement of the tropical cyclone (denoted by X), the translation speed (Y), i.e the velocity at which the cyclone in moving in the direction X and the maximum wind speed (Z) [20]. It is assumed, in line with the data measurements, that these characteristics remain constant over 6-hour time periods and the characteristics are instantaneously updated to created the next polygonal segment in the tracks trajectory.

Each simulated track is the sum of the values of the initial characteristics and the changes in those values. Each track can then be described by S_i, a 3-dimensional state vector after the i-th track segment,

$$S_i = S_0 + \sum_{j=1}^{i} \Delta S_j = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \sum_{j=1}^{i} \begin{pmatrix} \Delta X_j \\ \Delta Y_j \\ \Delta Z_j \end{pmatrix}, \tag{4.9}$$

the characteristics X_0, Y_0 and Z_0 as well as $\Delta X_j, \Delta Y_j$ and ΔZ_j are considered to be random variables [20].

The values of the characteristics are generated by sampling from local probability distribution functions constructed from the k-nearest historical measurements of the characteristics around a location. Note that the k parameter here is different to the k parameter in the generalised nearest neighbour estimator used for the genesis locations. An example of how these distribution functions are constructed is now given for X_0 at a location $\mathbf{t}$,

$$F_{X_0}(x, \mathbf{t}) = \frac{\#\left\{l : 1 \leq l \leq k_{X_0}, x_l^{(0)}(\mathbf{t}) \leq x\right\}}{k_{X_0}}, \tag{4.10}$$

where $x_l^{(0)}(\mathbf{t})$, $l = 1, \ldots, k_{X_0}$, denote the k_{X_0} historical realisations of X_0 closest to the location $\mathbf{t}$ [20]. Similar formulae are used in the construction of the location-dependant distribution functions for Y_0, Z_0. Similarly, for the changes in directions (and indeed similar for the changes in Y characteristic) the probability distribution is given by,

$$F_{\Delta X}(x, \mathbf{t}) = \frac{\#\left\{l : 1 \leq l \leq k_{\Delta X}, x_l(\mathbf{t}) \leq x\right\}}{k_{\Delta X}}, \tag{4.11}$$

where $x_l(\mathbf{t})$, $l = 1, \ldots, k_{\Delta X}$, denote the $k_{\Delta X}$ historical realisations of ΔX closest to the location $\mathbf{t}$ [20]. The distribution of a change in direction depends on all changes in direction, no matter what

step of the tropical cyclone they occurred. This is also true for the changes in translation speed.

To ensure that the model conforms to reality, certain boundary conditions are imposed both on direction (X) and translation speed (Y). For example, when calculating directions and changes in directions the values are made modulo $360°$, so the the value of X_j, where j is the step in the track, is within the interval $[0°, 360°)$, where $0°$ is considered due North. Another such boundary condition is implemented on the translation speeds, that is they must be non-negative at all times [20].

A different form of the local location-dependant probability distribution function is used to formulate the distribution of ΔZ. Here not only does the distribution depend upon the location $\mathbf{t}$ but also on the current wind speed z,

$$F_{\Delta Z}(x, \mathbf{t}, z) = \frac{\#\{l : 1 \leq l \leq k_{\Delta Z}, \Delta z_l(\mathbf{t}, z) \leq x\}}{k_{\Delta Z}}, \tag{4.12}$$

where $\Delta z_l(\mathbf{t}, z), l = 1, \ldots, k_{\Delta Z}$, denote the $k_{\Delta Z}$ historical realisations of ΔZ closest to the location $\mathbf{t}$ and in the same range as z, namely "low", "medium-low", "medium-high" and "high" [20]. How these ranges are defined is discussed in the implementation 4.2.1.

4.2.1 Implementation

The implementation of the propagation part of the model consists of nine main functions. Six functions for each of the initial characteristics and their changes, another function that calculates bearing (direction) between two respective locations, a function to generate a sample from the generated distribution functions and a function to calculate a new location given the sampled characteristics for the next step.

In the implementation, extensive use was made of the `geopy` package in Python. In particular, the `geopy.great_circle` function calculates the great-circle distance* between two locations, this is used for example when calculating distances between the current location and the nearest k neighbours. This function uses a spherical model of the earth, using an the average great-circle radius of 6372.795kms, which can result in an error of up to $\approx 0.5\%$ according to the `geopy` documentation. It is unclear from the model description in [20] what distance is used, so the great-circle distance is determined to be an accurate method to determine such distances.

When handling directions and the changes in directions the calculation result must be made modulo $360°$. Care must be taken in particular when calculating the changes in directions, for example say at step j the direction is $5°$ and at step $j+1$ the direction is $345°$ then $(345-5)\text{mod } 360$ is 340 which is the incorrect value for the change in direction it is actually $-20°$. This mistake can be avoided by taking the signed angle via the following correction $(\Delta x + 180)\%360 - 180$, where

*The great-circle (geodesic) is the shortest distance between two points on the surface of a sphere.

Δx is the calculated change in direction and $\%$ is the $modulo$ operation.

The calculation of translation speeds is made by using the following relationship,

$$\text{speed} = \frac{\text{distance}}{\text{time}} \tag{4.13}$$

the distances are calculated as described earlier using the `geopy.great_circle` function and the time is taken as 6 hours which is in accordance with the data set, as measurements in the Best Track data set are recorded at 6 hour intervals.

For the changes in wind speed (ΔZ) the local probability distribution functions are calculated by using wind speeds in the same range. The ranges are not defined clearly in [20] and it is decided that the quartiles of the speeds of the data set and they correspond to the "low", "medium-low", "medium-high" and "high" ranges.

To generate a random sample from the constructed local probability distribution functions the Inverse transform sampling method (ITM) is implemented. The method of sampling is not made clear in [20], so ITM is decided as an appropriate method as it is now described. ITM generates samples at random from a probability distribution given its cumulative distribution function (cdf). Mathematically, this can be described as follows: Let $F(x)$, $x \in \mathbb{R}$, denote any cumulative distribution function (continuous or not) and recall $F : \mathbb{R} \to [0, 1]$. That is F is non-negative and monotone and is *càdlàg*, where $F(\infty) = 1$ and $F(-\infty) = 0$. The generalised inverse of a cdf F is $F^{-1} : [0, 1] \to \mathbb{R}$, which is defined as,

$$F^{-1} = \inf\{x : F(x) \geq y\}, y \in [0, 1]. \tag{4.14}$$

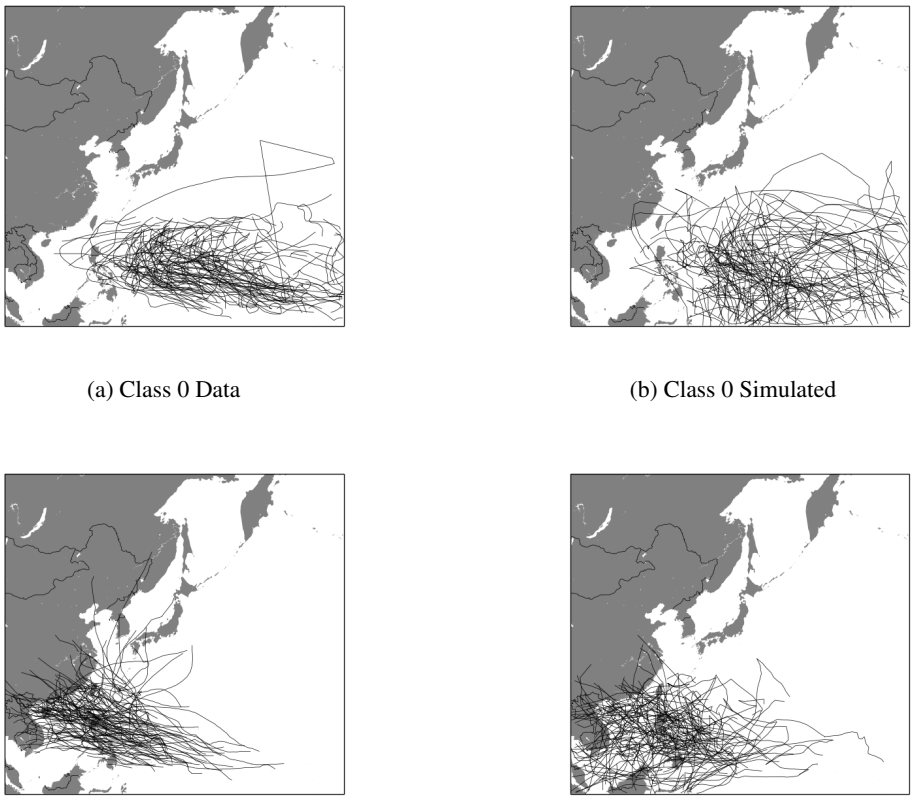
Proposition 4.2.1 *Let $F(x)$, $x \in \mathbb{R}$, denote any cumulative distribution function (cdf) (continuous or not). Let $F^{-1}(y), y \in [0, 1]$ denote the inverse function defined in (4.14). Define $X = F^{-1}(U)$, where U has the continuous uniform distribution over the interval $(0, 1)$. Then X is distributed as F, that is, $P(X \leq x) = F(x), x \in \mathbb{R}$ [22].*

Once all the updated characteristics have been generated, a function called `new_loc` takes the current position and the values of the updated characteristics and calculates the new location along the polygonal track.

4.3 Track Termination

As described in 4.2.1, each new location of the simulated tracks is updated in steps of 6 hours, after each step of the synthetic tracks it must be decided whether or not the track terminates at the current step or continues. This decision is determined stochastically via a Bernoulli experiment with a probability of success $p(\mathbf{t}, Z)$, depending on the storms current location and wind speed [20].

