

**LONG-TERM REGIONAL SIMULATION OF TROPICAL  
CYCLONES USING A GENERALIZED STOCHASTIC  
EMPIRICAL STORM MODEL.  
A CASE STUDY IN THE WESTERN NORTH PACIFIC**

**Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben;  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op  
dinsdag, 03 februari 2015 om 12:30 uur

door

**NGUYEN Binh Minh**

Master of Science in Water Science and Engineering  
geboren te Hanoi, Vietnam

Dit proefschrift is goedgekeurd door de promotoren:

Prof.dr.s.ir. J.K. Vrijling

Prof.dr.ir. P.H.A.J.M. van Gelder

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.s.ir. J.K. Vrijling,	Technische Universiteit Delft, promotor
Prof.dr.ir. P.H.A.J.M. van Gelder,	Technische Universiteit Delft, promotor
Prof.dr.ir. M. Kok,	Technische Universiteit Delft
Prof.dr.ir. N.C. van de Giesen,	Technische Universiteit Delft
Prof.dr.ir. R.D.J.M. Steenbergen,	Ghent University & TNO
Dr.ir. S. Caires,	Deltares
Dr. A.D. Nguyen,	Viet Nam National Mekong Committee

This research has been financially supported by the Ministry of Education and Training in Vietnam and Delft University of Technology in the Netherlands.

This thesis should be referred to as: Nguyen, B. M. (2015). "Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific." Ph.D. Thesis, Delft University of Technology, Delft, the Netherlands.

*Front & Back cover:* "Hurricane Elena" by Image Science and Analysis Laboratory, NASA-Johnson Space Center. [Public domain], via Wikimedia Commons

*Printed by:* Ipskamp Drukkers B.V., the Netherlands

ISBN: 978-94-6259-538-5

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>

Copyright © 2015 by Nguyen Binh Minh

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission of the author.

*To my loved ones*



# SUMMARY

## **“Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model.**

### **A case study in the Western North Pacific”**

by

**Nguyen Binh Minh**

Delft, 3<sup>rd</sup> February 2015

In coastal areas, Tropical Cyclones (TCs) are one of the greatest threats to humanity. Unfortunately, current risk reduction measures are not completely successful in lessening TC's consequences due to the remaining uncertainties in the estimates of key parameters, on which the designs of these measures rely. Because reliable observations of TCs, having affected many regions, are restricted to quite a small number, it is not feasible to derive accurate TC statistics solely based on historical records, without producing large errors.

This research presents a comprehensive methodology to effectively overcome the observed data scarcity problem. TCs are stochastically simulated over a period of thousands of years by a numerical model, which results in a long-term database of synthetic TCs, with specifications of the central track and intensity as well as the wind field at each time step. Because TC evolution is heavily dependent on local conditions, the simulation is carried out at a regional scale to maintain relative homogeneity within both the input and outcome, and to reduce computational demand. Since the model has a generalized theoretical framework and contains the worldwide historical weather data, it can be applied to any case study. Once users define the Area Of Interest (AOI), a stepwise calibration procedure is automatically performed by a computer program to achieve the most suitable approach and to specifically determine every single detail of the model for this user-defined AOI.

The method is validated through comparisons of observed and simulated TC statistics in the AOI. For a case study of Vietnam in the Western North Pacific, this evaluation proves the model's ability to reproduce the actual TC characteristics and to generate a useable long-term database with an acceptable accuracy for practical projects. Finally, the wind speed maps and the annual exceedance probability maps are provided as possible applications of the model results.

Key words: tropical cyclone, long-term simulation, stochastic model, regional domain, generalized methodology

# SAMENVATTING

**"Regionale simulatie op lange termijn van tropische cyclonen met behulp van een gegeneraliseerd stochastisch empirisch storm model.**

**Een case study in de Noordwestelijke Stille Oceaan"**

door

**Nguyen Binh Minh**

Delft, 3<sup>rd</sup> Februari 2015

In kustgebieden vormen Tropische Cyclonen (TC's) een van de grootste bedreigingen voor de mensheid. Helaas zijn de huidige risicobeperkende maatregelen niet geheel succesvol in de vermindering van de consequenties van TC's als gevolg van de resterende onzekerheden in de schattingen van de belangrijkste parameters waarop de ontwerpen van deze maatregelen zijn berekend. Omdat betrouwbare waarnemingen van TC's, van invloed in vele regio's, tot een heel klein aantal beperkt zijn, is het niet haalbaar om nauwkeurige TC statistieken, uitsluitend gebaseerd op historische gegevens, hieraan te ontleen zonder grote fouten te maken.

Dit onderzoek presenteert een omvattende methodologie om het schaarste-probleem van waargenomen gegevens effectief te overwinnen. TC's worden door middel van een numeriek stochastisch model gedurende duizenden jaren gesimuleerd, hetgeen resulteert in een lange-termijn database van synthetische TC's met specificaties van zowel het centrale spoor en intensiteit als het windveld bij elke stap in de tijd. Omdat de evolutie van TC's sterk afhankelijk is van lokale omstandigheden, wordt de simulatie op regionale schaal uitgevoerd om de relatieve homogeniteit binnen zowel input en resultaat te behouden, en om de computationele druk te verminderen.

Aangezien het model een algemeen theoretisch kader heeft en wereldwijde historische weergegevens omvat, kan het op iedere casus worden toegepast. Zodra gebruikers de 'Area Of Interest' (AOI, ofwel, het specifieke gebied van onderzoek) hebben bepaald, wordt automatisch een stapsgewijze kalibratie procedure uitgevoerd door een computerprogramma om de meest geschikte benadering te bereiken en om ieder detail van het model vast te stellen voor de specifiek gedefinieerde AOI. De methode wordt gevalideerd door vergelijkingen van de waargenomen en de gesimuleerde TC statistieken in de AOI. Voor een case study van Vietnam in de Noordwestelijke Stille Oceaan bewijst deze evaluatie dat het model in staat is om de werkelijke TC kenmerken te reproduceren en een bruikbare lange termijn database met een aanvaardbare nauwkeurigheid van concrete projecten te genereren. Ten slotte worden kaarten met windsnelheid en kaarten met kansberekeningen van jaarlijkse overschrijdingen geleverd voor mogelijke toepassingen van het model.

Trefwoorden: tropische cycloon, lange termijn simulatie, stochastisch model, regionaal domein, algemene methodiek

# CONTENTS

<b>SUMMARY</b> .....	<b>III</b>
<b>SAMENVATTING</b> .....	<b>IV</b>
<b>CONTENTS</b> .....	<b>V</b>
<b>1 INTRODUCTION</b> .....	<b>1</b>
1.1 TROPICAL CYCLONES .....	1
1.2 RISK REDUCTION .....	3
1.3 STOCHASTIC LONG-TERM REGIONAL SIMULATION .....	5
1.3.1 <i>Extreme winds</i> .....	5
1.3.2 <i>Long-term simulation</i> .....	6
1.3.3 <i>Stochastic regional model</i> .....	7
1.4 A CASE STUDY .....	8
1.4.1 <i>The Western North Pacific</i> .....	8
1.4.2 <i>Vietnam</i> .....	11
1.5 GENERALIZED STOCHASTIC EMPIRICAL STORM MODEL.....	11
1.5.1 <i>Research objectives</i> .....	11
1.5.2 <i>Research questions</i> .....	12
1.5.3 <i>Layout of this dissertation</i> .....	13
<b>2 MODEL SETUP</b> .....	<b>14</b>
2.1 DATA COLLECTION .....	14
2.1.1 <i>Best track data</i> .....	15
2.1.2 <i>Other model input data</i> .....	19
2.2 MODEL COVERAGE.....	19
2.2.1 <i>Area Of Interest</i> .....	19
2.2.2 <i>Threat Area</i> .....	21
2.3 COMPUTATIONAL GRIDS .....	28
2.3.1 <i>Shape</i> .....	28
2.3.2 <i>Size</i> .....	30
<b>3 TRACK AND CENTRAL INTENSITY MODEL</b> .....	<b>31</b>
3.1 BACKGROUND .....	32
3.2 PROBABILISTIC MODEL .....	33
3.2.1 <i>Single site probabilistic simulation</i> .....	33
3.2.2 <i>Empirical track modelling</i> .....	34
3.3 TRACK MODELLING.....	35
3.3.1 <i>Original technique</i> .....	35

3.3.2 Clarifications and possible improvements .....	37
3.3.3 Method background and clarifications .....	37
3.3.4 Possible improvements .....	41
3.4 INTENSITY SIMULATION OVER WATER .....	55
3.4.1 Representative of intensity.....	55
3.4.2 Relative intensity.....	56
3.4.3 Modelling intensity over the sea .....	61
3.5 DECAY MODEL .....	72
3.5.1 Existing methodology.....	72
3.5.2 The GSESM's approach .....	73
<b>4 SPATIAL WIND FIELD MODEL .....</b>	<b>78</b>
4.1 GENERAL PARAMETRIC WIND FIELD MODEL .....	79
4.1.1 Existing techniques.....	79
4.1.2 Parametric wind field methodology .....	80
4.1.3 The GSESM's approach for modelling the parametric wind field .....	86
4.2 RADIUS OF MAXIMUM WIND MODELLING.....	91
4.2.1 Existing methodologies.....	91
4.2.2 The GSESM's method for modelling the radius of maximum wind.....	92
4.3 HOLLAND PARAMETER MODEL .....	94
4.3.1 Existing methods .....	94
4.3.2 The GSESM's method for the modelling of the Holland parameter.....	95
<b>5 SIMULATION AND RESULTS.....</b>	<b>98</b>
5.1 SIMULATION PROCEDURE.....	99
5.1.1 Number of years.....	99
5.1.2 Annual occurrence rate.....	100
5.1.3 Model initialization .....	104
5.1.4 Track and central intensity development.....	105
5.1.5 Spatial wind field.....	106
5.1.6 Boundary conditions.....	107
5.2 MODEL VALIDATION .....	108
5.2.1 Testing over the entire AOI.....	109
5.2.2 Validation for each grid within the AOI .....	113
5.2.3 Evaluation of the number of years to be simulated in the model.....	116
5.3 POSSIBLE APPLICATIONS .....	117
5.3.1 Wind speed map.....	117
5.3.2 Annual exceedance probability map.....	121



<b>6 CONCLUSIONS AND RECOMMENDATIONS.....</b>	<b>123</b>
6.1 CONCLUSIONS .....	123
6.1.1 <i>Questions associated with the model setup.....</i>	<i>124</i>
6.1.2 <i>Questions related to the modelling of central track, intensity, and surface wind field.....</i>	<i>125</i>
6.1.3 <i>Questions connected with the model run .....</i>	<i>127</i>
6.2 RECOMMENDATIONS FOR FUTURE RESEARCH .....	128
<b>REFERENCES.....</b>	<b>130</b>
<b>LIST OF TABLES.....</b>	<b>141</b>
<b>LIST OF FIGURES.....</b>	<b>143</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>145</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>146</b>
<b>CURRICULUM VITAE.....</b>	<b>147</b>
<b>LIST OF PUBLICATIONS .....</b>	<b>148</b>



# 1 INTRODUCTION

This research presents a generalized methodology to model a long-term database of synthetic Tropical Cyclones (TCs) at regional scale. In the first chapter, the basic aspects are described, including the research subject (section 1.1), rationale (section 1.2), type of the model and its temporal and spatial scales (section 1.3), and selection of a case study (section 1.4). Finally, the research objectives, questions, and the layout of this dissertation are described in section 1.5.

## 1.1 Tropical cyclones

Natural hazards are one of the greatest threats to humanity. Every year, up to 340 million vulnerable people (Bankoff et al. 2004) are exposed to events that lead to disasters with tremendous human, environmental, and financial losses (Wisner et al. 2003). Between 1995 and 2004, natural disasters affected an estimated total number of 2.5 billion people, causing about 890,000 deaths, and \$570 billion losses (UNISDR 2008).

Among the various types of natural hazards, extreme weather events (e.g., extreme heat, droughts, river floods, and TCs) are the most destructive ones. The statistics data shows that these events are responsible for 71% of large-scale disasters, 45% of fatalities, 69% of economic losses, and 91% of insured losses (CRED 2008). Table 1.1 and Table 1.2 (Tompkins 2002) list the top five deadliest and costliest events that occurred between 1970 and 2005.. As can be clearly seen from the tables, in comparison with other extreme weather events, TCs are the most devastating disasters both in terms of recorded mortalities and financial losses.

TCs are cyclonic storm systems that originate over the oceans. Depending on their intensity and geographic positions, TCs are referred to by different names such as hurricanes, typhoons, tropical storms, and tropical depressions. In coastal areas, where more than 60% of the world's population lives within 150 km of the coastline (Green 2009), TCs are the most dangerous disasters that cause massive losses along their trajectories over the tropical belt (Ariffin and Moten 2009). TCs kill many people, affect numerous sectors (e.g., agriculture, aquaculture, and

industry), and cause billions of US dollars in property and infrastructure damages. From 1980 to 2000, a total of 251,384 people lost their lives due to TCs and every year about 119 million people are under their influences (UNDP 2004); more than one-third of the annual number of people, which are exposed to all types of natural disasters [i.e., 340 million people (Bankoff et al. 2004)].

In addition, TCs have an important impact on the long-term coastal morphology as storm surges and large waves generated by them can erode the beach and dunes, and reshape the landscape (Brettschneider 2006). Evidence from various researches [e.g., (Aixue Hu 2009; Emanuel 2001; Pasquero and Emanuel 2008; Sriviver and Huber 2007)] also suggested that TCs may be a critical element in the large-scale climate systems. Because TCs transfer enormous amounts of heat into the seas along their tracks, TC's activity can alter the oceanic meridional heat transport and overturning circulation.

**Table 1.1 Top five deadliest extreme weather events, 1970-2001 [from (Tompkins 2002)]**

Rank	Victims <sup>1</sup>	Insured Loss <sup>2</sup>	Date	Event <sup>3</sup>	Country
<b>1</b>	<b>300,000</b>	<b>N/a</b>	<b>14.11.70</b>	<b>TC Bhola</b>	<b>Bangladesh</b>
<b>2</b>	<b>138,000</b>	<b>3</b>	<b>29.04.91</b>	<b>TC Gorky</b>	<b>Bangladesh</b>
<b>3</b>	<b>15,000</b>	<b>106</b>	<b>29.10.99</b>	<b>TC 05B</b>	<b>India (Orissa), Bangladesh</b>
4	15,000	N/a	01.09.78	Flooding	Northern India
5	10,800	N/a	31.10.71	Flooding	India (Orissa), Bay of Bengal

**Table 1.2 Top five costliest extreme weather events, 1970-2001 [from (Tompkins 2002)]**

Rank	Victims <sup>1</sup>	Insured Loss <sup>2</sup>	Date	Event <sup>3</sup>	Country
<b>1</b>	<b>38</b>	<b>20,185</b>	<b>23.08.92</b>	<b>Hurricane Andrew</b>	<b>US, Bahamas</b>
<b>2</b>	<b>51</b>	<b>7,338</b>	<b>27.09.91</b>	<b>Typhoon Mireille</b>	<b>Japan</b>
3	95	6,221	25.01.90	Winterstorm Daria	France, UK et al.
4	80	6,164	25.12.99	Winterstorm Lothar	France, CH et al.
<b>5</b>	<b>61</b>	<b>5,990</b>	<b>15.09.89</b>	<b>Hurricane Hugo</b>	<b>Puerto Rico, US et al.</b>

In developing countries, where 80% of the world's largest cities are situated (Bendimerad 2004) and where 90% of global population growth is taking place (World Vision 2009), the tables show that the deadliest TCs usually happen. It also shows that there is an increase in exposure to TCs due to the population explosion, as more and more people are living in the insufficiently protected coastal areas (UNISDR 2008). For instance, landfalling TCs in Bangladesh were accounted for

<sup>1</sup> Dead and missing

<sup>2</sup> In USD millions, indexed to 2001

<sup>3</sup> TCs are in boldface

more than 60% of the global TCs' death toll over the period 1980-2000 (UNDP 2004). TC Bhola (1970), which was the deadliest TC in the history, killed more than 300,000 people (Southern 1979) by triggering violent storm surges in the densely populated Ganges Delta. The high fatality rate indicated that in this case, high vulnerability coincided with high physical exposure.

On the other hand, while loss of life has been significantly decreased in the developed world by robust damage reduction measures and effective preparedness systems, rich countries often pay the highest financial losses due to the great economic value of existing properties and a high insurance level in storm-prone regions (Emanuel et al. 2006). The U.S. is a typical example, in which, TCs are the most expensive natural disasters in the whole U.S. history (CBO 1995). The costliest TC ever recorded was Hurricane Katrina, which caused \$81.2 billion in property damages (Pielke et al. 2008).

In summary, TCs have extensive negative impacts on numerous aspects of human society and the ecological environment. As a consequence, there is an urgency to lessen the destructive effects of TCs by using various techniques. These damage reduction approaches will be discussed in the following section.

## 1.2 Risk reduction

An easy to understand definition of TC's risk is the estimated degree of threat facing a vulnerable group of people through exposure to this event (World Vision 2009). Therefore, the more susceptible the society is (in various means such as physically, economically, environmentally, or socially), the more expensive and deadly the TCs will be. While TCs are unavoidable, their risk can be considerably lessened either by decreasing the community's vulnerability and/or by increasing the capacities to withstand TCs of the affected population, which is their resilience.

In the developing world, because of limited available funds, equipment, and technology, risk reduction programs are mainly focused on non-structural ways. These include appropriate building technologies and regulations, proper land use planning based on vulnerability and capacity assessments, installation of early warning systems and signals, community awareness and preparedness plans, and evacuation practices (UNISDR 2008). Nevertheless, while these approaches show their advantages to some extent, suitable constructional techniques should be considered as long-term TC prevention methods. Because great numbers of poor people's settlements are situated in hazardous areas, effective coastal protection measures can substantially reduce a high death toll and property damage.

Regarding the rich countries, most of them rely on a comprehensive system to resist TCs, including both structural and non-structural methods. Structural solutions, also referred to as hard engineering methods, are basically appropriately designed coastal defence structures (e.g., sea walls, storm barriers, revetments, and offshore breakwaters) that can cope with the impacts of severe TC winds and storm surges at an acceptable level. Non-structural techniques are based on the early warning programs and predictive models of morphological effects induced by extreme TCs. They include soft engineering approaches (e.g., beach nourishment, sand dune stabilization, and beach drainage), advanced building codes and their enforcements, master plans of coastal zones based on precise risk assessments,

increasing the effectiveness of preparative solutions (e.g., evacuation strategies), and giving more warning time before TCs hit land.

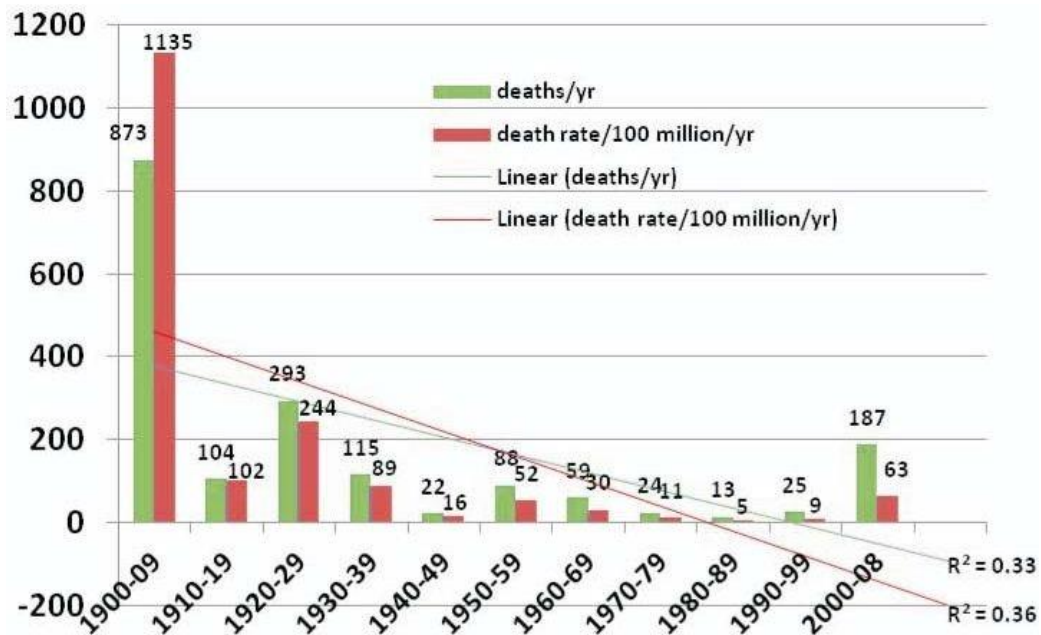
However, despite the remarkable developments in risk reduction techniques, there still remain some uncertainties concerning the reliability of these methods, which compromise the efficiency of risk reduction measures in many cases. Table 1.3 (Goklany 2009) gives the statistic data of annual global deaths and death rates caused by different types of extreme weather events for two contiguous periods, 1900-1989 and 1990-2008. As can be seen from the table, among the top three deadliest categories, which accounted for more than 99% of total mortalities, only TCs have an increase in annual fatalities. While the total annual deaths for the six types decreased by 84% between the 1900–1989 and 1990–2008 periods, the number of people that lost their lives due to TCs increased more than twofold at the same time. Even in the U.S., where a lot of effort and resources were put into TC risk reduction programs (Lee and Rosowsky 2007), this upward trend over the most recent decades can also be observed. Figure 1.1 (Goklany 2009) provides the trend for deaths and death rates due to TCs in the U.S. The figure begins with very high mortalities caused by hurricane Galveston (1900) and continues with a very steep downward trend in fatalities during the following decades. However, the large spike for deaths due to hurricane Katrina (2005) suggested that a TC is still a "hard to tackle" subject, even with the modern techniques.

To conclude, TCs are responsible for most of the mortalities and economic losses due to natural disasters in many regions. Unfortunately, unlike other types of extreme weather events, current risk reduction measures for TCs are not completely successful in lessening their consequences. The rationale for this study is to develop an advanced methodology of TC simulation, which provides an important input to establish proper building codes and to assess TC risks in the storm-prone regions.

**Table 1.3 Annual global deaths and death rates for various categories of disasters, 1900-1989 and 1990-2008 [from (Goklany 2009)]**

Events <sup>4</sup>	Deaths per year		Death rates per year (per million people)	
	1900-1989	1990-2008	1900-1989	1990-2008
Droughts	130,044	225	58.19	0.04
River floods	75,169	7,676	31.87	1.28
<b>TCs</b>	<b>11,018</b>	<b>20,079</b>	<b>4.00</b>	<b>3.35</b>
Mass movement-wet	441	780	0.15	0.13
Extreme temperatures	124	5,144	0.03	0.82
Wildfires	22	69	0.01	0.01
Total	216,819	33,973	94.24	5.63

<sup>4</sup> TCs are in boldface



**Figure 1.1 Deaths and death rates due to hurricanes in the U.S., 1900–2006 [from (Goklany 2009)]**

## 1.3 Stochastic long-term regional simulation

### 1.3.1 Extreme winds

When TCs hit the coast, they bring along many devastating direct impacts, which can be distinguished into four types.

Firstly their intense winds, which can reach up to 408 km/h (Courtney et al. 2012), can easily demolish homes and buildings either by blowing them away or by throwing debris (e.g., broken trees, signs, or other stuffs) on top of these properties.

Secondly, TCs produce storm surges, which are rising walls of ocean water that sweep through the entire affected coastal regions. These wind-induced surges can be as high as 14.6 m at their peak (Whittingham 1958) and up to 80 to 160 km wide (FEMA 2013). The wind-induced surges are one of the deadliest effects of TCs, accounting for nearly 90% of all TC-related injuries and fatalities throughout the history (Pava et al. 2010).

Thirdly, TCs can generate tornadoes. However, current risk assessments do not often include their influence, because of the infrequent occurrence and the relatively small contribution to the total losses due to TCs. Over the period 1948-1972, tornadoes were observed in about 25% of all TCs making landfall in the U.S. Providing that they happened due to a TC, tornadoes were responsible for only a small number, up to 10%, of the mortalities caused by the TC that generated them (Novlan and Gray 1974). In other regions, the tornado occurrence rates are even smaller.

Finally, after TCs come ashore, they generate heavy precipitation that cause tremendous freshwater flooding in rivers and urban areas. These floods can

afterwards trigger landslides along riverbanks and in mountainous and hilly regions. In several cases (e.g., hurricane Floyd or hurricane Mitch) loss of lives and destruction of properties mainly originated from flooding. Although the impacts of TC-induced rainfall and its implications are visible, most of the existing studies associate TC risk only with winds. The underlying reason is, that currently, the knowledge of TC-induced rains has not yet reached to a level, which can be a basis for reliable rainfall predictions. Furthermore, observations of TC-induced precipitation are not sufficient to make proper evaluations of flooding risks (Emanuel et al. 2006). In contrast, measurements of TC winds are much more complete and, historically, much of the death toll and financial losses are caused by TC winds and wind-induced surges.

Therefore, extreme TC winds are considered as a critical input for many TC-related research, such as setting up building regulations or estimating TC risks in storm-prone regions.

### 1.3.2 Long-term simulation

As described in the previous section, extreme winds and surges are two major culprits of TCs.

Unfortunately, reliable observations on the tracks and winds of TCs having affected many regions are restricted to quite a small number. In many cases, TCs have been monitored and recorded for a relatively short period, such as about 40 years in Australia (James and Mason 2005) or 60 years in the Western North Pacific [WNP (Rumpf et al. 2007)]. However, even with longer historical data, such as the 100-year record in the U.S. (Powell et al. 2005), the measurements are still not adequate to directly utilize the observed data to estimate design levels of coastal defence measures with acceptable exceedance probabilities during a normal functioning period (e.g., 20, 50, or 100 years) (James and Mason 2005). This is because TCs are both relatively infrequent and small in terms of the length of coastlines affected by these TCs each year. Therefore, it is not feasible to derive accurate key parameters for the most intense TCs, solely based on historical records, on which risk analyses, building codes, and designs of coastal defence structures will rely (Hallegatte 2007), without producing large errors. In some particular places, such as New England in the U.S., although the locations have hardly been stricken by TCs, the possible consequences of a landfalling TC are massive because of the densely populated urban area, the high economic value of existing properties, and high insurance levels (Emanuel et al. 2006). However, due to a limited compilation of reported TCs, reliable risk assessments for these locations are currently not available (Lin et al. 2010).

The observed samples of storm surges, available to determine extreme surges associated with TCs, are even more limited because of the small number of locations at which such data are measured. This is especially the case for places like Boston and New York in the U.S., which have experienced rare but extremely devastating TCs, or for offshore sites, where properties and infrastructures for the tourism or energy industry (e.g., pontoons, offshore windmills or oil rigs) are located or proposed.



An effective and widely accepted technique to overcome the data scarcity problem is to enlarge TC samples by using a numerical TC simulation. The outcome of this model is a long-term database of synthetic TCs, over a period of hundreds or even thousands years, with specifications of the surface wind and atmospheric pressure fields at time step along their tracks. On the condition that statistical characteristics derived from these simulated TCs are proven to imitate those of the population of real TCs, results from a long-term TC model can be utilized as a complete input for any TC-related study.

In this case, the benefits are twofold. Firstly, synthetic wind speed records can be used directly to estimate the risk of extreme winds. Secondly, it provides the detailed and reliable wind and pressure fields that are needed to drive a numerical storm surge model (Weisberg and Zheng 2006). After calibration against reference observations, such model determines the extreme surge at every time step of the simulation for all centres of the computational grids. The calculations are repeated for every TC in the long-term data set and produce a compilation of extreme surges for the entire research area (James and Mason 2005).

In conclusion, a long-term simulation is necessary for most of TC hazard analyses and structural risk assessments.

### 1.3.3 Stochastic regional model

Numerical weather prediction uses mathematical models of the atmosphere and the oceans combining the basic formulae of various physical principles and processes, based on current weather conditions. These models provide the groundwork to forecast the development of tracks and intensities of specific TCs several days or hours beforehand. However, predictive capacity is not a topic in this study, because the model intends for simulating the statistical attributes of the TC population. Moreover, at seasonal and longer timescales, as is the case with risk research, weather is unpredictable and TCs must be treated stochastically (Hall and Jewson 2007).

In most studies, the TC key parameters at landfall (or within a short seaward distance from the coastline in case of offshore sites such as oil rigs) are of the most important concern within risk assessments. Therefore, a logical way to simulate landfall characteristics is to base the research only on observed data at landfall locations [e.g., (Jagger et al. 2001)]. Nevertheless, it is impossible to evaluate risks at numerous coastal locations when there are very few, or even no, historical records. An effective approach to compensate for this data shortage is to utilize the whole TC track, from the initial points to the lysis positions. By using this method, the number of data used to formulate a stochastic model increase significantly.

As the simulation outcomes are necessary to drive the numerical surge models, the research area must be large enough to include the accumulation of surge reactions caused by remote impacts. This is because, although surges occur locally, circulation physics at the ocean's bottom allow for non-local influences (Weisberg and Zheng 2006). Furthermore, the model domain must also be quite extensive to capture all the possible effects of historical TCs in the primary Area Of Interest (AOI) and to allow for variety in TC development scenarios.

However, the research area should not be too large. There are three reasons for this suggestion. Firstly, because most of the time risk-related studies are only interested in a relatively small region (e.g., the coastal zone of a state or a country), thus TC's activity located too far away are not relevant for assessing risk (Hall and Jewson 2007). Secondly, if a large research area is used, the model will consist of a great number of TCs with different behaviours, as TC's evolution is heavily dependant on local conditions. Therefore, in a large-scale study (e.g., a basin-wide one), various filter criteria are applied sooner or later to maintain the homogeneity of usable TCs. An example is presented by Rumpf et al. (2007), in which a strong inhomogeneity in the patterns of TC geographic trajectories in the WNP were explored. The researcher, later on, divided the historical tracks into 6 classes, which were claimed to be more homogeneous than the parent compilation and thus enhanced model performance. Finally, a huge model domain can significantly increase computational demand. This unnecessary requirement can be a crucial limitation, especially when other computationally demanding studies, such as an in-deep loss estimation, must be carried out based on model results, or simulations must be repeated many times with various alternatives (Legg et al. 2010).