The probability of termination is constructed as the maximum of two probabilities p_Z and $p_{\mathbf{t}}$, these probabilities correspond to the storms wind speed and location respectively.



(a) Class 0 Data

(b) Class 0 Simulated

(c) Class 1 Data

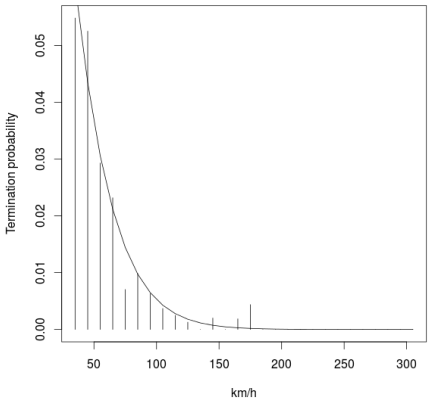
(d) Class 1 Simulated

Figure 4.1: Comparison of simulated tracks and tracks from the data set.

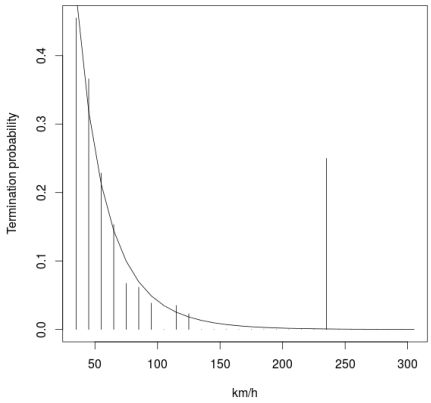
The wind speeds at the terminations points of the tropical cyclones vary considerably in the data set. It is expected that termination will occur at lower wind speeds and increasingly less so at higher wind speeds and indeed this is the case. It is also well known that the behaviour of tropical cyclones differs over land than over sea. In order to take account of this fact, terminations over land and sea are fitted individually. A curve of the form

$$p_Z = c \cdot e^{-\lambda Z^{\alpha}}, \tag{4.15}$$

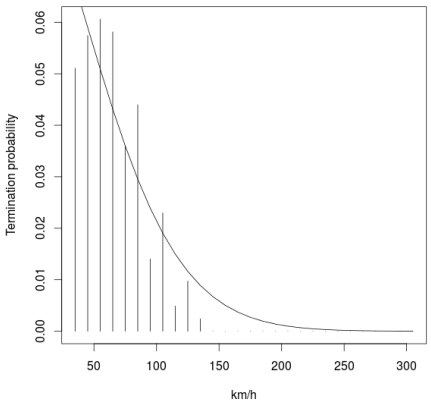
is fit to the historical measurements of termination wind speeds where p_Z depends on the current wind speed Z. In Figure 4.2, the curve 4.15 is fitted to the termination data of class 1 and 2 for both termination over land and sea.



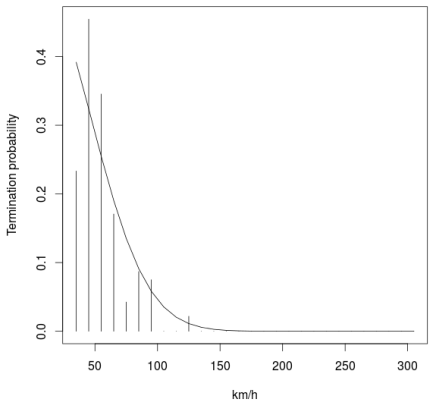
(a) Class 1 Sea



(b) Class 1 Land



(c) Class 2 Sea



(d) Class 2 Land

Figure 4.2: Curve fitted to historical termination probabilities for (a) Class 1 over sea, (b) Class 1 over land, (c) Class 2 over sea and (d) Class 2 over land.

Figure 4.2(b) shows the fitted curve for class 1 land termination data, there is one high wind speed of 230km/h that is recorded as a termination speed in the data, this value appears to be an outlier. One possible reason for such a large termination wind speed is that measurements of the tropical cyclone were not carried out after this time, even though it may still have been an active cyclone. There are may possible explanations for this, for example faulty equipment. However, this outlier remains a part of the data set as it is not see to have a negative effect on the fitting of the wind speed termination curve. Another option would be to bin the data so that there are the same number of points in each bin, this would remove artificial outliers.

The termination of tropical cyclones is highly influenced by its geographical location, for example over land or in the higher latitudes. To account for this the second probability $p_{\mathbf{t}}$ is calculated as the fraction of termination points among the k nearest points of observation from the location $\mathbf{t}$ in the historical data [20]. For example, suppose at the current location there are 2 termination points among the nearest 10 observations, this gives that $p_{\mathbf{t}} = 0.2$.

Thus the probability $p(\mathbf{t}, Z)$ of termination at the current location $\mathbf{t}$ is given by the combination of these two probabilities [20]

$$p(\mathbf{t}, Z) = \max\{p_Z, p_{\mathbf{t}}\}. \tag{4.16}$$

This stochastic method for determining if a track terminates accounts for both weak tropical cyclones with relatively low wind speeds, for tropical cyclones that transition over land by ensuring that they do not penetrate unrealistically deep into the continent.

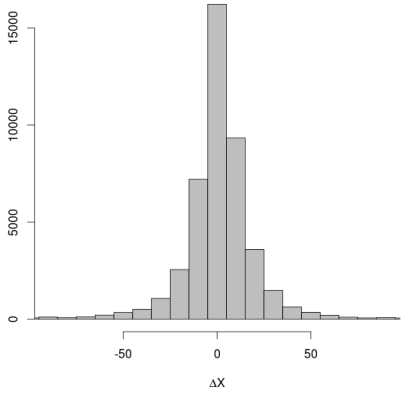
4.3.1 Implementation

There are three main functions implemented for the termination part of the model, not surprisingly named `pT`, `pZ` and `p_termin`. The function `pT` calculates the probability $p_{\mathbf{t}}$. It takes the great-circle distance from the current location $\mathbf{t}$ to all the historical points around it. It then takes the nearest k points to $\mathbf{t}$ and counts the number of termination points and returns the probability $p_{\mathbf{t}}$.

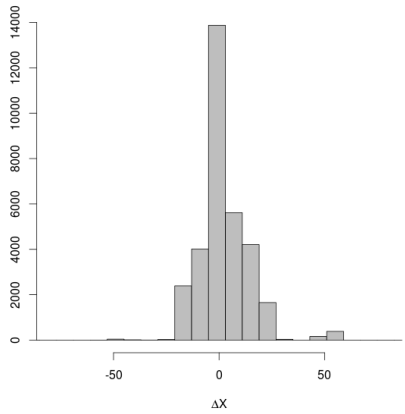
The function `pZ` calculates the probability p_Z. This function consists of the fitted curve 4.15 for each class for both over land and sea, including the estimated coefficients for each class estimated using R. Note that class 0 and class 3 have no termination points over land. The estimation of these coefficients is achieved as follows; firstly all the points of measurement and their corresponding wind speeds are binned into 10km/h wide bins. The fraction of the number of termination points in each bin is then calculated. To avoid artefacts resulting from imperfect data, points with wind speeds less than 30km/h are omitted [20]. The curve is fitted via least-squares by making use of the `optim` function in R.

Finally, now that the probabilities $p_{\mathbf{t}}$ and p_Z have been calculated, it needs to be decided if the track will terminate at the current point $\mathbf{t}$, this is determined stochastically by the function `p_termin`. Firstly, the maximum of the two probabilities $p_{\mathbf{t}}$ and p_Z is calculated, this is the *prob-*

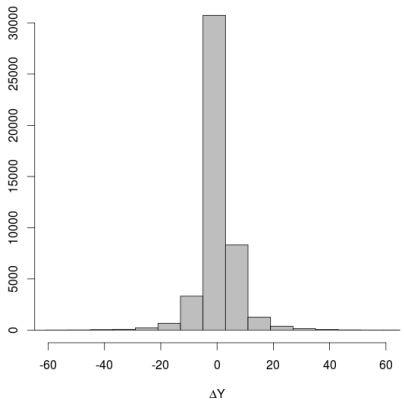
ability of termination. Then a random draw from the standard uniform distribution is taken, using the `Numpy` function `np.random.uniform()`, and it is compared to the probability of termination. If the probability of termination is greater than the random sample from the standard uniform distribution the track terminates at the current location **t**, otherwise the track continues to the next step.



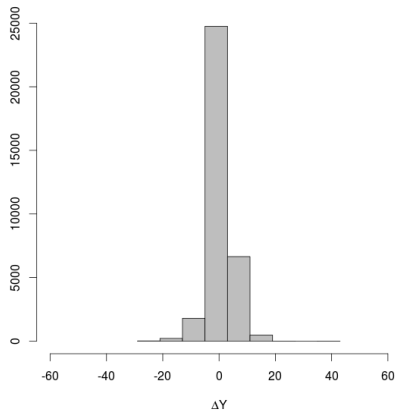
(a) ΔX Data



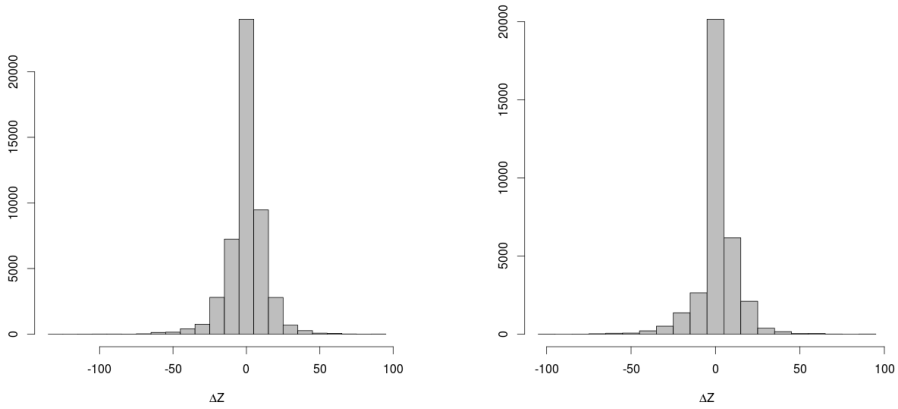
(b) ΔX Simulation



(c) ΔY Data



(d) ΔY Simulation



(e) ΔZ Data (f) ΔZ Simulation

Figure 4.1: A comparison between the changes in the characteristics between the data (left) and the simulation (right).