To sum up, a stochastic regional model is suitable method to generate a long-term database of synthetic TCs. The research area must be large enough to contain any possible effects of winds and surges, but should not be too large to keep the homogeneity of TC tracks and to reduce computational demand. A detailed definition of the model domain, which is supported by sound arguments, will be given in section 2.2.

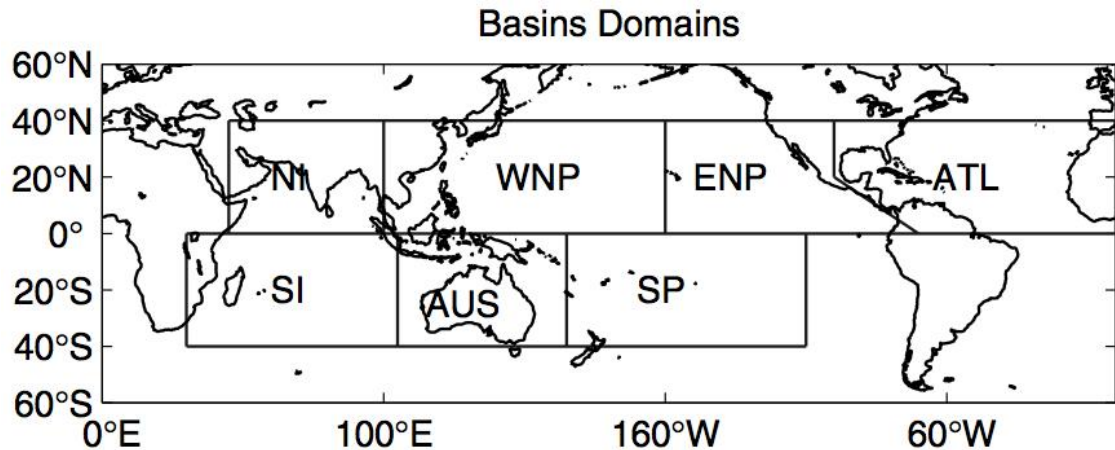
## 1.4 A case study

Since the model developed in this research has a generalized theoretical framework and contains the worldwide historical weather data, it can be applied to any case study. Once users define the AOI, a stepwise calibration procedure is automatically performed by a computer program to achieve the most suitable approach and to determine every single detail of the model, specifically for this user-defined AOI. However, a case study must be given in order to verify the theoretical framework and to evaluate the model performance.

### 1.4.1 The Western North Pacific

Figure 1.2 (Camargo et al. 2005) gives an overview of worldwide TC basins. Among them, the WNP, which covers the latitude and longitude ranges from 0 to 60 degrees North and from 100 to 180 degrees East respectively, experienced the most intense TCs throughout history. In the available records, a total of about 70 TCs reached a central pressure of less than 900 hPa, most of which took place in the WNP. Proof is given in Table 1.4 [adapted from (Wikipedia 2014a)]. This table presents the top five strongest TCs measured by minimum central pressure (the reasons to choose this parameter as an indicator of TC intensity will be provided later, in subsection 3.4.1), based on observations in the International Best Track Archive For Climate Stewardship [IBTrACS (Knapp et al. 2010b)]. As can be seen from Table 1.5 [adapted from (Wikipedia 2014a)], in comparison with other basins, the most extreme TC ever recorded in the WNP was Typhoon Tip, which

attained a central pressure of 870 hPa on 12th October, 1979 (Dunnavan and Diercks 1980), and which was much stronger than the most intense TC in other basins (with geographic extents as given in Figure 1.2). The statistical data in Table 1.6 (Neumann 1993) also shows that WNP is the most active TC basin in the world.



**Figure 1.2 Global TC basins<sup>5</sup> [from (Camargo et al. 2005)]**

**Table 1.4 Top five most extreme TCs [adapted from (Wikipedia 2014a)]**

Rank	Minimum central pressure (hPa)	Season	Name of the TC	Basin <sup>6</sup>
<b>1</b>	<b>870</b>	<b>1979</b>	<b>Tip</b>	<b>WNP</b>
2	875	1973	Nora	WNP
3	875	1975	June	WNP
4	877	1958	Ida	WNP
5	880	1966	Kit	WNP

**Table 1.5 Strongest TCs by basins [adapted from (Wikipedia 2014a)]**

Rank	Minimum central pressure (hPa)	Season	Name of the TC	Basin <sup>6</sup>
<b>1</b>	<b>870</b>	<b>1979</b>	<b>Tip</b>	<b>WNP</b>
2	882	2005	Wilma	ATL
3	890	2002	Zoe	SP
4	895	2003	Gafilo	SI
5	900	1998	Gwenda	AUS
6	902	1997	Linda	ENP
7	912	1999	Paradip	NI

<sup>5</sup> Worldwide TC basins: Atlantic (ATL), Australia (AUS), Eastern North Pacific (ENP), North Indian (NI), South Indian (SI), South Pacific (SP), and Western North Pacific (WNP).

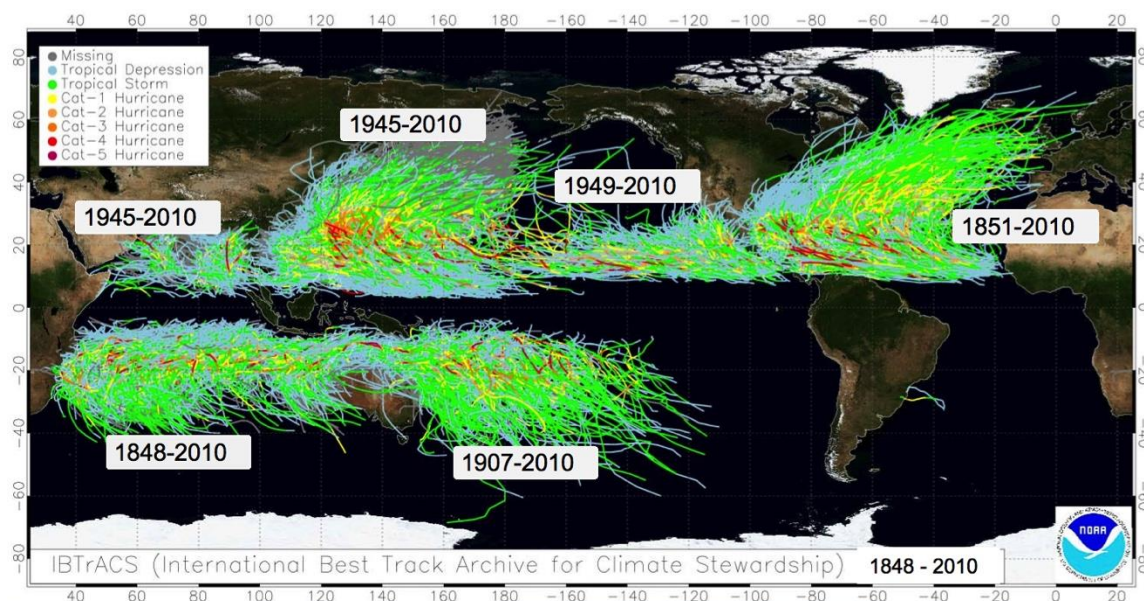
<sup>6</sup> The TCs occurred in the WNP are in boldface

**Table 1.6 Annual TC occurrence by basins [from (Neumann 1993)]**

Rank (by average annual TC occurrence)	Basin <sup>7</sup>	Tropical Storm or stronger (greater than 17 m/s sustained winds)			Hurricane/Typhoon/Severe Tropical Cyclone (greater than 33 m/s sustained winds)		
		Most	Least	Average	Most	Least	Average
<b>1</b>	<b>WNP</b>	<b>39</b>	<b>14</b>	<b>26.0</b>	<b>26</b>	<b>5</b>	<b>16.5</b>
2	ENP	28	8	16.6	16	3	8.9
3	ATL	28	4	12.1	15	2	6.4
4	SP	20	4	9.9	12	1	5.2
5	SI	14	4	9.3	8	1	5.0
6	AUS	16	3	7.5	8	1	3.6
7	NI	10	2	4.8	5	0	1.5

Moreover, TC risk analyses in the WNP suffer greatly from a data scarcity problem due to the relatively short observation period. As shown in Figure 1.3 (Knapp et al. 2010b), measurements have been implemented since 1945, after World War II. The time length of the historical data is therefore much shorter than the ones in other basins, such as half and one-third of the TC records in Australia and the Atlantic region, respectively.

To summarize, the WNP as the research basin in this study is not only chosen because the basin has the highest TC occurrence rate and experiences the strongest TCs, but also due to the serious shortness of historical data, which is the exact problem that the model developed in this study intends to overcome.



**Figure 1.3 Globally recorded TC's activity [from (Knapp et al. 2010a)]**

<sup>7</sup> The WNP is in boldface

## 1.4.2 Vietnam

Vietnam is located in one of five storm-prone areas of the WNP, and has frequently been affected by TCs. On average, the country has experienced 6 to 8 typhoons each year (UNDP 2007). Furthermore, with a long coastline of approximately 3440 km (Luong et al. 2011) and the densely populated coastal areas, Vietnam is among the top five countries most affected by weather related loss events, particularly by TCs (Dasgupta et al. 2009). Rapid population growth, unplanned urbanization, and development within high-risk zones are the main causes that have increased the vulnerability of the coastal population (Holmes et al. 2005). As a result, the consequential socio-economic damages have increased continuously. For example, while the total sum of losses and losses per unit GDP caused by extreme weather events (mostly by TCs) in 2008 were US \$ 2423 million Purchasing Power Parity (PPP) and 1.01% respectively (Harmeling 2009), those values were US \$ 2943.05 million PPP and 1.15% respectively in 2009 (Harmeling 2010).

In Vietnam, TC risk studies are especially faced with numerous difficulties. Generally, there is no complete TC database as well as a systematic method or the tools to store, maintain, and analyse such records (Luong et al. 2011). Observations have been measured, kept, and assessed locally at the provincial level. This collection and management approach leads to an inconsistent and, usually, non-electronic local compilation of data. In addition, the potential problems with a sole reliance on observed data are more serious, because there are hardly any local nearshore measurements, which are valuable sources for model verification. This is due to budget constraints, the lack of suitable techniques, the use of obsolete equipment, and also, the aftermath of war in the country from 1945 to 1975.

## 1.5 Generalized Stochastic Empirical Storm Model

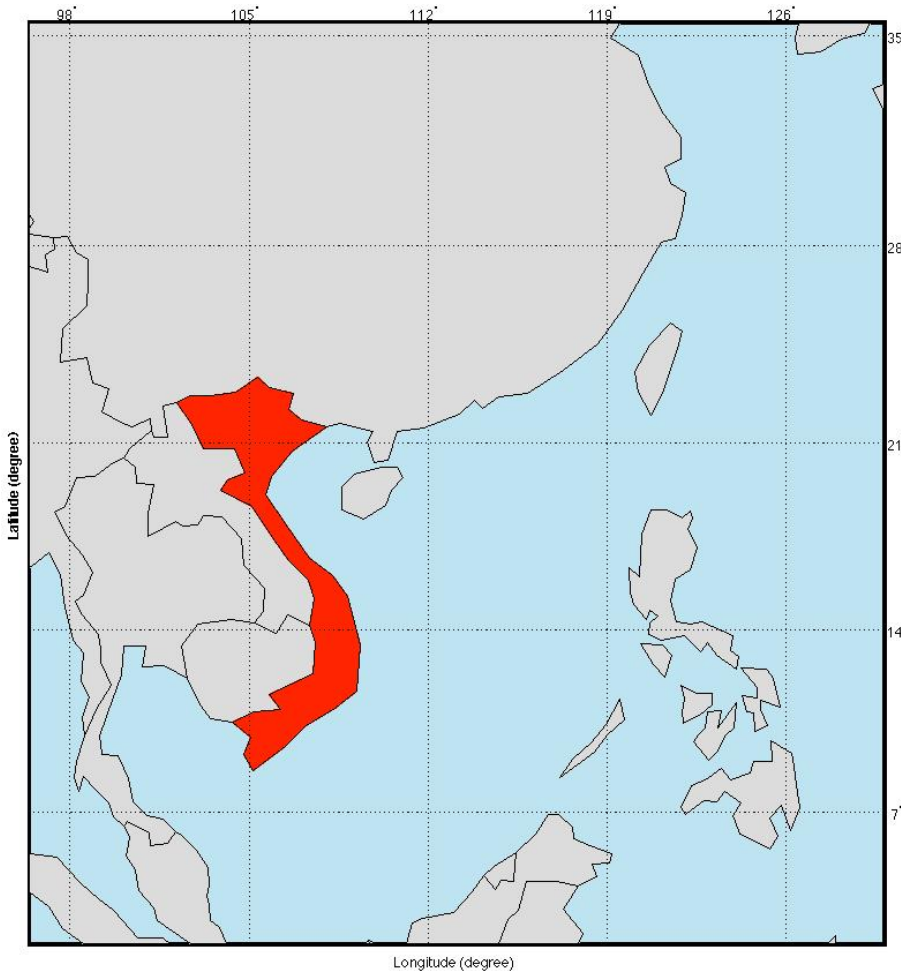
### 1.5.1 Research objectives

Considering all the aforementioned reasoning, this research has two objectives:

- To develop a model of TC activity, which is called Generalized Stochastic Empirical Storm Model (GSESM), with the following properties:
  - To simulate the full TC tracks, from genesis to lysis,
  - To contain specifications of the surface wind and atmospheric pressure fields at each TC centre and time steps along their tracks,
  - A long-term simulation period
  - A stochastic model
  - To produce reasonable research area
  - Usable for any case study at any location

The synthetic TC database provided by this model can be used to assess risks, to design coastal defence structures, to establish building codes, or to be used in other TC-related studies.

- To apply the model to a case study of Vietnam (Figure 1.4) in the WNP basin, in order to:
  - Verify the theoretical model and assess its accuracy
  - Present some products that usable for other researches



**Figure 1.4 Geographic extent of Vietnam and the nearby water areas<sup>8</sup>**

### 1.5.2 Research questions

The following questions are proposed, including:

Questions associated with the model setup (chapter 2):

- What type of data must be collected to construct the Generalized Stochastic Empirical Storm Model (GSESM)? Which sources are available for each required type of data? How to evaluate the quality of those sources? If there are several accessible sources, which one will be chosen and why?
- What is the Area Of Interest (AOI)? How to define its geographic range?

---

<sup>8</sup> The Vietnamese geographical area is drawn in red color

- What is the Threat Area (TA)? Which criteria can be used to determine if a data point should be included in the research or not? How to utilize this indicator to define the boundary of the TA?
- What is the shape and the size of the computational grids?

Questions related to the modelling of central track and intensity (chapter 3), and the surface wind field (chapter 4):

- Which are the current theoretical frameworks for modelling key parameters? What are their pros and cons? Which one should be chosen as a basis for the Generalized Stochastic Empirical Storm Model (GSESM)?
- What are the limitations of the (chosen) existing method? How can the GSESM overcome these limitations? In comparison with the original technique, what are the GSESM's improvements?

Questions connected with the model run (chapter 5):

- How long should a reasonable length of synthetic Tropical Cyclone (TC) database be?
- Which basic discrete distribution should be employed to approximate the TC annual occurrence rate?
- How to define a set of initial points for TCs in the simulation?
- What is the Damage Distance Threshold (DDT)? How to define the DDT?
- Which physical boundary conditions should be included to introduce the realistic limits of the parameters as well as to defined the lysis of a TC?
- How to validate the model?
- What are the possible applications of the model outcomes?

### 1.5.3 Layout of this dissertation

The layout of this dissertation follows the research questions. Each chapter will deal with a set of questions as shown above.

After the overview of the field of study in chapter 1, the dissertation continues with a discussion of every aspect of the model setup in chapter 2.

In chapter 3 a summary of available methods for modelling central track and intensity is provided and a new approach with improvements over the existing ones will be given.

In chapter 4 a similar approach is followed as in chapter 3, but instead of central track and intensity, here the modelling of surface wind field is presented.

The simulation is carried out in chapter 5. Key parameters of both historical and simulated TCs are collected and compared to validate the theoretical model. Furthermore, the wind speed maps and the annual exceedance probability maps are provided as the possible applications of the model outcomes.

Finally, the conclusions and recommendations for future research are given in chapter 6.

## 2 MODEL SETUP

### Research questions:

- What type of data must be collected to construct the Generalized Stochastic Empirical Storm Model (GSESM)? Which sources are available for each required type of data? How to evaluate the quality of those sources? If there are several accessible sources, which one will be chosen and why?
- What is the Area Of Interest (AOI)? How to define its geographic range?
- What is the Threat Area (TA)? Which criteria can be used to determine if a data point should be included in the research or not? How to utilize this indicator to define the boundary of the TA?
- What is the shape and the size of the computational grids?

The GSESM will be formulated because of the rising demand for an advanced technique that can compensate for the lack of Tropical Cyclone (TC) observations in many regions. Research must be done to ensure a proper model configuration. Tasks to be considered include: selection of required data types and sources for each type (section 2.1), a definition of the AOI and the TA and their geographic range (section 2.2), and analysis and choice of shape and dimension of the computational grids (section 2.3).

### 2.1 Data collection

When one has to determine the possible situations that could be happen at a given location, there are two ways to fulfil that job (Brettschneider 2008). The first approach is the "persistence" technique, in which previous conditions are extrapolated to find out the next ones. The rationale behind the method is that, logically, there must be a close relationship between the situations at two contiguous time steps, providing that the interval between them is small enough. An example of that methodology is the "CLIPER" model [CLImatology &



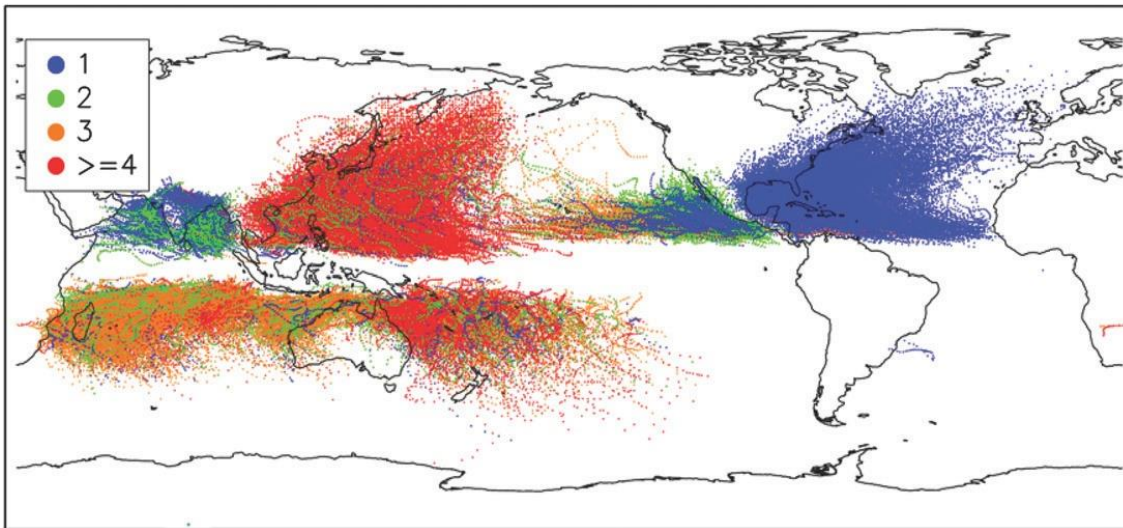
PERsistence (NHC 1997)] and its updated version "CLIPER5" (NHC 2006), which were constructed to forecast future TC track and intensity. In this study, the "persistence" method cannot be directly applied because the model is built without predictive ability (subsection 1.3.3) and synthetic TC tracks are intentionally created based on historical data, but yet somewhat independently from them. However, the logical "persistence" technique still gives a valuable idea for the equations of TC's evolution, which will be used in chapter 3. The second approach is the "climatology" method, in which an average value can be an acceptable quantity of a parameter at a specific position. For instance, if the TC annual occurrence rate derived from a long-term database is 3.4 TCs per year, one can expect that more than 3 TCs occur at that location in any year. Using this technique, both the fundamentals and the accuracy of the model are not significantly changed even if additional TCs are introduced to the historical record. The "climatology" method is therefore suitable for any long-term TC studies and it will be used in this study as well.

The GSESM involves the modelling of TC track and intensity together with surface wind field. As a result, the historical sets of those parameters, which contain values measured directly or values derived from other sources, must be collected.

### 2.1.1 Best track data

The basic and most important input for a TC climatology study is the so-called Best Track Data (BTD). A BTD generally contains position (i.e., latitude and longitude coordinates of the TC centre) and intensity (i.e., surface wind speed or pressure), which is measured every 6 hours for each historical TC. In the Western North Pacific (WNP), unfortunately, there is no ultimate BTD like the HURDAT database (Jarvinen 1984) in the Atlantic. Because TCs are monitored by various agencies, there are at least 4 different BTD sources, as shown in Figure 2.1 (Kruk et al. 2009). The following organizations provide data in the WNP:

- Japan Meteorological Agency (JMA)
- U.S. Defense Joint Typhoon Warning Center (JTWC)
- Chinese Meteorological Administration's Shanghai Typhoon Institute (STI)
- Hong Kong Observatory (HKO)

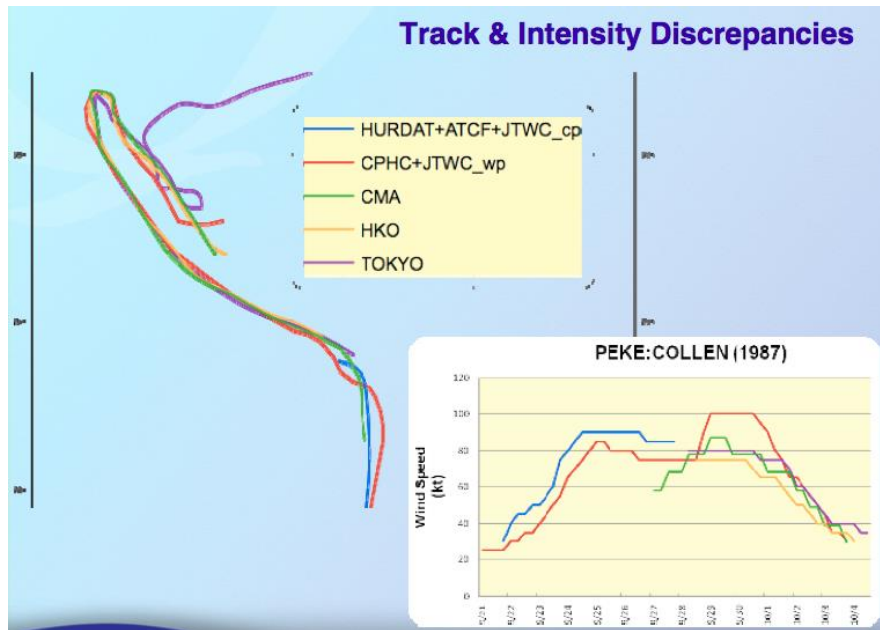


**Figure 2.1 Number of available BTD sources [from (Kruk et al. 2009)]**

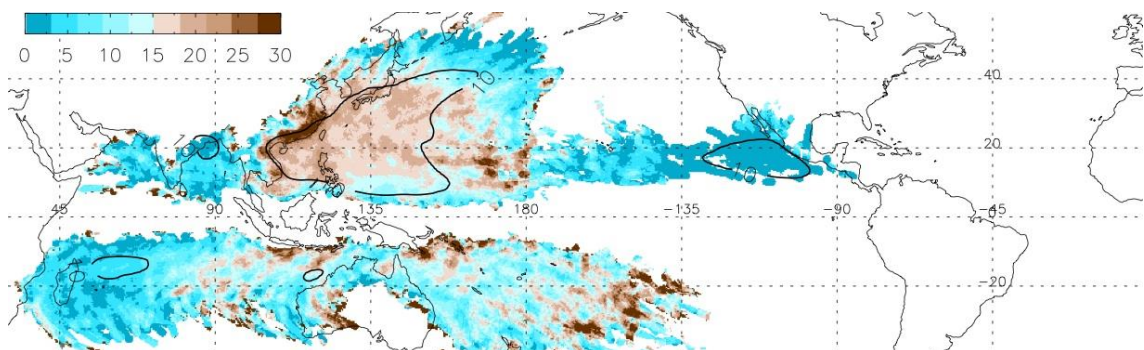
Many researchers [e.g., (Kamahori et al. 2006; Ren et al. 2011; Song et al. 2010; Wu et al. 2006)], examined the quality of all, or some, of the above BTD sources and provided two common conclusions.

Firstly, the data recorded before the first use of weather satellites are unreliable and therefore unacceptable for application in statistical analyses. That is because ship reports and damage information of landfalling TCs were the only way to assemble data at that time. The same situation can also be seen in other basins [e.g., the Atlantic (Brettschneider 2006)], which leads to an exclusive use of BTD derived from satellite imagery in most cases.

Secondly, there are strong discrepancies in estimations of TC parameters between different agencies in the WNP (Barcikowska 2012; Knapp and Kruk 2009). An example is presented in Figure 2.2 (Knapp et al. 2010b), which shows the interagency differences in both track and intensity of typhoon Peke (1987). Generally, up to 30% difference can be observed in nearly every TC, as shown in Figure 2.3 (Knapp et al. 2010b). The main reason for those discrepancies are the considerable limitations (Velden et al. 2006) of the Dvorak method (Dvorak 1975), which has been the main approach for assembling BTD sets, especially since reconnaissance flights over the WNP mostly ended in 1987. Although the same basic principles of the Dvorak methodology are applied by all organizations, the subjective classifications of cloud patterns in satellite observations (Kossin and Velden 2004) and the diverse guidelines for applications of this methodology (Barcikowska et al. 2012) result in interagency differences. This has led to efforts to combine various BTD sets (Kruk et al. 2009) into a global one [e.g., the Munich Re record that is used in Rumpf et al. (2007) or the IBTrACS archive (Knapp et al. 2010b)]. However, in many cases, the disparities are irreconcilable (Knaff and Sampson 2006; Lander 2008) and thus a replacement dataset is required to estimate the explicit values of TC key parameters.



**Figure 2.2** Interagency differences in estimations of track and intensity, TC Peke (1987) [from (Knapp et al. 2010a)]



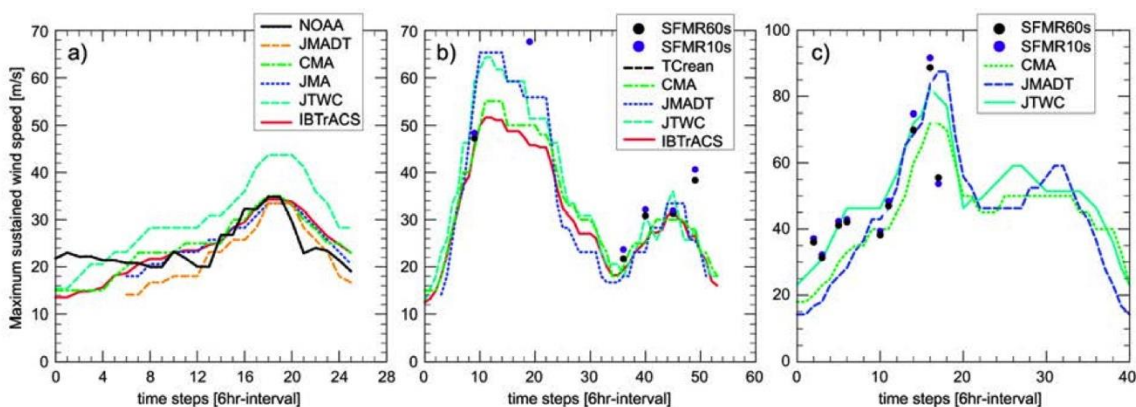
**Figure 2.3** Discrepancies (in percentage) in observed surface maximum sustained winds between different agencies [from (Knapp et al. 2010a)]

Because such a definitive BTD source is not yet available, the only approach is to choose one of the obtainable databases. Since the BTD compiling methods at all the agencies are not fully and detailed described, the selection of one set over another one, without supporting arguments, certainly introduces arbitrariness and can be detrimental to the model. Furthermore, the so-called "global databases" such as the IBTrACS (Knapp et al. 2010b), are not helpful at all, as they only choose one database for each basin among the available ones, without any explanations.

An attempt to evaluate the reliabilities of the available records was recently carried out by Barcikowska et al. (2012), in which different BTD were compared with independent reference data. Two trustworthy sources were chosen as references, namely the Blended Sea Winds database (Zhang et al. 2006) for TCs with low intensities, and the aircraft measurements collected during the THORPEX Pacific Asian Regional Campaign [TPARC-2008 (NOAA 2008)] for extreme

conditions. Figure 2.4 (Barcikowska 2012) presents the comparisons between wind speeds derived from the BTD and the references for different TCs. As can be seen in the figure, the JMA's BTD (denoted as JMADT in the figure) is closer to the references than other records. The JMA is successful in keeping the homogeneity within its database by using the same method and information sources during the entire monitoring period. It makes JMA's record more reliable than other ones to derive TC statistics. Furthermore, the JMA's BTD also provides the observations of a valuable parameter, which is an advantage over other sources in defining model coverage. This benefit will be described in subsection 2.2.2.