4.4 Algorithm

Here the algorithm for the generation of the synthetic tropical cyclones is described. This section unites the different parts of the model described previously and elucidates the interaction between the various components. Recall that the algorithm is run individually for each of the six classes as mentioned in Chapter 3.

1. **Initialisation:** Find all required estimators and probabilities as previously defined, (4.4), (4.10), (4.11), (4.12), (4.15) and (4.16). Then go to step 2.

2. **Genesis points:** Generate a realisation of the inhomogeneous Poisson process with intensity parameter as defined in (4.4) and go to step 3.

3. **Choose a point:** From the realisation of the inhomogeneous Possion process in step 1, pick a genesis point that does not yet have a track associated with it and go to step 4. Otherwise, is there are no such points left terminate the algorithm.

4. **Initial segment:** For the selected genesis point in step 3, generate a realisation of S_0 from the distribution functions defined in (4.10), where the location **t** is the selected genesis point. Then find the new location after its first segment given by the randomly sampled initial characteristics. Next go to step 5.

5. **Probability of termination:** Conduct a Bernoulli experiment with probability of success given by (4.16) given the storms current location **t** and current wind speed Z. If the result is

a “success”, terminate the track and go to step 7. Otherwise, go to step 6.

6. **Additional segment:** Generate a realisation of ΔS_j from the distribution functions given by (4.11) and (4.12). Add ΔS_j to the previous segment S_{j-1} to get S_j, from this the next location of the track can be calculated. Then return to step 5.

7. **Class verification:** Determine the class of the generated storm track as described in Chapter 3. If the class of the storms track matches that of which the algorithm is being performed, then the track is accepted for the selected point of genesis and return to step 3. Otherwise, reject the storm track and go to step 4 with the same point of genesis [20].

Note the possibility of a storm being rejected as described in step 7 of the algorithm is not just theoretical, in fact quite a few rejections do happen during the simulation [20]. One such reason for this is the value of the parameter k that is discussed in Chapter 5.

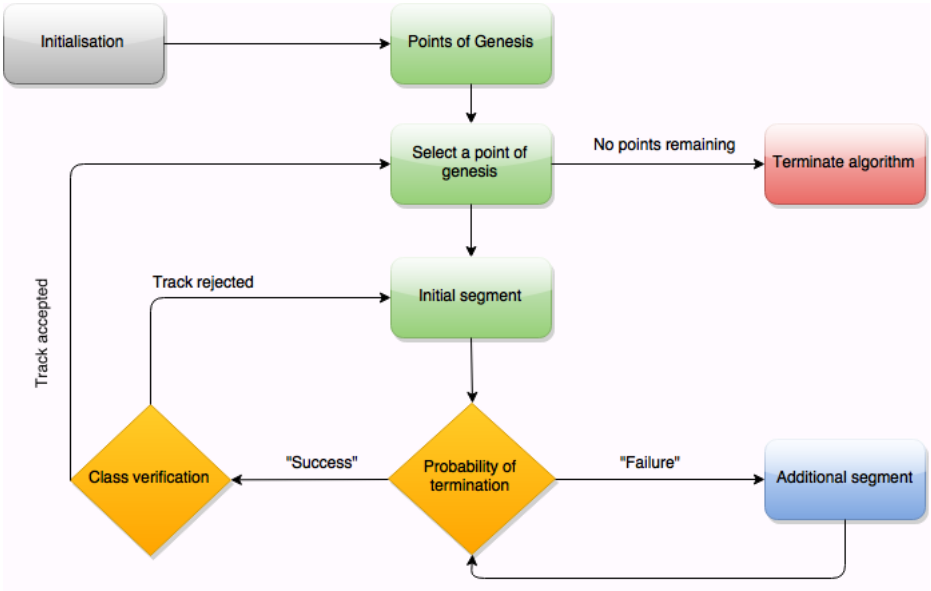


Figure 4.2: Flowchart of algorithm.

Chapter 5

Model Evaluation

As was one of the aims of this project, in this chapter the model is evaluated by comparing various landfall statistics of the model to the data set. This evaluation is carried out for different values of the parameter k. The selection of this parameter was evaded in [20], here an investigation is carried out into a suitable value of this parameter. To ensure that the model correctly estimates the mean landfall rates across several locations in the observation window. The parameter k determines how many of the nearest neighbours are to be used in the calculation of the local location-dependent probability distribution functions and also the number of neighbours to be used in the calculation of the $p_{\mathbf{t}}$ probability.

To conduct the one sample test, 75 runs of the algorithm was generated without classes based on the entire data set of 60 years. The number of storms on each run is Poisson-distributed with a parameter (N) given by the mean number of storms per year in the data. This value can be calculated by taking the number of tracks in the data set divided by the number of years of data. For this data set the value is given by

$$N = \mathbb{E}\left[\text{mean } \# \text{ of storms per year}\right] = \frac{1710}{60} = 28.5. \tag{5.1}$$

Thus in general, the expected number of storms over a period of m years for this data set is given by $28.5 \times m$.

The null hypothesis is that there is no difference between the means, i.e.

$$H_0 : \mu_{\text{sim}} = \mu_{\text{p}} \qquad \text{vs.} \qquad H_{\text{a}} : \mu_{\text{sim}} \neq \mu_{\text{p}} \tag{5.2}$$

where μ_{p} is the "population" mean, that is the mean landfall value from the data is taken as the population mean and μ_{sim} is the mean landfall rate from the simulation. For now, assume that $X_1, \ldots, X_n$ are independently distributed as $\mathcal{N}(\mu, \sigma^2)$, and the aim is to test the hypothesis given

in 5.2 using the test statistic,

$$t_n = t_n(X_1, \ldots, X_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_{\text{p}})}{S_n} = \frac{\sqrt{n}(\overline{X}_n - \mu_{\text{p}})}{\sigma} \Bigg/ \frac{S_n}{\sigma} \tag{5.3}$$

where the sample variance is given by $S_n^2 = \Sigma(X_i - \overline{X}_n)^2/(n-1)$ [5]. When $\mu_{\text{p}} = \mu_{\text{sim}}$ and the X's are normal, t_n has the t-distribution with $n-1$ degrees of freedom. However, suppose that the normality assumption fails and the X's are instead distributed according to some other distribution F with mean μ_{p} and finite variance [5]. Then according to the Central Limit Theorem, $\sqrt{n}(\overline{X}_n - \mu_{\text{p}})/\sigma$ has limiting distribution $\mathcal{N}(0,1)$. Note that S_n/σ converges to 1 in probability, thus by Slutsky's theorem, t_n has limiting distribution $\mathcal{N}(0,1)$ regardless of the distribution function F [5]. In particular, the t-distribution with $n-1$ degrees of freedom tends to $\mathcal{N}(0,1)$ as $n \to \infty$.

The hypothesis test is carried out at a significance level of $\alpha = 0.05$. Since this is a two-sided test there will be a 95% acceptance region with a critical region of $\alpha/2 = 2.5\%$ in each tail.

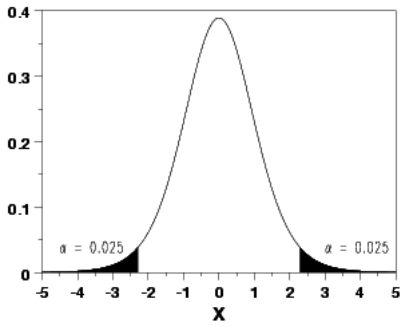


Figure 5.1: t-distribution with $\alpha = 0.05$

Firstly, on investigation, it is found that the value of the parameter $k = 5$ is too small. At $k = 5$, the simulated tracks exhibit unrealistic behaviour. In some cases, the tropical cyclone's translation speed (Y) hits 0km/h and remains at this speed for a number of steps until p_{Z} is large enough that the track terminates or the translation speed picks up. This phenomenon is due to the value of k being too small. The tropical cyclone may transition into an area of the observation window where the changes in translation speed (ΔY) are small or negative and since k is small, the model will only sample from these small ΔY values. Thus, the value of $k = 5$ is unrealistic. However, in order for the model to capture the local behaviour of the tropical cyclones in the data at each point, the model requires a "small enough" value of k, but not too small that the tropical cyclones become stationary. Therefore sensitivity analysis is carried out for various values of the parameter k.

The next value of the parameter k tested is $k = 10$. For this value, the tracks rarely exhibit the behaviour of getting stuck at a zero translation speed. This value of k is seen to be realistic. Therefore, the hypothesis test 5.2 is conducted on this value at several locations in the observation

window. The details of the results is given in Table 5.1.

The locations of Japan, Hong Kong (including surrounding metropolitan area), Shanghai, Taiwan and the Philippines were chosen are they are considered to be of both financial interest due to their size and economy and also due to their large population sizes.

Table 5.1: Table of results for hypothesis test 5.2 with $k = 10$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Japan	3.2933	3.6666	1.6421	A
Hong Kong	0.3733	0.4333	0.6319	A
Shanghai	0.133	0.0333	0.4746	A
Taiwan	1.1333	2.15	1.2	R
Philippines	4.28	8.2203	1.8785	R

For the locations of Japan, Hong Kong and Shanghai, the model captures the mean landfall rates in the respective areas well. However, for the areas of Taiwan and the Philippines the model underestimates the mean landfalls rates of the data set, in both case by approximately half. Possibly a larger sample size, i.e. > 75, may over come this issue. Also, a larger data set to based these mean landfall rates may result in different values of μ_{p}.