In conclusion, the historical record from JMA (JMA 2014) is selected as the BTD in this research due to its superior accuracy. Although the data has been given since 1951, only the observations from 1977, which contain the measurements for both central pressure and maximum sustained wind, are taken into account. The choice of time range (i.e., from 1977) is due to the need for the values of wind speed of historical TCs in the model and the significant improvements in accuracy of estimation methods at JMA since 1977. This time range is popular among the different researchers and is found to be agreed upon in the literature and other studies [e.g., (Barcikowska et al. 2012; Kamahori et al. 2006; Ott 2006)].



**Figure 2.4 Comparisons between data derived from BTD sets and references for different TCs<sup>9,10</sup> [from (Barcikowska 2012)]**

<sup>9</sup> TCs in the comparisons: TC Dolphin in 2008 (Figure 2.4a), typhoon Sinlaku in 2008 (Figure 2.4b), and typhoon Megi (2010) (Figure 2.4c)

<sup>10</sup> "NOAA" line (Figure 2.4a) is satellite-based data from the Blended Sea Winds. SFMR10s and SFMR60s (Figure 2.4b, c) are aircraft observations [1 second values from the Stepped-Frequency Microwave Radiometer (SFMR)] averaged over the periods of 10 and 60 seconds, respectively.

### 2.1.2 Other model input data

In addition to the BTD record, the following data are essential for the estimation of various required parameters as well as to present the model results:

- Data from the Twentieth Century Reanalysis Project (Compo et al. 2011), which are given in 2x2 degrees global square grids, will be used in chapter 3 and chapter 4, to include:
  - Mean monthly values of atmospheric pressure at the mean sea level,
  - Mean monthly values of relative humidity at the "near surface" level,
  - Temperature at the top of the troposphere.
- Mean monthly values of Sea Surface Temperature (SST), which are provided in 2x2 degrees global square grids, and which will be used in chapter 3, are taken from the Extended Reconstructed SST V3b record (NOAA 2014a).
- Digital maps from the 1:110m Cultural Vectors (Natural Earth 2014).

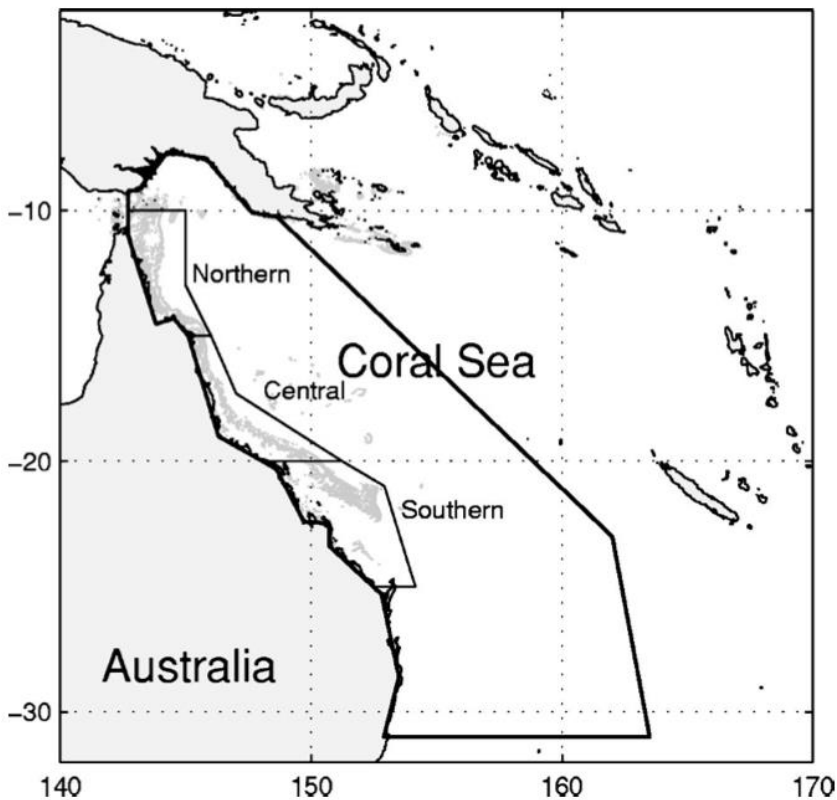
## 2.2 Model coverage

As described earlier in subsection 1.3.3, the model should have a reasonable domain. The "not too small, not too large" research area is defined in two steps. Firstly (subsection 2.2.1), the geographic extent of an AOI is determined, based on location of the case study. Secondly (subsection 2.2.2), the TA (i.e., model domain) is defined, using the AOI in the previous step and the relevant BTD for the case study. Although the concepts of the AOI and the TA were first described in the literature by Powel et al. (2005), the idea has been used in most TC studies, with or without researchers' acknowledgement.

### 2.2.1 Area Of Interest

The AOI is the region where TC's parameters are derived from both historical and simulated data, in order to evaluate model results and assess risks due to TCs. Most of the time, it is a sub-region or a group of different subareas separated by a relatively short distance from the coastline.

However, because reasonable explanations for the selection of the AOI were not given, its geographic range was ambiguous and inconsistent among different studies. For instance, while Hall and Jewson (2007) used the 100-km-radius areas from the landfall locations, Vickery et al. (2000) utilized the 250-km-radius sub-regions from the mileposts along the coastline. Another example is presented in Figure 2.5, in which James and Mason (2005) defined the AOI by the polygonal boundary of the marine park outside the coast.

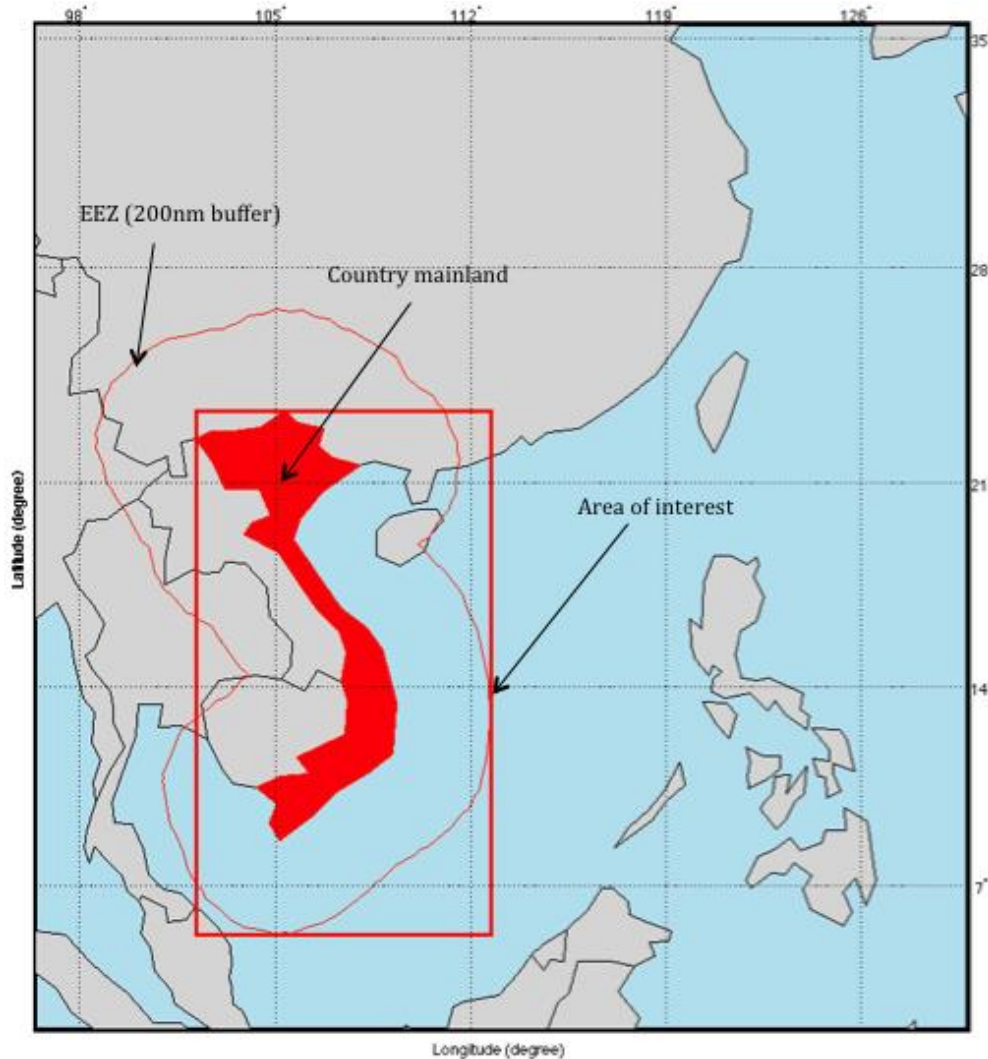


**Figure 2.5 The AOI and the TA for the Queensland coast<sup>11,12</sup> [defined by James and Mason (2005)]**

In this study, the AOI is determined by considering its underlying meaning. That is, basically, an AOI is the region which contains inhabited areas or properties, and which researchers would like to assess risks due to TCs. Those properties can be coastal houses, buildings, infrastructures, and offshore structures, already constructed or planned for the future projects. Therefore, the AOI consists of two parts. The first part is the mainland of a state or country that research will focus on. The second part is the oceanic region where offshore constructions (e.g., windmills or oil rigs) are located or proposed. A suitable choice for this location is the Exclusive Economic Zone (EEZ), which stretches out up to 200 nm from the coast, as defined in the 1982 United Nations Convention on the Law of the Sea [UNCLOS (UN 1982)]. Using this definition, the AOI for the case study of Vietnam is provided in Figure 2.6. As can be seen from that figure, although the seaward boundary of the EEZ is a complicated shape, the AOI is limited by a rectangular, which bounds the EEZ outline. That is because a simple shape of the AOI not only avoids unnecessary complexity in the model, but also ensures that the AOI will be fit for the modelling using polygonal computational grids.

<sup>11</sup> The AOI is defined by the boundary of the Great Barrier Reef Marine Park (GBRMP), which is the polygon adjacent to the coastline.

<sup>12</sup> The seaward border of the TA is about 600 km outside the boundary of the GBRMP (i.e., the AOI)



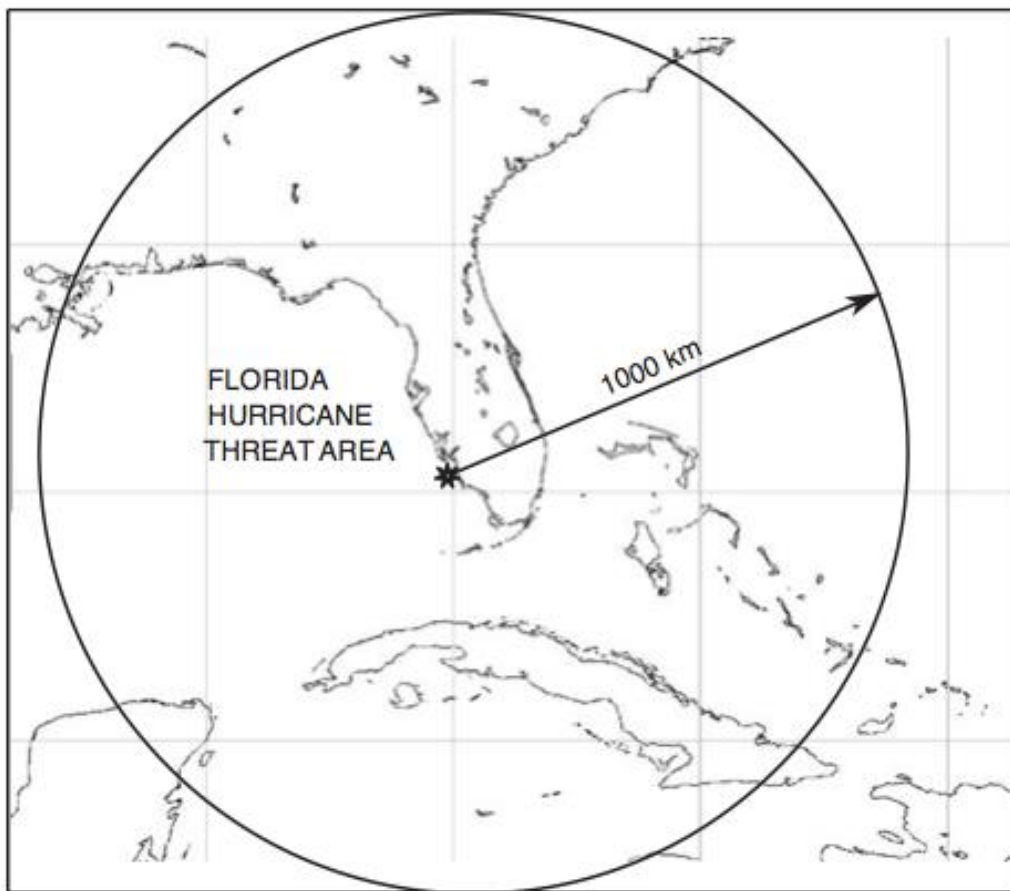
**Figure 2.6 The AOI for the case study of Vietnam**

### 2.2.2 Threat Area

In most studies, the main focus is on the TCs that are capable of influencing conditions in the AOI (Powell et al. 2005). The region, which covers all the centres of TCs affecting the AOI, is called the TA. As already discussed in subsection 1.3.3, if a model is only concerned with historical observations inside the TA, the result will be a homogeneity among the input database, together with a considerable reduction in required time for BTDA analyses. Furthermore, because nearly all TCs originated outside the TA do not enter the TA area and therefore cannot be expected to affect situations in the AOI, the computational demand for model simulation can also be significantly reduced by initiating synthetic TCs only within the TA (James and Mason 2005).

However, in previous studies, no criterion was introduced to determine whether a TC has an impact on the AOI or not. Thus, the researchers either used the entire

basin as the TA for their model [e.g., (Lee and Rosowsky 2007; Vickery, Skerlj, and Twisdale 2000; Wang and Rosowsky 2012)], or defined the region by using a shape separated by a seaward distance from the AOI. In the latter case, because arbitrariness was presented in the choice of TA, its size and shape differed among various studies. For instance, while James and Mason (2005) used an irregular polygon with a seaward distance of about 600 km outside the AOI (see Figure 2.5), Powell et al. (2005) drew a 1000-km-radius circle from a point within the AOI, as shown in Figure 2.7.



**Figure 2.7 The TA for the State of Florida [defined by (Powell et al. 2005)]**

Therefore, there is a desire and a necessity to find a new method to reliably specify the geographic range of the TA in the research. One parameter, which can be a very useful indicator of TC effects on a given region, is the maximum radius of 34 kt wind speed (1-min average). There are two reasons for choosing that criterion.

Firstly, a maximum radius of 34 kt wind speed, is the minimum avoidance distance for all vessels in the vicinity of a TC, as stated in most navigation guidelines [e.g., (NOAA 2014b) shown in Figure 2.8]. An exposure to seas inside the threatening area can dangerously hamper ship manoeuvrability and stability. One important point to keep in mind is that, because the 34 kt wind field is usually asymmetric,



presenting this field as a circle is an oversimplification. However, the definition of a symmetrical wind field is still widely used in "the safer the better" approach.

Secondly, 34 kt wind speed, which is used in most TC scales such as the well-known Saffir-Simpson Hurricane Scale [SSHS (NHC 2014)], is also a threshold to determine the lowest intensity of a tropical storm (NOAA 2014c), as presented in Figure 2.9 (Wikipedia 2014b)]. Thus, this value is a limit of the extreme winds, which most possibly can cause structural damages to the exposed properties.

For the WNP basin, fortunately, the JMA has included such a crucial TC parameter in its BTD since 1977, which formulates a basis for determining the TA in this study. The procedure is carried out in two steps. Firstly, DPs, which are observed after 1977 and which had maximum sustained surface wind speed of 30 kt (10-min average, equivalent to 34 kt 1-min average) or more, are extracted from the BTD. Secondly, a circle is drawn from each of the TC centres by using its maximum 30 kt wind radius. If that circle lies within, entirely covers, or intersects with the AOI, the concerned DP surely has affected the conditions in the AOI. The region, which covers all the centres of TCs (i.e., DP) affecting the AOI, is defined as the TA.

However, all DPs located inside the TA must be taken into account and treated as input BTD for the GSESM even if, historically, they had no impact on the AOI. Because TC's activity depends heavily on local conditions, the DPs still influence TC's characteristics in formation and evolution. Thus, these DPs contribute considerably to the overall TC's behaviour, which form the fundamentals for the simulation. Figure 2.10 provides the step-by-step flow chart outlining the approach used in this research. An example for the case study of Vietnam in the WNP is shown in Figure 2.11 and Figure 2.12. The figures present the geographic extent of both AOI and TA specifically defined for the case study, along with TC centres and their maximum 30 kt wind radii in 1982.

To conclude, a new approach to define the TA and model's input BTD is introduced. The new methodology will prove to be objective following the various supporting arguments. For the first time, a maximum 30 kt wind radius is utilized to determine whether a TC centre has an impact on the AOI or not. While this technique still best captures the statistical characteristics of historical DPs that have affected the AOI, it also effectively reduces the computational demand by removing a large proportion of TC centres, which are completely irrelevant to the research. For instance, Table 2.1 summarizes a number of DPs for various types of TC centres for the case study of Vietnam. As can be seen from that table, the size of the BTD used in this case significantly decreases from 35,106 DPs to 9,275 DPs. This number is only about a quarter of length of the BTD over the entire WNP.

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

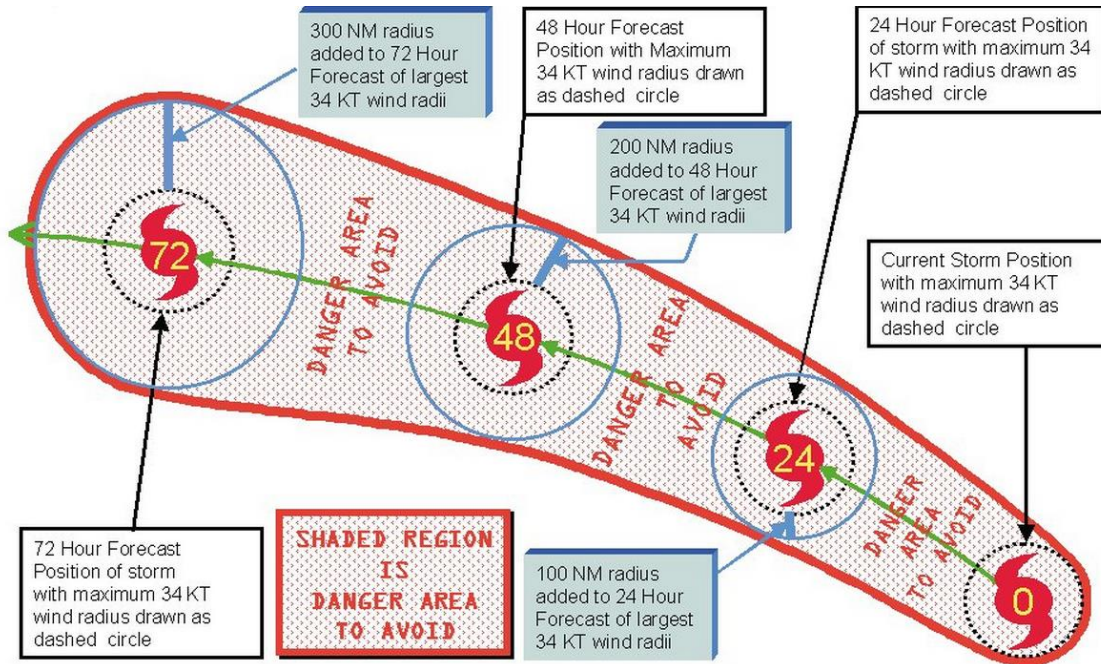


Figure 2.8 Rule of TC avoidance [from (NOAA 2014b)]

Category	Wind speeds
Five	≥70 m/s, ≥137 knots ≥157 mph, ≥252 km/h
Four	58–70 m/s, 113–136 knots 130–156 mph, 209–251 km/h
Three	50–58 m/s, 96–112 knots 111–129 mph, 178–208 km/h
Two	43–49 m/s, 83–95 knots 96–110 mph, 154–177 km/h
One	33–42 m/s, 64–82 knots 74–95 mph, 119–153 km/h

Related classifications	
Tropical storm	18–32 m/s, 34–63 knots 39–73 mph, 63–118 km/h
Tropical depression	≤17 m/s, ≤33 knots ≤38 mph, ≤62 km/h

Figure 2.9 Saffir-Simpson Hurricane Scale [from (Wikipedia 2014b)]

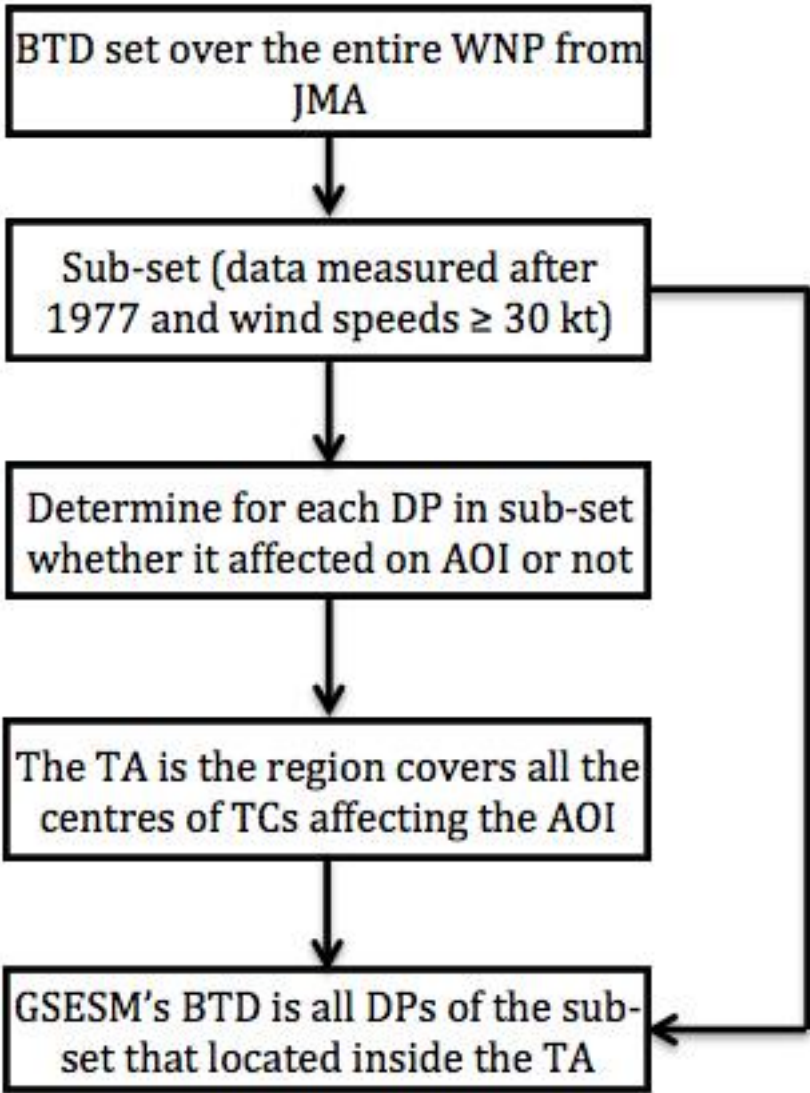
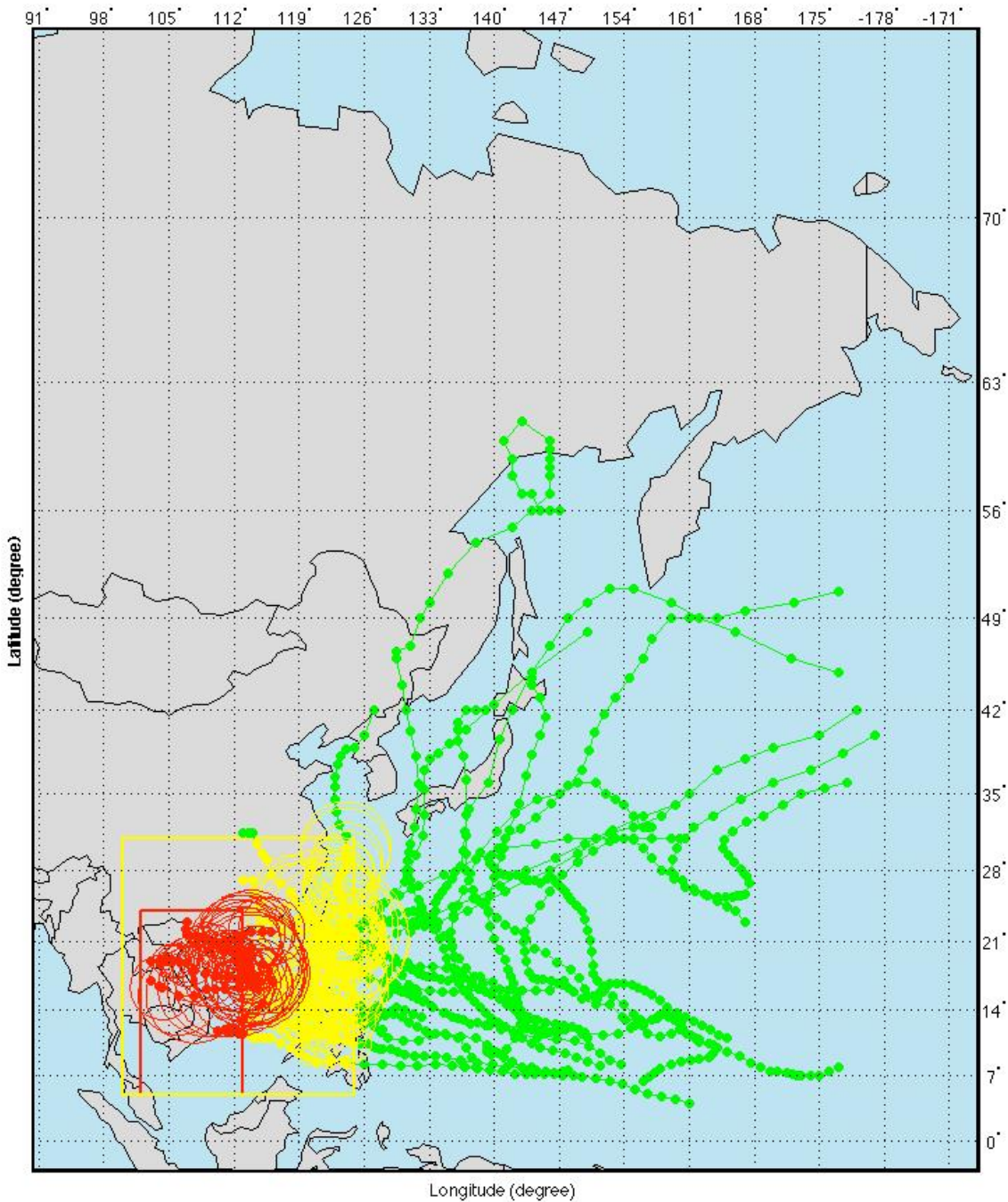


Figure 2.10 Flow chart of the approach to define the TA for the case study

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
Nguyen Binh Minh - 2015



**Figure 2.11 TCs in the WNP of the 1982 season (zoom-out map)<sup>13,14,15,16,17</sup>**

<sup>13</sup> The AOI is the red rectangular and the TA is the yellow rectangular

<sup>14</sup> The DPs, which were not inside the TA, and thus irrelevant for the study, are drawn in green

<sup>15</sup> The DPs, which had impacts on the AOI, and their maximum 30 kt wind radii, are drawn in red

<sup>16</sup> The DPs, which were inside the TA but had no impact on the AOI, and their maximum 30 kt wind radii, are drawn in yellow

<sup>17</sup> The BTD used in this research is a combination of both red and yellow TC's centres

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

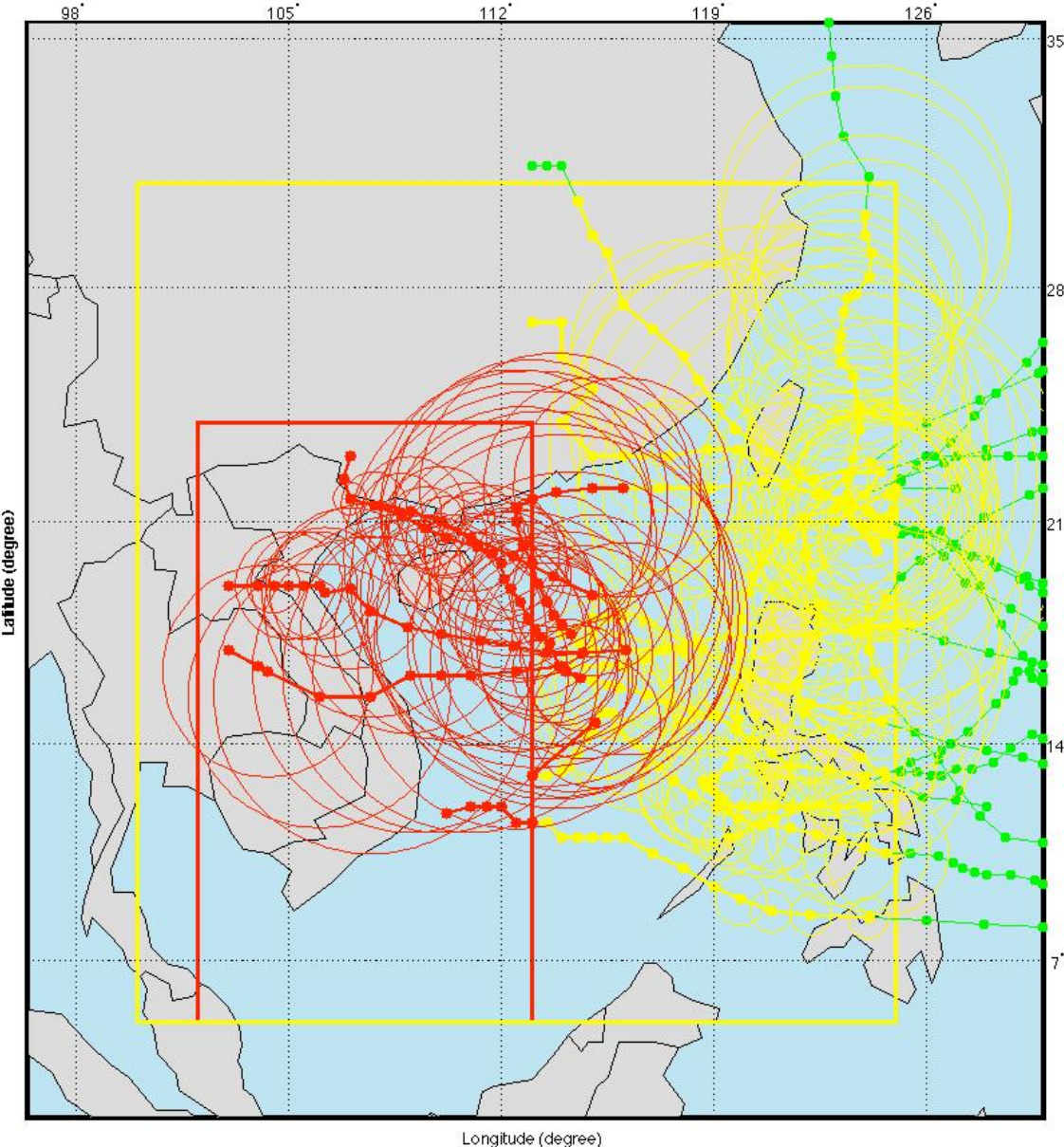


Figure 2.12 TCs in the WNP of the 1982 season (zoom-in map)<sup>18</sup>

Table 2.1 TC's centres for the case study of Vietnam<sup>19</sup>

Type of TC's centres	Entire WNP BTD (6+7+8)	6 (green)	7 (red)	8 (yellow)	GSESM's BTD (7+8)
Number of DPs	35106	25831	3456	5819	9275