In addition to analysing specific individual locations, it is important for both financial applications in hazard assessment (discussed later Chapter 6) and model evaluation to test if the model captures tropical cyclones that not only affect one of these defined areas but two at once. That is, does the model capture the behaviour of tropical cyclones that transition into two areas of interest [*] along their path. In order to evaluate this, pair of locations in the observation window are considered. Note, there are some pairs of locations that do not have any such probabilities, therefore these areas are not examined. The results of this test are summarised in Table 5.2.

Table 5.2: Table of results for hypothesis test 5.2 with $k = 10$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Taiwan & Philippines	0.52	0.7666	0.7946	R
Taiwan & Japan	0.16	0.1	0.4362	A
Taiwan & Shanghai	0.0266	0.0166	0.1622	A
Philippines & Hong Kong	0.2267	0.2166	0.4814	A
Philippines & Japan	0.4133	0.25	0.7369	A

It is clear from Table 5.2 that the model captures well the "clash" behaviour of the tropical

[*]These are often referred to in the literature as "clash probabilities".

cyclones as all tests on each paired region with the exception of Taiwan & Philippines, do not reject the null hypothesis at the given level of significance $\alpha = 0.05$. It is no surprise that the joint region of Taiwan & Philippines is rejected as both individual locations landfall rates are underestimated, as is the joint region.

The same analysis is conducted but changing the value of the parameter k from 10 to $k = 20$. It is expected that the results will vary but it is hoped that the results will not vary too much. If the results are completely different then this means that the model is highly sensitive to the model parameter k and this introduces problems when using the model as a very accurate estimate of this parameter would be required. The results for the single location landfall rates is given in Table 5.3

Table 5.3: Table of results for hypothesis test 5.2 with $k = 20$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Japan	2.0533	3.6666	1.4227	R
Hong Kong	0.2133	0.4333	0.4439	R
Shanghai	0.0266	0.0333	0.1622	A
Taiwan	0.92	2.15	0.9118	R
Philippines	4.08	8.2203	1.784	R

As observed from the table of results, the model fails to reproduce the correct landfall rates for Japan and Hong Kong that was captured by the model when the value of the parameter $k = 10$. The model again, as when $k = 10$, fails to correctly capture the landfall rates for Taiwan and Philippines and again underestimates them by roughly half as before. However, on balance, there is no "dramatic" change in the mean landfall rates from the change in the value of k from 10 to 20, despite the negative test results. This indicates that a value of ~ 10 is a good value for the parameter k as it captures the general behaviour of the cyclone tracks. Values larger than this can result in the loss of the local behaviour exhibited in the data. The results of the paired locations for the model with $k = 20$ is given in Table 5.4.

Table 5.4: Table of results for hypothesis test 5.2 with $k = 20$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Taiwan & Philippines	0.4133	0.7667	0.5717	R
Taiwan & Japan	0.1067	0.1	0.3108	A
Philippines & Hong Kong	0.16	0.2167	0.404	A
Philippines & Japan	0.24	0.25	0.5157	A

Similar to the model run with $k = 10$, it correctly estimates the landfall rates of the joint regions

detailed in Table 5.4 with the exception of Taiwan & Philippines. Note that for the joint region of Taiwan & Shanghai no tracks were simulated with $k = 20$. This confirms that the value of $k = 10$ is a more suitable value.

It is important not only to test the mean landfall rates but also to confirm if the model accurately captures the variability of the landfall rates. This is done by an F-test. Full details of this statistical test is given in [24], an outline of the test is given here. A two-sided F-test checks to see if the variances of two populations are equal or not. This corresponds to the following null and alternative hypotheses,

$$H_0 : \sigma^2_{\text{sim}} = \sigma^2_{\text{p}} \qquad \text{vs.} \qquad H_{\text{a}} : \sigma^2_{\text{sim}} \neq \sigma^2_{\text{p}} \tag{5.4}$$

where σ^2_{p} is the variance of landfall rates calculated from the data, and σ^2_{sim} is the variance of landfall rates calculated from the simulation. The test statistic for this test is given by $\mathbf{F} = S^2_{n_1}/S^2_{n_2}$, where $S^2_{n_1}$ and $S^2_{n_2}$ are the sample variances. The test is conducted at a level of significance of $\alpha = 0.05$. For the two-sided test carried out here, the null hypothesis is rejected if,

$$\mathbf{F} < F_{1-\alpha/2,N_1-1,N_2-1} \quad \text{or} \quad \mathbf{F} > F_{\alpha/2,N_1-1,N_2-1} \tag{5.5}$$

where F_{α,N_1-1,N_2-1} corresponds to the critical value of the F-distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom, N_1 and N_2 are the respective sample sizes, and a level of significance α. In essence, the more $\mathbf{F}$ deviates from 1, the stronger the evidence is that the null hypothesis should be rejected.

The results of this test for $k = 10$ with $\alpha = 0.05$ is given in Table 5.5. Note in Table 5.5, $S^2_{n_2}$ refers to the sample variance from the data and $S^2_{n_1}$ refers to the sample variance from the simulation.

Table 5.5: Table of results for hypothesis test 5.4 with $k = 10$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$S^2_{n_1}$	$S^2_{n_2}$	TR
Japan	3.0929	3.1555	A
Hong Kong	0.4295	0.4455	A
Shanghai	0.2151	0.0322	A
Taiwan	1.4425	1.6941	A
Philippines	3.3484	11.2933	R
Taiwan & Philippines	0.6317	0.7122	A
Taiwan & Japan	0.1993	0.1233	A
Taiwan & Shanghai	0.0263	0.01638	A
Philippines & Hong Kong	0.2389	0.2031	A
Philippines & Japan	0.4804	0.3208	A

From the results in Table 5.5, it is clear that the model captures well the variability among the landfall rates with the exception of the Philippines in which the variance is underestimated.

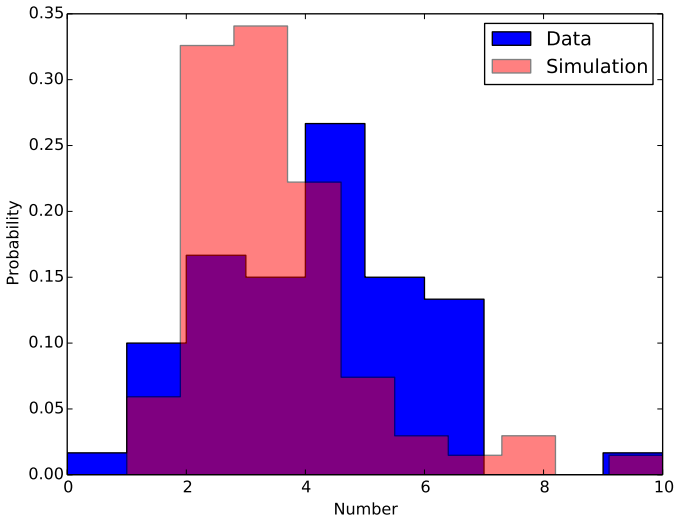


Figure 5.2: Comparison of the distribution of landfall rates of data against simulation for Japan.

Chapter 6

Hazard Assessment

In this chapter, the model discussed and analysed previously is now applied to hazard and risk assessment as per the aim of this thesis. Important figures and statistics for insurance and reinsurance companies are considered, for example return periods. Also, the impact of an ENSO* event is considered and the potential impact that they could have on the behaviour of the tropical cyclone tracks.

6.1 Clash Probabilities

There are events both natural and human catastrophes that are so large that they can cause enormous levels of losses both financial and social, these are known as mega-catastrophes. Though definitions vary, a mega-catastrophe typically generates more than $\$50$bn of direct and indirect losses, and very often occurs when a single peril triggers associated damage-inducing events [3]. In the case of a tropical cyclone not only is damage inflicted by the wind intensity but also from associated triggered events such as storm surge, rainfall and tornadoes may even develop over land.

As mentioned in Chapter 5, "clash" probabilities are of immense importance for actuarial use in reserving and pricing against large-scale destructive events. Clash losses, or losses impacting several areas of insurable risk simultaneously, often accompany a mega-catastrophe [3]. The direct and indirect losses that these events can potentially have can result in insurance, reinsurance companies and governments having large liability costs and as mentioned in the literature survey in Chapter 2 it is expected that these larger, more destructive storms are to become more frequent.

As outlined in Chapter 5, Table 6.1 shows the clash probabilities of the defined joint areas of importance with $k = 10$ and calculated over a 75 year simulation.

Table 6.1 shows that in fact, a tropical cyclone affecting both Taiwan and the Philippines is very likely in any given year with a probability of ~ 0.49. This has implications for both insurers and re-

*El Niño Southern Oscillation.

Table 6.1: Clash probabilities calculated over a sample of 75 years.

Location	Clash Probability	Standard Deviation
Taiwan & Philippines	0.4933	0.7948
Taiwan & Japan	0.1733	0.4464
Taiwan &Shanghai	0.0266	0.1622
Philippines & Hong Kong	0.24	0.4888
Philippines & Japan	0.3733	0.6931

insurers. If the tropical cyclones affecting both of these regions turn out to be strong cyclones, e.g. Category $3+$, may result in devastating losses. However, there is also relatively high probabilities that a tropical cyclone will affect the joint regions of Philippines and Hong Kong and Philippines and Japan in any given year. It is important to note that the variability of these probabilities is high with standard deviations shown in Table 6.1, confidence intervals can be easily calculated.

In Section 6.4.2, comparisons are drawn between the clash probabilities calculated here and clash probabilities where the model is run under ENSO conditions.

6.2 Wind Profile

So far only landfall rates have been considered. These landfalls are based on the centre of the tropical cyclone making landfall. However, the centre of a storm may not directly pass over a location of interest but it can still affect the specific location by strong winds. Therefore, to gauge a more accurate picture of potential damage that could be caused by the tropical cyclone, it is necessary to consider the wind distribution or the wind footprint from the centre of the storm to the radius of gale force winds.

As a first approach to modelling the wind distribution a modified Rankine vortex of the form,

$$v(r) = \begin{cases} v_{\text{max}} \cdot \frac{r}{r_{\text{max}}} & \text{if } 0 \leq r < r_{\text{max}} \\ v_{\text{max}} \cdot \left(\frac{r}{r_{\text{max}}}\right)^{-x} & \text{if } r \geq r_{\text{max}} \end{cases} \tag{6.1}$$

where the wind speed $v(r)$ (in km/h) at a distance r from the centre of the storm is given as a function of r, the radius at which the maximum wind speed is attained r_{max} and the maximum wind speed v_{max} (in km/h) for some $x \in [0, 1]$ [21]. The exponent x determines the shape of the wind distribution outside of the radius of the maximum wind speed. Reflecting the empirical nature of the model instead of determining x directly, it is appropriate to determine the radius of maximum wind speed r_{max} and the gale-force radius r_{gale}, where r_{gale} is the maximum distance from the centre of

the storm at which the wind speeds are of at least gale force, i.e. $v_{\text{gale}} = 63\text{km/h}^{\dagger}$ [21]. Assuming that r_{max} and r_{gale} are known, x can be calculated as follows:

$$x = \frac{\ln(v_{\text{max}}) - \ln(63)}{\ln(r_{\text{gale}}) - \ln(r_{\text{max}})}. \tag{6.2}$$

In the data set considered in this project, the measurements of r_{gale} and r_{max} are relatively very few. These measurements were only recorded for tropical cyclones from roughly 2002 on with some exceptions. This reduces the available data set considerably. As a result, r_{max} and r_{gale} are calculated as empirical functions of wind speed using the following relations:

$$r_{\text{max}} = \frac{v_{\text{max}}}{a \cdot \exp(b v_{\text{max}})} \tag{6.3}$$

$$r_{\text{gale}} = c \cdot \ln(v_{\text{max}}) - d \tag{6.4}$$

where the coefficients a, b, c, d are determined by least-squares regression [21]. The coefficients are calculated independently for Category 1 cyclones, Category 2 & 3 cyclones and Category 4 & 5 cyclones, according to the Saffir-Simpson Hurricane Wind Scale. Then for each cyclone track, the values of r_{max} and r_{gale} are calculated as functions of v_{max} in (6.3) and (6.4) for the values of a, b, c, d corresponding to the maximum wind speed attained along the respective track.

Due to the very small set of data that can be used to fit the expressions (6.3) and (6.4), the author concedes that the fits may be of poor quality and more data is required for better fits.

As a first approximation of the wind distribution the modified Rankine vortex provides a reasonable estimate. However, the modified Rankine vortex neglects to take into account important factors of the shape of a tropical cyclone, for example the storms latitude and the effect of the Coriolis effect [‡]. To incorporate these factors, an alternative model for tropical cyclone wind profiles is described, named the λ model.

The wind velocity at a radius r is given by the λ model as a function of minimum pressure, latitude and either R_{th} the radius of V_{th} which is a threshold wind speed or by the radius of maximum wind speed (RMW), as follows,

$$V(r) = F(P_{\text{min}}, \varphi, R) \tag{6.5}$$

where P_{min} is the minimum pressure, φ is the latitude of the centre of the cyclone and R is either R_{th} or RMW as described [31]. The tangential velocity of a radius r at the top of the boundary

[†] It is assumed that $v_{\text{max}} > v_{\text{gale}}$ which implies $r_{\text{max}} < r_{\text{gale}}$.

[‡] A deflection of moving objects when the motion is described relative to a rotating reference frame, for example the Earth.

layer is then given in full form by the following relation,

$$V = \sqrt{2\alpha(P_{\text{env}} - P_{\text{min}})} \cdot \sqrt{\frac{2\lambda^2}{r^2}(1 - e^{-\frac{r^2}{2\lambda^2}}) - e^{-\frac{r^2}{2\lambda^2}}} - \frac{1}{2}fr \tag{6.6}$$

where α is the specific volume, P_{env} is the ambient pressure in the environment, f is the Coriolis parameter, P_{min} is the minimum surface pressure and λ is the single shape parameter of the model [31].

Note that the specific volume $\sim 1\text{m}^3/\text{kg}$, P_{env} for the western North Pacific $\sim 1010\text{hPa}$[§] and the Coriolis parameter is given by $f = 2\Omega \sin\varphi$ where Ω is the Earth's angular velocity $\sim 7.292115 \times 10^{-5}$ radians per SI second and φ is the latitude of the centre of the cyclone.

Finally, the shape parameter λ is given by

$$\lambda = \begin{cases} \frac{R_{th}(fR_{th}+2V_{th})}{4\sqrt{\alpha(P_{\text{env}}-P_{\text{min}})}} & \text{if } R = R_{th} \\ \frac{1}{1.89}\text{RMW} & \text{if } R = \text{RMW} \end{cases} \tag{6.7}$$

where R is as described in Equation 6.5 [31].

In the data set considered for this project, the observations of minimum pressure P_{min} are very few and so an empirical relationship can be used to approximate the minimum pressure from the maximum wind speed. One such relationship for the western North Pacific basin is given by [2] and is a simple relationship as follows,

$$v_{\text{max}} = 3.4(1010 - p_{\text{min}})^{0.644} \tag{6.8}$$

where v_{max} is the 1-min mean wind speed at a 10-m elevation in ms^{-1}, p_{min} is the minimum central pressure and 1010hPa is a typical ambient pressure to use in the western North Pacific and is widely used in the literature [2]. This approximation was derived from the raw data without first binning to remove a bias toward the more frequently sampled lower intensities, and this led [16] to derive a new version after binning the original data [12], defined as follows,

$$\text{MSLP} = 23.286 - 0.483V_{\text{srm}} - \left(\frac{V_{\text{srm}}}{24.254}\right)^2 - 12.587S - 0.483\varphi + P_{\text{env}} \tag{6.9}$$

where MSLP is the minimum sea level pressure, V_{srm} is the storms relative maximum surface wind speed adjusted for storm motion, S is the normalised size parameter (discussed below), φ and P_{env} are as described before [16]. It has been found that the storms motion or translation speed has a small effect on the MSLP, so in order to account for this, the relative maximum surface wind speed is estimated by $V_{\text{srm}} \approx V_{\text{max}} - 1.5c^{0.63}$, where c is the translation speed of the storm [2]. The final

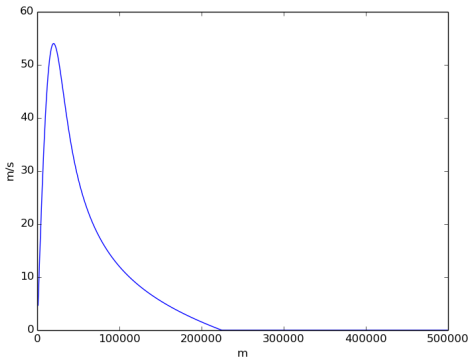
[§]Same value as used in [12]

parameter to estimate is $S = V_{500}/V_{500c}$ where

$$V_{500c} = V_{\text{max}} \left(\frac{R_{\text{max}}}{500} \right)^{x} \tag{6.10}$$

i.e. based on a modified Rankine vortex as described earlier and V_{500} is estimated by calculating the azimuthal mean of the wind speeds in the annulus of 400-600kms [16]. For the purposes of this analysis, it suffices to use the simple empirical relationship given in Equation 6.8.

For the purposes of demonstration, the λ employed to reconstruct the wind profile for a given trajectory. Using the λ model, the wind profile with a maximum velocity of 152km/h translates into



(a) Lambda model

Figure 6.1: Wind profile constructed using Lambda model.

960hPa using the relationship 6.8 at a latitude of 24° with a radius of maximum wind speed (RMW) 20km.

On investigation of the λ model, it is observed that when the RMW is small, the λ model fails to capture the wind profile before the maximum wind speed. For example, if the reconstruction of the wind profile using the λ model is down with RMW of 2km *ceteris paribus*, and zoom-in on the plot to the first few hundred metres this is demonstrated in Figure 6.2.

In order to apply the λ model to the output of the tropical cyclone track model described in Chapter 4, an estimate of the RMW is required. For the estimate of RMW of the simulated tropical cyclones, the exponential fit used in the modified Rankine vortex is applied, (Equation 6.3).

In both of these models to reconstruct the wind profiles, the underlying assumption is that the cyclones are symmetric. It is well understood that despite tropical cyclones having a strong defined shape, they are in general, asymmetric.

In the following section, both wind profile models are applied to calculate return periods and the differences between the two methods of calculation are highlighted.

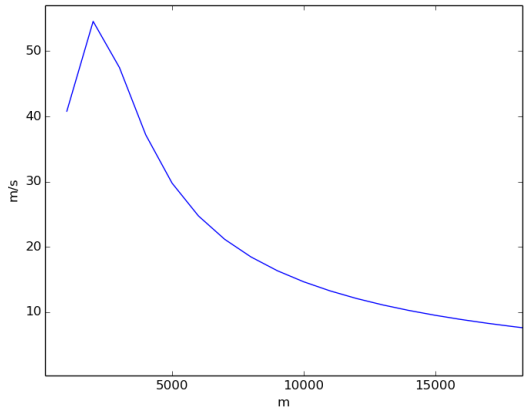


Figure 6.2: Wind profile constructed using Lambda model with RMW = 2km, zoomed.