<sup>18</sup> Map is zoomed-in for a close-up of the AOI and the TA

<sup>19</sup> Notations for different types of TC's centres comply with footnote numbers in Figure 2.11

## 2.3 Computational grids

In most TC studies, especially risk-related ones, a system of grids (cells) is developed to cover the entire research area (i.e., the TA in this study). The centre of each grid represent all points located inside that cell. Such a system not only remarkably lessens model complicity, and therefore significantly reduces computational demand, but also conforms to research objectives, as risks due to TCs should not be assessed at individual points. Relatively large grids are good enough for primary analyses. However, on condition that detailed calculations are required, finer grids can also be used for a particular sub-region within the model domain, to archive more refined model results. This section deals with the formation of grids used in this study, including estimating the shape (subsection 2.3.1) and size (subsection 2.3.2) of the cells. Unfortunately, in other research, very little attention was given to these factors.

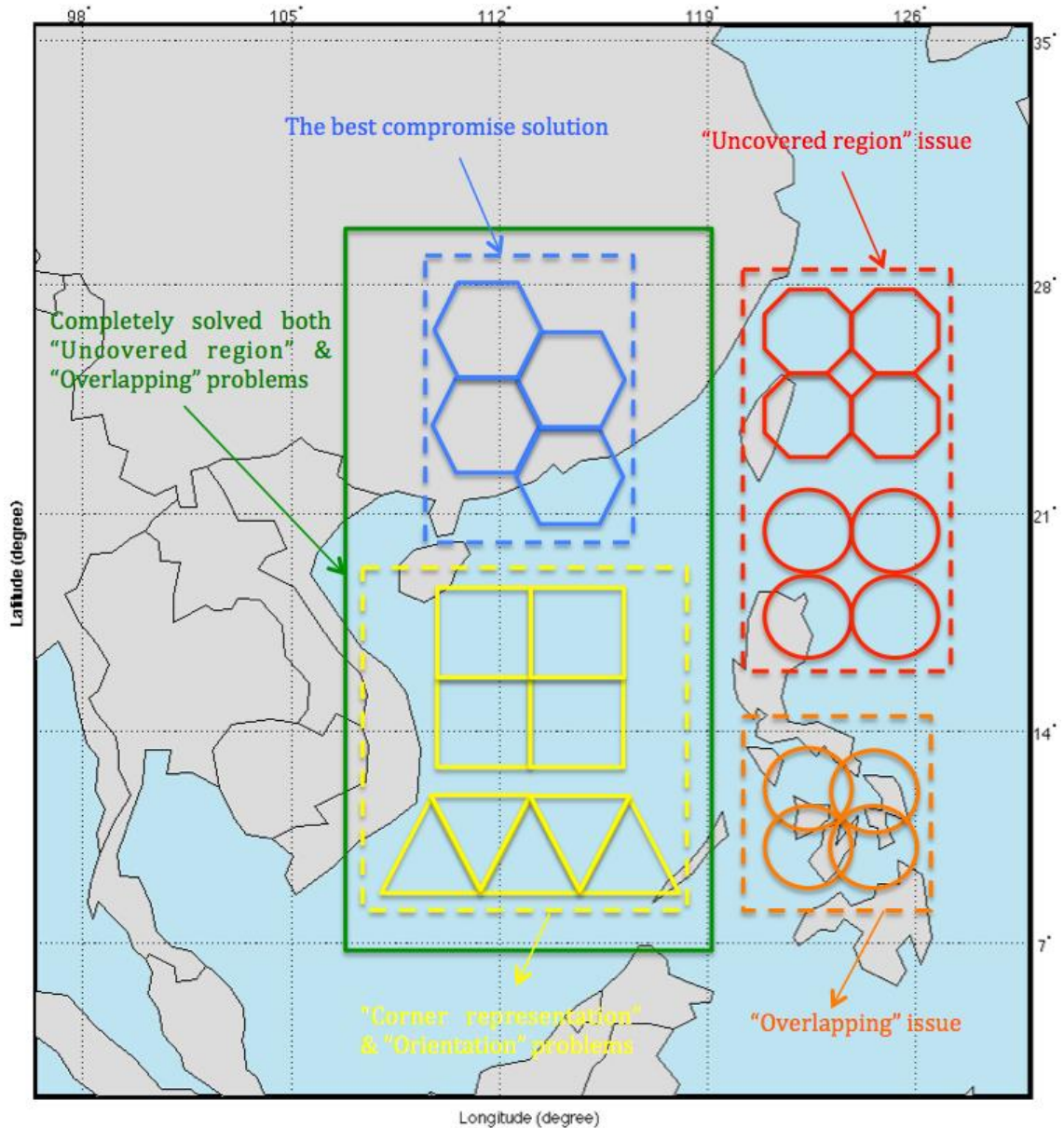
### 2.3.1 Shape

As a grid's centre is the representative for all other points located inside this grid, the shape of the computational cells is important. Several shapes were used in previous studies.

Usually, square grids are utilized in most TC studies [e.g., (Elsner and Kara 1999; Hope and Neumann 1971; Vickery, Skerlj, and Twisdale 2000; Wang and Rosowsky 2012)]. Because locations of TC centres in the BTD sets are recorded in a rectangular coordinate system (i.e., latitude and longitude), dividing model domains by rectangular shape is a quite likely idea (Brettschneider 2008). Nevertheless, using square grids results in two problems. The first drawback is a so-called "corner representation" issue, in which a square's centre cannot equally represent all DPs inside that shape. Take an edge for example, while the shortest distance from the centre to that edge (i.e., length of perpendicular line) is 1 unit, the longest distance to that edge (i.e., length of line connected to corners) is much larger, about 1.4 units. Thus, the corners are overestimated in calculations. Secondly, it also suffers from the "orientation" problem, that is, the number of DPs in a square is changed when one rotates that shape while still keeping its centre.

Another shape, although rarely seen, which is used in other research [e.g., (Ho et al. 1975)], is an octagon. The rationale behind that selection of shape of the computational grids is to solve the above "corner representation" issue. However, it also creates an additional, and even more serious problem, which is the "uncovered region" issue. When one draws an octagon inscribed in a square, that octagon only accounts for 87.5% of the area of the square. A system of octagons generates the diamond-shaped gaps, which occur at every intersection of four nearby octagons, as can be seen in Figure 2.13. Such uncovered regions represent one-eighth of the entire research area (Brettschneider 2006).

Uses of the triangle can also be found in the literature [e.g., (Weisberg and Zheng 2006)]. In that case, like the square, there is no uncovered region. However, both "corner representation" and "orientation" problems still exist and even become more critical than with the square.



**Figure 2.13 Several alternatives for the shapes of computational grids**

Other simple shapes should also be examined. The circle, for instance can be an ideal shape to deal with "corner representation" and "orientation" issues. However, to cover the entire domain with circles, they must overlap with each other. It is obviously not a preferable option since controversies and inconsistencies exist in estimating TC parameters at a great number of DPs in the overlaps. If one decides that no overlap is allowed and adjoining circles is only touch each other, the "uncovered region" issue will appear. The uncovered area in this case is 21.5%. That percentage is nearly double the value when the octagons are used (i.e., 12.5%).

Considering all aforementioned reasoning, the hexagon is used in this study, as it is the best compromise solution for the shape of computational grids. Like the square and triangle, a hexagon absolutely does not have the limitation of uncovered or

overlapping areas. Furthermore, "corner representation" and "orientation" issues are also minimized because, like the octagon, a hexagon has equal lengths of edge and inside angles, and therefore it is a good approximation to the ideal shape of a circle. A summary of all the analysed shapes is presented in Table 2.2. There are still a small number of DPs, which lay exactly on the edges of hexagons. However, in such cases, a DP can be represented by one of the two centres of hexagons, which share that edge. The consistency is still maintained since TC's characteristics heavily depend on local conditions, and therefore the situation of the two contiguous cells must be similar to the other.

**Table 2.2 Summary of various shapes of computational grids**

Grid type		Square	Octagon	Triangle	Hexagon	Overlapped circle	No overlapped circle
Popularity in literature		Many	Few	Fewer	<b>Least</b>	No use	
Rationale		TC positions are reported in latitude and longitude	Reduce the corner representation & orientation issues	Simplest shape	<b>Resolve the uncovered region problem of the octagon</b>	Perfectly solve the corner representation & orientation issues	
Problem type	Corner representation	Worse	Better	Worst	<b>Good</b>	No (Best)	No (Best)
	Orientation	Worse	Better	Worst	<b>Good</b>	No (Best)	No (Best)
	Uncovered region	No (Best)	Worse	No (Best)	<b>No (Best)</b>	No (Best)	Worst
	Overlapping	No (Best)	No (Best)	No (Best)	<b>No (Best)</b>	Worst	No (Best)

### 2.3.2 Size

In other studies, an arbitrarily, but well liked grid size of 5 degrees, was used for a basin-wide domain [e.g., (Lee and Rosowsky 2007; Vickery, Skerlj, and Twisdale 2000; Vickery, Wadhwa, Twisdale, et al. 2009; Wang and Rosowsky 2012)] without any given reasons. It is an important aspect, which this research would like to address. However, unlike the shape of computational grids, the dimensions of these cells cannot be specified at the preparation stages of the model, because the influences of grid size on model performance are only revealed in the evaluation of several fundamental research equations. Thus, the most suitable value of grid size will be defined in chapter 3 and chapter 4.

Another point to keep in mind is that, there is no need for creating a system of equal area grids, as suggested by some researchers [e.g., (Brettschneider 2008)]. Although areas of the cells defined by spherical coordinates (i.e., the unit of grid size is degrees) are changed with latitude, the choice of using a regional domain nearly eliminates that issue. The reasons are that varied areas can only be observed with large differences in latitude range (e.g., more than 30 degrees, which occur in the basin-wise studies), and defining computational grids by spherical coordinates is definitely a more natural and logical approach.



# 3 TRACK AND CENTRAL INTENSITY MODEL

## Research questions:

- Which are the current theoretical frameworks for modelling key parameters? What are their pros and cons? Which one should be chosen as a basis for the Generalized Stochastic Empirical Storm Model (GSESM)?
- What are the limitations of the (chosen) existing method? How can the GSESM overcome these limitations? In comparison with the original technique, what are the GSESM's improvements?

There is no doubt that a Tropical Cyclone (TC) is a natural phenomenon, which consists of numerous complicated processes. Unfortunately, the current knowledge is not yet sufficient to reveal all physical principles that govern these procedures. Even when TCs can be fully described, TCs' complexities still prevent researchers from including all of these physical phenomena in any long-term simulation, as they would behave in reality. In this case, TCs must be represented by several relatively simple (in comparison with the reality) mathematical expressions, which obviously have to be introduced along with certain assumptions. The formulae are categorized into two groups, which can be combined together to approximate the real TC's behaviour with an acceptable accuracy for practical applications.

This chapter deals with the first set of equations for modelling the central track and intensity. In section 3.1, background information is provided as well as the most common research assumptions. Following, in section 3.2, various available methodologies are described, analysed and discussed. One of the existing approaches is then chosen as a basis for the model developed in this study (i.e., the GSESM), based on conclusions from this section. Finally, the complete GSESM's

methodology is formulated by clarifying and gradually improving the existing method with detailed examinations and comparisons of model performances, in the modelling of both central track (section 3.3) and intensity (section 3.4).

### 3.1 Background

After a TC is initiated, the interest centres on providing enough information to answer the two critical and most visible questions that are, at any given moment, where the TC is and how strong it is. The first question, unfortunately, has no explicit solution since a TC actually does not concentrate on any single point. Instead, a TC spreads out its effects over a broad region. However, the problem can be diminished by simplifying such a region into a more compact area (i.e., the TC's eye), namely where the most extreme conditions occur. A TC's strength (i.e., maximum sustained surface wind speed or minimum central pressure) is determined at its eye and can be considered as an estimation of intensity. Moreover, in all databases, a TC's eye is further represented by only one point, which is the eye's centre.

Because the above procedure cannot be performed continuously at every moment, it is carried out periodically at each time step. The interval that separates those steps is selected to balance the rise in computational demand (when one reduces that period to produce a lower variation in conditions between two contiguous steps) with the decrease in data accuracy and continuity (when the interlude is stretched longer). In this study, a 6-hour interval is used since it is a standard, which is employed not only in all historical records [i.e., the Best Track Data (BTD) sets] but also in every single simulation.

Thus, information of location and strength provided by each Data Point (DP) in a TC compilation are, in fact, positions of the eye's centre and the most extreme conditions (i.e., maximum sustained surface wind speed or minimum central pressure), which can be obtained within the eye of a TC. Furthermore, in a literal sense, there is no track at all because the database only gives a series of DPs, which are recorded at every 6-hour time step. However, the implicit concept of a track, which can be understood as an implied line connecting all DPs of the same TC, is still widely used as it helps people to visualize TC's activity geographically in an intuitive way. Because a trajectory is interpreted as a line, linear interpolation can be utilized between the 6-hour DPs in many studies [e.g., (Vickery, Skerlj, and Twisdale 2000)] when more precise locational information is required. However, the simple notion of a linear track is clearly a simplification since numerous fluctuations must exist between the TC's positions at two contiguous time steps (Brettschneider 2006).