6.3 Return Periods

A return period is theoretically the inverse of the probability that an event will be exceed a given threshold in any given year. Physically, this is the inverse of the expected number of occurrences in any given year.

For the purposes of hazard assessment for this project, return periods are calculated for specific locations or points of interest in the observation window for different wind speed thresholds. In order to calculate the hazard at a particular point of interest t_0 for each tropical cyclone, the distance from t_0 to all points of measurement of a tropical cyclone is calculated [21]. By applying both wind profile models described in Section 6.2, the wind speeds with which the location of interest t_0 is affected by each step of the tropical cyclone is recorded. The maximum of these wind speeds is then regarded as the storms 'wind impact' on t_0 [21]. Any wind speeds of less than gale force are discarded. Then, the expected return period for any given wind speed threshold v_0 can be estimated from those wind speeds by dividing the number of years m by the number of impacts on t_0 greater than or equal to v_0 [21],

$$\widehat{\mathrm{RP}}(v_0) = \frac{m}{\#\{\text{wind impacts} \geq v_0\}}, \tag{6.11}$$

where $\#$ represents the cardinality of the set.

In Table 6.2, the calculated probabilities that a wind speed will exceed the stated thresholds over 100 years of simulation are given using the modified Rankine vortex as a model for the wind profile. The return periods can be obtained by taking the reciprocal of these probabilities. Note that the 100 year simulation is based on the entirety of the data set.

Using the modified Rankine vortex as a model for the wind profile when calculating the exceedance probabilities, the values appear to be relatively low for the given thresholds. In fact at higher thresholds, the exceedance probabilities are zero which, given the locations of interest, does not seem plausible. In Table 6.3, the exceedance probabilities are calculated using the λ model.

Table 6.2: Probabilities of exceedance in any given year over 100 year simulation where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h
Toyko	0.12	0.04
Seoul	0.13	0.07
Osaka	0.12	0.04
Hong Kong	0.12	0.06
Shanghai	0.12	0.06
Beijing	0.12	0.07
Taipei	0.13	0.06
Guam	0.13	0.04
Manila	0.12	0.06

Table 6.3: Probabilities of exceedance in any given year over 100 year simulation where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h	WT: 180km/h	WT: 250kmph
Tokyo	0.16	0.16	0.12	0.11
Seoul	0.17	0.14	0.09	0.07
Osaka	0.18	0.18	0.1	0.1
Hong Kong	0.15	0.15	0.09	0.07
Shanghai	0.2	0.19	0.09	0.05
Beijing	0.13	0.13	0.11	0.04
Taipei	0.2	0.18	0.14	0.11
Guam	0.18	0.15	0.1	0.09
Manila	0.19	0.11	0.09	0.09

By applying the λ model to reconstruct the wind profile, more realistic exceedance probabilities can be calculated, even at higher thresholds. These values are of important use in the field of actuarial science, in particular in cat models, reserving and premium calculations. Using these values, insurance and reinsurance companies can compute an exceedance probability (EP) curve. EP curves visualise the probabilities of any given financial loss threshold being exceeded.

An important note to make is that return periods calculated in this way assumes that the conditions in which the tropical cyclones are simulated remains the same over the lifetime of the simulation. These issues have been highlight in [11], landfall rates cannot be assumed to be constant in time due to the variability of influences [11].

A possible analysis to use only data from a period that where tropical cyclone tracks exhibits similar behaviour due to a certain influence. One such period could be when tracks are influenced by ENSO, this analysis is carried out in the next section.

6.4 ENSO

In this section, the model is run under ENSO conditions both El Niño and La Niña. ENSO is the effects of a band of sea surface temperatures which are either above average warm or cold for significant periods of time. El Niño-Southern Oscillation events are a coupled ocean-atmosphere phenomenon. El Niño involves warming of tropical Pacific surface waters from near the International Date Line to the west coast of South America, weakening the usually strong SST gradient across the equatorial Pacific, with associated changes in ocean circulation [27]. In order to run the model under these conditions, the data must be subset into El Niño and La Niña periods.

The Oceanic Niño Index (ONI) has become the de-facto standard that NOAA uses for identifying ENSO events in the tropical Pacific. ENSO events are defined as 5 consecutive overlapping 3-month periods $|0.5^{\circ}|$ anomaly. The ONI is broken down into Weak (with a 0.5 to 0.9 SST anomaly), Moderate (1.0 to 1.4), Strong (1.5 to 1.9) and Very Strong (≥ 2.0) events [¶]. For the purposes of this project, anomalies of greater or equal to 1 are used.

Table 6.4: Years used ENSO conditions.

El Niño years	La Niña years
57-58, 63-64, 65-66, 72-73, 82-83, 86-87, 87-88, 91-92, 97-98, 02-03	55-56, 70-71, 73-74, 75-76, 88-89, 98-99, 99-00

6.4.1 Genesis & Track Shape

During El Niño and La Niña periods the expected number of tropical cyclones increases and decreases receptively with,

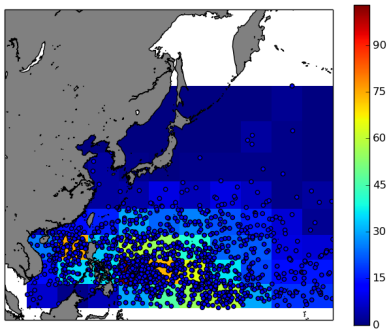
$$N = \mathbb{E}\left[\text{mean } \# \text{ of storms per El Niño year}\right] = \frac{554}{19} = 29.158$$

$$N = \mathbb{E}\left[\text{mean } \# \text{ of storms per La Niña year}\right] = \frac{357}{13} = 27.462$$

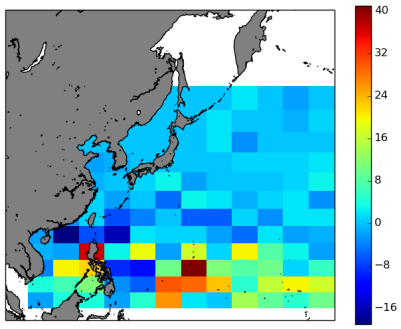
compared with a mean number of tracks of 28.5 over the entire data set. This increase and decrease in the number of tropical cyclones formed is directly related to the change in SST experienced during ENSO periods.

Figure (6.3(b)) shows that there is a shift eastward of the high density of the genesis points under El Niño conditions and conversely Figure (6.3(c)) shows that the highest density of genesis points moves slightly westward in the observation window under La Niña conditions. This is expected.

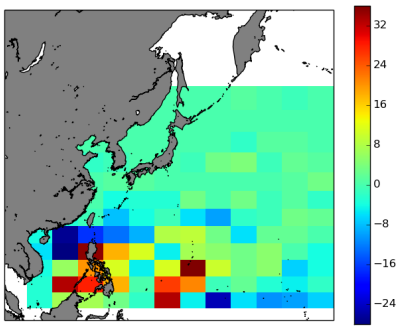
[¶] As defined by NOAA.



(a) Simulation with entire data set.



(b) Anomaly simulation under El Niño conditions



(c) Anomaly simulation under La Niña conditions

Figure 6.3: Comparison of genesis locations under El Niño and La Niña conditions.

In a study carried out by [30] on how strong ENSO events affect tropical storm activity, it is noted that there is a southeastward (northwestward) shift of tropical cyclone activity in El Niño (La Niña) years [30].

In general, the shape of the tropical cyclones are convex to the continent; they tend follow an elongated 'C' shape. The shape of these tracks is generally explained by the sum of two terms, namely environmental steering, which is the dominant of the two terms and the "beta drift". The environmental steering can be broken down into two components the flow of the cyclone itself, and the local ambient flow around the storms location. Tropical cyclone motion may be represented to first-order simply as the advection of the storm by the local environmental flow [19]. Most tropical cyclones form on the side of the subtropical ridge closer to the equator and will on average tend to drift towards the pole and westward, forming the base of this 'C' shape. This tendency is known as beta drift. This is due to the interaction between the hurricane vortex and the Earth's vorticity gradient, namely the beta effect (variation of the Coriolis parameter with latitude) [19]. For a more mathematical rigorous understanding of tropical cyclone track shapes, see [14], [19].

When the subtropical ridge position shifts due to El Niño, so does the preferred tropical cyclone tracks. Areas west of Japan and Korea tend to experience much fewer tropical cyclone landfalls during El Niño periods. During El Niño periods, the break in the subtropical ridge tends to lie near 130°E, which would favour the Japanese archipelago [32]. During La Niña periods, the formation of tropical cyclones, along with the subtropical ridge position, shifts westward across the western Pacific ocean, which increases the landfall threat to China [32]. Conversely, when the subtropical ridge shifts westward across the western Pacific ocean as a result of La Niña shorter tropical cyclones are observed. The shift westward may increase the landfall threat to China [32].

6.4.2 Landfall Rates & Clash Probabilities

In this section, comparison and differences are highlighted between the model run under "normal conditions" and the model run under ENSO conditions in terms of landfall rates and clash probabilities. The statistical test carried out is to analyse the mean landfall rate differences (the same test carried out in hypothesis 5.2), under the different scenarios. Here μ_{p} is taken as the mean landfall rates calculated by simulation and shown in Tables 5.1 & 5.2. In Table 6.5, the summary of results for this test comparing the model run under normal conditions against El Niño conditions.

Table 6.5: Table of results for hypothesis test 5.2 with $k = 10$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Japan	2.5067	3.2933	1.1434	R
Hong Kong	0.1467	0.3733	0.3562	R
Shanghai	0.0667	0.1333	0.2511	R
Taiwan	0.5333	1.1333	0.6844	R
Philippines	3.1067	4.28	1.9561	R
Taiwan & Philippines	0.24	0.52	0.4888	R
Taiwan & Japan	0.04	0.16	0.1973	R
Taiwan & Shanghai	0.0267	0.0267	0.1622	A
Philippines & Hong Kong	0.1067	0.2267	0.3108	R
Philippines & Japan	0.24	0.4133	0.6333	R

From Table 6.5, all with the exception of the joint region Taiwan & Shanghai, reject the null hypothesis that the means are equal. We can conclude that there is a statistically significant difference at a $\alpha = 0.05$ level of significance. It is observed that the mean landfall rates under El Niño are smaller across the entire range of locations considered. This result conforms with the result that the points of genesis of the tropical cyclones shift eastward discussed in Section 6.4.1. Next, the same test is conducted comparing the model run under normal conditions against La Niña conditions, Table 6.6 gives the summary of results for this test.

Table 6.6: Table of results for hypothesis test 5.2 with $k = 10$ and $\alpha = 0.05$, TR - test result; A - the null hypothesis is not rejected otherwise it is rejected (R).

Location	$\overline{X}_n$	μ_{p}	S_n	TR
Japan	1.2133	3.2933	1.0173	R
Hong Kong	0.0667	0.3733	0.2511	R
Shanghai	0.04	0.1333	0.1973	R
Taiwan	1.0667	1.1333	0.8751	A
Philippines	3.5733	4.28	1.9114	R
Taiwan & Philippines	0.4267	0.52	0.5736	A
Taiwan & Japan	0.08	0.16	0.2731	R
Philippines & Hong Kong	0.0667	0.2267	0.2511	R
Philippines & Japan	0.1733	0.4133	0.3811	R

In Table 6.6, firstly notice that the null hypothesis is not rejected in the case of Taiwan and the joint region of Taiwan & Philippines. This means that at a level of significance of $\alpha = 0.05$ the mean landfall rates for Taiwan and the joint region of Taiwan & Philippines are not significantly different in La Niña versus normal conditions. In all of the other locations the null hypothesis is rejected at the given level of significance and the mean landfall rates under La Niña are less than those under normal conditions. In the literature [30] [32], it is generally agreed that over La Niña periods landfall rates of locations in the south western area of the observation window are higher than during El Niño periods and under normal conditions.

For the analysis conducted here, it is observed that indeed it is the case that compared to El Niño periods have smaller mean landfall rates than under La Niña periods in the south western part of the observation, for example for Taiwan the mean landfall rates are 1.0667 and 0.5333 for La Niña and El Niño periods respectively. However, when La Niña periods and normal conditions are compared the mean landfall rates during the cooler years of SST are smaller. There are several possible reasons for this; one such explanation is the underlying data which the tracks under La Niña conditions. There are very few observations (i.e. historical tracks) in the sub-setted data, a total of 357 tracks. Another reason, also associated with the underlying data, is that a more sophisticated method for sub-setting the data into El Niño and La Niña periods could be used rather than sub-setting by years according to ONI.

6.4.3 Return Periods

As in Section 6.3, exceedance probabilities of given wind speed thresholds are calculated under ENSO conditions. Again, these are calculated by using both the modified Rankine vortex and the λ model to reconstruct the wind profiles. In Tables 6.7 & 6.8, the calculated exceedance probabilities by the modified Rankine vortex model are given under El Niño and La Niña conditions respectively.

Table 6.7: Probabilities of exceedance in any given year over 100 year simulation under El Niño conditions using modified Rankine vortex, where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h
Tokyo	0.05	0.03
Seoul	0.05	0.03
Osaka	0.06	0.03
Hong Kong	0.05	0.03
Shanghai	0.06	0.04
Beijing	0.05	0.03
Taipei	0.07	0.04
Guam	0.09	0.02
Manila	0.06	0.03

Table 6.8: Probabilities of exceedance in any given year over 100 year simulation under La Niña conditions using modified Rankine vortex, where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h
Tokyo	0.07	0.01
Seoul	0.05	0.02
Osaka	0.07	0.01
Hong Kong	0.08	0.01
Shanghai	0.08	0.02
Beijing	0.08	0.02
Taipei	0.08	0.02
Guam	0.09	0.02
Manila	0.09	0.03

Again as described in Section 6.3, although there are differences under ENSO conditions with each other and compared to normal conditions, the probabilities are relatively small. Instead, the same analysis is conducted using the λ model. Tables 6.9 & 6.10 give the corresponding exceedance probabilities under El Niño and La Niña conditions respectively.

Applying the λ model to calculated return periods under ENSO conditions highlights some interesting results. At the lower wind speed thresholds, namely 80km/h and 100km/h, the exceedance probabilities at all locations considered are higher under ENSO conditions. As expected, there is more variability at higher wind speed thresholds. In order to summarise the results, Table 6.11 shows the changes in the exceedance probabilities (at least $\pm 10\%$ change) against the *status quo* of the exceedance probabilities calculated under normal conditions given in Table 6.3.

An interesting location that Table 6.11 highlights is Beijing. According to this analysis, the exceedance probabilities at higher wind speed thresholds under ENSO conditions are smaller than

Table 6.9: Probabilities of exceedance in any given year over 100 year simulation under El Niño conditions using λ model, where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h	WT: 180km/h	WT: 250kmph
Tokyo	0.21	0.2	0.14	0.1
Seoul	0.22	0.18	0.12	0.08
Osaka	0.22	0.21	0.19	0.13
Hong Kong	0.2	0.17	0.09	0.08
Shanghai	0.24	0.21	0.14	0.08
Beijing	0.16	0.13	0.05	0.03
Taipei	0.26	0.23	0.14	0.1
Guam	0.26	0.26	0.19	0.15
Manila	0.22	0.2	0.16	0.12

Table 6.10: Probabilities of exceedance in any given year over 100 year simulation under La Niña conditions using λ model, where WT is the wind threshold.

Location	WT: 80km/h	WT: 100km/h	WT: 180km/h	WT: 250kmph
Tokyo	0.21	0.2	0.13	0.1
Seoul	0.19	0.16	0.12	0.09
Osaka	0.2	0.2	0.14	0.1
Hong Kong	0.23	0.19	0.14	0.07
Shanghai	0.24	0.22	0.17	0.11
Beijing	0.16	0.14	0.04	0.02
Taipei	0.27	0.26	0.22	0.17
Guam	0.26	0.25	0.23	0.2
Manila	0.27	0.25	0.19	0.13

under normal conditions. In general, in most locations and in particular at lower wind speed thresholds, exceedance probabilities are larger. This could explained by the fact that despite tropical cyclones during El Niño periods tend to start in the south eastern part of the observation window, the cyclones tend to be more intense, i.e. have higher maximum wind speeds/ smaller minimum pressure due the the increase in SST. Conversely, during La Niña periods the cyclogenesis tends to shift westward which will cause more of the locations of interest to be affected by the storm despite tropical cyclones during these periods tending to be weaker. Manila appears to have great sensitivity to ENSO periods with exceedance probabilities considerably larger than under normal conditions.

Another interesting comparison to draw is the changes in exceedance probabilities between El Niño and La Niña periods. On comparison it is found that generally the exceedance probabilities during periods of lower SST, i.e. La Niña periods, are greater than El Niño. This type of information is of significant importance to insurance and reinsurance companies when running scenarios on cat

models, in particular for calculating reserves as they can better prepare for potential loss making events differently depending on prevailing conditions.

Table 6.11: Changes in exceedance probabilities under El Niño and La Niña conditions compared to normal conditions, where + is a positive change, - is a negative change and = indicates no change (of at least $\pm 10\%$).

	WT: 80km/h		WT: 100km/h		WT: 180km/h		WT: 250kmph	
Location	El Niño	La Niña	El Niño	La Niña	El Niño	La Niña	El Niño	La Niña
Tokyo	+	+	+	+	+	=	=	=
Seoul	+	+	+	+	+	+	+	+
Osaka	+	+	+	+	+	+	+	=
Hong Kong	+	+	+	+	=	+	+	=
Shanghai	+	+	+	+	+	+	+	+
Beijing	+	+	=	=	-	-	-	-
Taipei	+	+	+	+	=	+	=	+
Guam	+	+	+	+	+	+	-	+
Manila	+	+	+	+	+	+	+	+

Chapter 7

Summary

This project implements a stochastic model for the simulation of synthetic tropical cyclone tracks in the western North Pacific. The implementation is discussed in detail in Chapter 4, including a flowchart of the algorithm implemented. In Chapter 5 a detailed model evaluation is given by verifying mean landfall rates and their variability. Subsequently, the model is applied to hazard assessment in Chapter 6 under different climatic scenarios. The results of the hazard assessment shows that during La Niña periods the exceedance probabilities and correspondingly the return periods tend to be higher than during other periods.

The model to simulate the synthetic tropical cyclones assumes that the genesis points are randomly distributed under an inhomogeneous Poisson process based on historical data. A better approach may be to use a more general process like a Gibbs process or possibly to condition a genesis point being created on SST. Also the division of the historical tropical cyclones based on track shape into classes is not an approach that is immediately transferable to other basins. At each step including the initial step when the characteristics are determined, it is assumed that new tropical cyclones will have the same behaviour as historical cyclones. This may not be the case, as global temperatures increase, this could influence the track shapes that may not be taken in to account in this approach. All these assumptions can and should be questioned in order to give a more flexible model yet still statistically justified. Future research may include an investigation in to possible correlation between the characteristics in the model.

Possible future investigation into the parameter k in the model may explain the underestimation of mean landfall rates over La Niña periods. The parameter k determines the number of nearest neighbours to use in the construction of the local distribution functions. When the data is subsetted into La Niña periods, there is relatively fewer tracks in the data, therefore with the parameter $k = 10$, found to be a suitable parameter value in Chapter 5 may not be suitable in instances where there is a smaller about of data. In the case of the sub-setted La Niña data with $k = 10$, the local distribution functions may be including points that are some distance from the current point but are

still within the closest 10 neighbours. This may result in simulated tracks not exhibiting the correct behaviour. One possible solution could be to decrease the size to k so that the local distributions generated are truly local, however this could result in tropical cyclones getting "stuck", i.e. the storm has zero translation speed and cannot sample a positive change, also discussed in Chapter 5. This inflexibility is one of the short falls of this model. One option to over come this problem is to impose a parametric distribution for the direction characteristic that could be location dependant.

When reconstructing the wind profile of the tropical cyclones using the models described in Chapter 6, the assumption is that the tropical cyclones are symmetric. One approach to resolve this to some degree is, for example to apply the λ model in for each quadrant of the tropical cyclone. In order to apply this, the model requires wind speed measurements and radius measurements for each quadrant. In the JTWC data set, some (although currently very few) of these measurements have been recorded for a threshold wind speed, this would imply that in the estimation of λ instead of using RMW, R_{th} must be used. With more data a model could be fit to estimate the R_{th} at the threshold wind speed in each of the quadrants in order to obtain a more accurate wind profile. Another interesting result from this work is the sensitivity of the return periods to the wind profile. The exceedance probabilities vary significantly when they are calculated using both the modified Rankine vortex and λ models. Further research and analysis is required to investigate the impacts of several wind profile constructions on return period calculations.

In Chapter 6, the method for calculating the exceedance probabilities (return periods) may produce the same estimate for different threshold wind speeds. One such solution to this problem is to impose a strictly monotonic parametric distribution function $F(v)$ (e.g. Gamma-, Weibull- or other extreme value distributions) could be fitted to the calculated wind impacts [21]. Then the estimate $\widehat{\text{RP}}(v_0)$ for the expected return period of threshold wind speed v_0 could be calculated as

$$\widehat{\text{RP}}(v_0) = \frac{m}{n_{t_0}(1 - F(v_0))}$$

where m is the number of years and n_{t_0} is the number of wind impacts [21]. However, the underlying assumption with the calculation of return periods is on the conditions over the period of the data that is used to simulate the track. A warming climate may have an impact not only on the track shapes but also the cyclogenesis. More accurate return periods could be calculated if the model had some coupled model that could simulate a warming or cooling of SST. Another area of possible research could be to calculate spatial distributions of return periods for certain locations. Due to the scale of some tropical cyclones and some large metropolitan areas, obtaining spatial distributions of such areas will provide a more in-depth understanding of the wind impacts that these locations may experience. This demonstrates that area continues to provide a fertile ground for future research.

Appendix

The theorems stated here were used in this thesis but not directly stated:

Theorem A.1 Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with mean $\mathbb{E}[X_i] = \mu \in (-\infty, \infty)$ and finite variance $V(X_i) = \sigma^2 > 0$. Let:

$$S_n = \sum_{i=1}^{n} X_i$$

Then:

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } n \to \infty$$

i.e. converges in distribution to a standard normal.

Theorem A.2 Slutsky's Theorem

Let $\{X_n\}$, $\{Y_n\}$ be sequences of scalar/vector/matrix random elements. If X_n converges in distribution to a random element X and Y_n converges in probability to a constant c, then

$X_n + Y_n \xrightarrow{d} X + c$;

$X_n Y_n \xrightarrow{d} cX$;

$X_n / Y_n \xrightarrow{d} X/c$, *provided that c is invertible*

where $\xrightarrow{d}$ denotes convergence in distribution.

The Saffir-Simpson Hurricane Wind Scale was referred to in this thesis but not directly stated:

Saffir-Simpson Hurricane Wind Scale		
Category	Sustained Winds	Description
1	64-82 kt 119-153 km/h	Very dangerous winds will produce some damage
2	83-95 kt 154-177 km/h	Extremely dangerous winds will cause extensive damage
3 (major)	96-112 kt 178-208 km/h	Devastating damage will occur
4 (major)	113-136 kt 209-251 km/h	Catastrophic damage will occur
5 (major)	137 kt or higher 252 km/h or higher	Catastrophic damage will occur

Table 1: Source: National Oceanic Atmospheric Administration (NOAA)

References

[1] *Insurance Industry Catastrophe Management Practices, American Academy of Actuaries*.

[2] Gary D Atkinson and Charles R Holliday. Tropical cyclone minimum sea level pressure/maximum sustained wind relationship for the western north pacific. *Monthly Weather Review*, 105(4):421–427, 1977.

[3] Erik Banks. *Catastrophic risk: analysis and management*. John Wiley & Sons, 2005.

[4] Jesse Dorrestijn, Daan T Crommelin, A Pier Siebesma, Harmen JJ Jonker, and Christian Jakob. Stochastic parameterization of convective area fractions with a multicloud model inferred from observational data. *Journal of the Atmospheric Sciences*, (2014), 2014.

[5] J Steinebach EL Lehmann, JP Romano. Testing statistical hypotheses. *Metrika*, 64(2):255–256, 2006.

[6] Kerry Emanuel, Sai Ravela, Emmanuel Vivant, and Camille Risi. A statistical deterministic approach to hurricane risk assessment. *Bulletin of the American Meteorological Society*, 87(3):299–314, 2006.

[7] C.B. Field, Intergovernmental Panel on Climate Change. Working Group I., IPCC., and Intergovernmental Panel on Climate Change. *Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Summary for Policymakers : a Report of Working Groups I and II of the IPCC*. Published for the Intergovernmental Panel on Climate Change, 2012.

[8] PN Georgiou, A Go Davenport, and BJ Vickery. Design wind speeds in regions dominated by tropical cyclones. *Journal of Wind Engineering and Industrial Aerodynamics*, 13(1):139–152, 1983.

[9] EJ Gumbel. On the frequency distribution of extreme values in meteorological data. *Bull. Am. Meteorol. Soc*, 23(3):95, 1942.

[10] Timothy M Hall and Stephen Jewson. Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus A*, 59(4):486–498, 2007.

[11] GJ Holland. Misuse of landfall as a proxy for atlantic tropical cyclone activity. *Eos*, 88(36):349–356, 2007.

[12] Greg Holland. A revised hurricane pressure-wind model. *Monthly Weather Review*, 136(9):3432–3445, 2008.

[13] Greg J Holland. An analytic model of the wind and pressure profiles in hurricanes. *Monthly weather review*, 108(8):1212–1218, 1980.

[14] Greg J Holland. Tropical cyclone motion: Environmental interaction plus a beta effect. *Journal of the Atmospheric Sciences*, 40(2):328–342, 1983.

[15] MK James and LB Mason. Synthetic tropical cyclone database. *Journal of waterway, port, coastal, and ocean engineering*, 131(4):181–192, 2005.

[16] John A Knaff and Raymond M Zehr. Reexamination of tropical cyclone wind-pressure relationships. *Weather and forecasting*, 22(1):71–88, 2007.

[17] Thomas P Leahy. Modelling the generation of synthetic tropical cyclone tracks: A review. 2015, unpublished.

[18] D. Crommelin N. Verheul. Data-driven stochastic representations of unresolved features in multiscale models. 2015.

[19] G.R. North, J.A. Pyle, and F. Zhang. *Encyclopedia of Atmospheric Sciences*. Number v. 1-6. Elsevier Science, 2014.

[20] Jonas Rumpf, Helga Weindl, Peter Höppe, Ernst Rauch, and Volker Schmidt. Stochastic modelling of tropical cyclone tracks. *Mathematical Methods of Operations Research*, 66(3):475–490, 2007.

[21] Jonas Rumpf, Helga Weindl, Peter Höppe, Ernst Rauch, and Volker Schmidt. Tropical cyclone hazard assessment using model-based track simulation. *Natural hazards*, 48(3):383–398, 2009.

[22] Karl Sigman. Lecture notes on Inverse Transform Method. `http://www.columbia.edu/~ks20/4404-Sigman/4404-Notes-ITM.pdf`, 2010.

[23] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[24] George W Snedecor and Witiiam G Cochran. Statistical methods 8. ed. *Iowa State Univ*, 1989.

[25] Dietrich Stoyan and Helga Stoyan. *Fractals, random shapes, and point fields: methods of geometrical statistics*. Wiley Chichester, 1994.

[26] Kendall W.S. Mecke J. Stoyan, D. *Stochastic Geometry and its Applications*. Wiley, 1995.

[27] P.D. Jones P. Ambenje R. Bojariu D. Easterling A. Klein Tank D. Parker F. Rahimzadeh J.A. Renwick M. Rusticucci B. Soden Trenberth, K.E. and P. Zhai. *Observations: Surface and Atmospheric Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)].* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.

[28] Peter J Vickery and Lawrence A Twisdale. Wind-field and filling models for hurricane wind-speed predictions. *Journal of Structural Engineering*, 121(11):1700–1709, 1995.

[29] PJ Vickery, PF Skerlj, and LA Twisdale. Simulation of hurricane risk in the US using empirical track model. *Journal of structural engineering*, 126(10):1222–1237, 2000.

[30] Bin Wang and Johnny CL Chan. How strong ENSO events affect tropical storm activity over the western North Pacific. *Journal of Climate*, 15(13):1643–1658, 2002.

[31] Shuai Wang, Ralf Toumi, Arnaud Czaja, and Adrian Van Kan. An Analytic Model of Tropical Cyclone Wind Profiles. *Quarterly Journal of the Royal Meteorological Society*, 2015.

[32] MC Wu, WL Chang, and WM Leung. Impacts of El Niño-southern oscillation events on tropical cyclone landfalling activity in the western North Pacific. *Journal of Climate*, 17(6):1419–1428, 2004.