Because a BTD is the data source for most TC climatology researches, especially the long-term simulations, several warnings related to BTD's accuracy must be pointed out. Firstly, when a TC is monitored in the BTD, sometimes only sporadic information is available. In those cases, a "best guess" or more specifically, TC climatology (i.e., previous understanding of TC statistics) is employed to fill in the gaps. Although this approach maintains the continuity among DPs within the database, it definitely introduces errors to the calculations. Secondly, since the main technique for assembling BTD sets (i.e., the Dvorak method) has many

limitations (see subsection 2.1.1), both the observed position and the intensity are rounded. While TC's coordinates make use of one decimal place, which correlates with the accuracy of 8.3 to 12 km depending on latitude, the wind speed and the central pressure are rounded to the nearest 5 kt and 1 hPa, respectively. On the other hand, TC simulations do not suffer from these issues. While the first problem (i.e., only sporadic information is available) is inherent in only historical measurements, the second limitation (i.e., values are rounded) does not exist at all in any model with efficient support of computer power.

However, because a TC database, regardless if it is observed or simulated, is a simplification of the real conditions, several assumptions must be presented. For a long-term simulation, two most important hypotheses, which not only reduce research complexity to make formulae feasible but also frame the basic feature of the model (i.e., a probabilistic study), are presented as follows:

- Every single TC is a discrete realization of the same fundamental stochastic procedure (James and Mason 2005; Lee and Rosowsky 2007).
- The evolution of a TC track and the intensity of the TC are two distinct processes and therefore, treated independently (Brettschneider 2006). However, these processes are often considered simultaneously as their combination gives a more complete picture of TC's activity.

## 3.2 Probabilistic model

As already discussed earlier, in subsection 1.3.3, only probabilistic models are applicable to TC hazard estimation. Nowadays, probabilistic models are utilized in various sectors for the design of properties and for assessment of risks associated with TCs. The simulations are also widely used by insurance firms and banks to assign insurance levels in storm-prone areas. In regard to building codes, probabilistic models are applied in the establishment of design wind speed maps, such as the ANSI A58.1 (ANSI 1982), ASCE 7-93 (ASCE 1993) and ASCE 7-05 (ASCE 2006) in the U.S., the CUBic (1985) in the Caribbean, and the AS/NZS1170.2-2002 (AS/NZS 2002) in Australia. Although numerous models have been employed, they are all based on two distinct underlying approaches (Apivotanagul et al. 2011). These underlying methods are the single site stochastic technique, which does not simulate the entire TC tracks, and the empirical track approach that imitates the full trajectories.

### 3.2.1 Single site probabilistic simulation

The first TC simulation described in the literature, was carried out by Russell (1968, 1971). Following the basic ideas of Russell's study, other researches were conducted using similar approaches [e.g., (Batts et al. 1980; Georgiou 1985; Georgiou et al. 1983; Neumann 1991; Tryggvason et al. 1976; Vickery and Twisdale 1995b)]. First, statistical distributions of key parameters, typically including heading, translation speed, radius to maximum wind, central pressure deficit, and distance of the closest approach or coast crossing position, are estimated for a particular location. Each distribution is then sampled by the Monte Carlo method to form the initial condition of a synthetic TC. Such a simulated TC is

assumed to make landfall forward in a straight-line course with a constant intensity until landfall. After the moment of landfall, the TC is weakened as specified by a decay model [e.g., (DeMaria et al. 2006)]. Thus, although those studies vary in some of the details (e.g., wind field and decay models used, size of domain, or types of fit distributions), the basic principles were not changed.

Although effort has been put into expanding and improving the Russell's pioneering technique, single site simulations are facing some critical, and in some cases unacceptable, problems. Firstly, Russell's method is only valid for a small region since all necessary statistics are derived from site-specific information. Thus, it is impossible to carry out TC analyses for a relatively large area (e.g., a state or a small country) or for several sub-regions at the same time. Secondly, basic hypotheses of a straight-line path (i.e., constant heading) and an invariable intensity until landfall, limit the essential variations in TC developments and can be a considerable detriment to the model. Moreover, a single site method forces the key parameters to comply with theoretical distributions *a priori*, without the full knowledge of their properties and possible fluctuations. The Monte Carlo sampling also contains an inherent issue. Because TC's parameters are individually sampled, the resulting simulated scenarios can include incoherent values, which are completely different than the realistic combinations. Providing that one can accept all aforementioned drawbacks, the serious shortages of observed data in many important areas still prevent the effective applications of Russell's technique in deriving meaningful site-specific TC statistics (see subsection 1.3.2).

### 3.2.2 Empirical track modelling

In the Empirical Track Modelling (ETM), introduced by Vickery et al. (2000), a full track of each synthetic TC is simulated, from its initial point over water to the lysis. As the TC propagates, the heading, translation speed and central pressure are estimated at every 6-hour time step. A wind field model is utilized to compute surface wind speeds at any positions inside the vicinity of the TC. The ETM overcomes the disadvantages of the single site approach. Firstly, because the whole tracks are used, the number of available DPs for estimating TC's parameters has significantly increased, therefore providing more reliable statistics. Since tracks are present over a broad region, the research area is not restricted to a small site. Secondly, because all key parameters are updated at every 6-hour time step, this technique retains the important variations in the evolution of the central track and intensity. Thirdly, the method does not predefine any theoretical distributions and therefore does not require thorough understanding of a TC's behaviour, which is currently not available. Finally, using the ETM, initial points of the synthetic TCs are directly based on those of the historical TCs. Moreover, mathematical formulae of TC development express close correlations between nearby DPs. Thus, the ETM maintains the coherence in synthetic combinations of key parameters, the inherent temporal relationships between different DPs of the same TC, and the spatial dependence of TC's activity on local conditions.

The only potential drawback of the ETM compared to the single site approach is that, it involves creating a large modelled database of thousands or millions of events, which may require an enormous computational demand. Nevertheless, this disadvantage is becoming less crucial due to the remarkable advances in computer

power. Although one cannot expect any miracles, because the ETM is still far from reality, this approach represents the state-of-the-art in TC long-term simulation (Vickery, Masters, et al. 2009). Most recent studies are based on ETM [e.g., (Emanuel et al. 2006; Hall and Jewson 2007; James and Mason 2005; Lee and Rosowsky 2007; Powell et al. 2005; Rumpf et al. 2007; Vickery, Wadhera, Twisdale, et al. 2009; Wang and Rosowsky 2012)]. Thus, the ETM is chosen as a basis for the GSESM, whose general approach is described as follows:

- First of all, the total number of synthetic TCs is estimated. This quantity can directly be indicated [e.g., 100,000 events in (James and Mason 2005)] or it can indirectly be defined by first determining a number of years in the model [e.g., 10,000 years in (Wang and Rosowsky 2012) or 15,000 years in (Lee and Rosowsky 2007)]. Then one can sample the distribution of annual occurrence rate of historical TCs to get the number of TCs that have to be simulated each year.
- A set of initial conditions (or more specifically, starting location, time, heading, translation speed, and intensity) of all TCs in the BTM is used as a source to generate the first DP of each synthetic TC.
- For the subsequent time step, mathematical expressions are utilized to calculate the new central position and intensity based on changes in those parameters over the current 6-hour interval. The simulation is carried out for all synthetic TCs and constitutes a full model's database.
- Since the procedure is repeated until a simulated TC comes to its lysis, the whole track of this TC is generated.
- When needed, the details of the surface wind field at each DP along the track can be derived from the local conditions by using a wind field model.

### 3.3 Track modelling

The foremost component of an ETM-based study is the track modelling. This element is so important that some researchers primary focus on it [e.g., (Rumpf et al. 2007)]. Other people even exclusively examine this element [e.g., (Hall and Jewson 2007)].

#### 3.3.1 Original technique

In the ETM approach as used by Vickery et al. (2000), giving the "known" conditions at time step  $i-1$  and  $i$ , the "unknown" situation at the time step  $i+1$  can be estimated from equations 3.3 and 3.4 by first quantifying the changes in translation speed and heading over the current period as follows:

$$\Delta \ln c = a_1 + a_2 \psi_i + a_3 \lambda_i + a_4 \ln c_i + a_5 \theta_i + \varepsilon \quad (3.1)$$

$$\Delta \theta = b_1 + b_2 \psi_i + b_3 \lambda_i + b_4 c_i + b_5 \theta_i + b_6 \theta_{i-1} + \varepsilon \quad (3.2)$$

Where:  $c$  - translation speed (km h<sup>-1</sup>);  $\theta$  - heading (degrees, clockwise from north);  $\psi$  and  $\lambda$  - latitude and longitude (degrees) of the TC's centre, respectively;  $\varepsilon$  - random error term. Symbol  $\Delta$  denotes the change over the current interval. Thus, TC parameters at time step  $i+1$  are obtained from:

$$\Delta \ln c = \ln c_{i+1} - \ln c_i \quad \rightarrow \quad \ln c_{i+1} = \ln c_i + \Delta \ln c \quad (3.3)$$

$$\Delta \theta = \theta_{i+1} - \theta_i \quad \rightarrow \quad \theta_{i+1} = \theta_i + \Delta \theta \quad (3.4)$$

Subscripts  $i-1$ ,  $i$ , and  $i+1$  of the values denote which time step is mentioned in the corresponding parameters.  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ , etc. are coefficients, depending on conditions at the current location.

As can be seen from the formulae, the compositions of these expressions satisfy all requirements as described earlier:

- Relatively simple formulae in comparison with the reality (i.e., linear equations).
- Retaining inherent temporal relationships between different DPs of the same TC, because the condition at time step  $i+1$  is derived from conditions of the previous time steps (i.e.,  $i-1$  and  $i$ ). The idea is similar to the underlying meaning of the "persistence" technique (see section 2.1), because logically, there must be a close correlation between the situations at two contiguous time steps, providing that the interval between them is small enough.
- Maintaining spatial dependences of TC's activity on local conditions, since all coefficients in the equations are estimated based on the condition at current location.

The ETM defines the sets of coefficients  $a$  and  $b$  in equations 3.1 and 3.2 through the following process:

- First, the whole research area is divided into various  $5^\circ \times 5^\circ$  square grids.
- For each grid, all historical DPs located in that cell are collected. Because different sets of coefficients for westbound and eastbound DPs are used, that collection of DPs is further separated according to TC's headings.
- The multiple linear regression analyses are performed on the observed records to determine two sets of coefficients, which will be applied to any westbound and eastbound simulated DPs that occur in the current grid.
- The above steps are repeated for every cell in the model domain. For those grids with little or even no historical DPs, which are generally called lack-data cells, the coefficients are assigned as the corresponding coefficients of the nearest grids.

An identical technique was utilized in other researches without any changes [e.g., (Lee and Rosowsky 2007; Wang and Rosowsky 2012)]; others involved certain modifications [e.g., (Emanuel et al. 2006; Hall and Jewson 2007; James and Mason 2005)]. However, none of these studies included either a detailed analysis of the ETM (Vickery, Skerlj, and Twisdale 2000) or any direct comparison with the ETM's methodology. Thus, ETM's weaknesses, which are the basis for possible expansions or improvements, were not revealed and the provided modifications (if any) could not be regarded as beneficial or harmful. Furthermore, because the ETM was developed particularly for simulating TCs in the Atlantic, it is questionable whether the approach is applicable to other basins.

### 3.3.2 Clarifications and possible improvements

The remaining ambiguousness of the ETM can be clarified by answering several questions, including:

Questions related to method background and clarifications (subsection 3.3.3):

- Which criterion (standard) can be used to determine whether a scenario has a better performance than the others?
- If there are several grids that have the same distances to a lack-data one, which cell should be chosen as a source to provide coefficients in equations 3.1 and 3.2 for DPs occurring within that lack-data grid?
- Are there any insurmountable limitations of the mathematical expressions? Which DPs cannot be included in the computations?

Questions associated with possible improvements (subsection 3.3.4):

- Because there is no convincing evidence to prove any differences between the characteristics of the eastbound and westbound DPs, is it necessary to use two distinct sets of coefficients in equations based on headings?
- As the specific type of linear regression was not described in the ETM, are there any benefits of utilizing a different regression solution in the GSESM?
- Do the compositions of variables in equations 3.1 and 3.2 give the best performance? Which combinations of the available parameters are the most effective ones for the simulation?
- Can the method be improved by employing an alternative system of grids with a different shape and size, instead of the one used in the ETM?
- In comparison with the ETM, how much improvement is gained in the GSESM?

These questions will be addressed in the subsequent subsections. Using a step-by-step approach, the ETM is clarified and gradually improved in order to establish the complete methodology of the GSESM. An important point to reiterate is that, the GSESM is a universal method. Thus, in this dissertation, if any details are indicated, they are only applicable for the specific case study. Once users assign another Area Of Interest (AOI), every single particular will be automatically redefined based on this user-defined AOI by using a computer program.

### 3.3.3 Method background and clarifications

#### 3.3.3.1 Indicator of enhancement (Principle 1)

To determine whether a scenario has a better performance than the others, one most important criterion must be selected as an indicator of enhancement. Since the given formulae are in linear form, the coefficient of determination ( $R^2$ ) is a logical choice in the GSESM. It is also widely used in many works to directly compare various possible options [e.g., (Vickery 2005)]. In case of numerous combinations of available variables are being examined, the adjusted  $R^2$  is employed instead of the original  $R^2$ . While the original  $R^2$  always automatically and

spuriously increases when extra variables are added to the formulae, the adjusted  $R^2$  is modified for the number of explanatory terms relative to number of DPs. Thus, the concept of adjusted  $R^2$ , introduced by Theil (1961), is a comparative measure of suitability of different nested sets of variables, and it is (or will be) especially helpful in the feature selection phase of model development.

Another standard, which was described in the literature [e.g., (Vickery, Wadhwa, Twisdale, et al. 2009)], is the mean value of the error term. However, this benchmark is less crucial than the  $R^2$  and it is often suitable for a comparison between alternatives with relatively large differences. Furthermore, because performance of linear regression must be evaluated simultaneously for both equations 3.1 and 3.2, which include two distinctly observed responses in different units, the absolute mean error cannot be employed to compare different options in track modelling.

Since a specific scenario, in comparison with other scenarios, can perform better at some grids while performs worse at other grids, an average value of the criterion must be used to evaluate distinct options over the entire research area. This average is derived from:

$$A = \frac{\sum_{i=1}^N V_i n_i}{\sum_{i=1}^N n_i} \quad (3.5)$$

Where:  $A$  – average value of the standard (e.g.,  $R^2$ , or adjusted  $R^2$ ) for the whole model domain;  $N$  – total number of grids;  $V_i$  and  $n_i$  are value of the criterion and number of DPs in grid  $i$ , respectively.

### 3.3.3.2 Selection of substitutive grids (Principle 2)

The ETM stated that a nearest grid is used as a substitute (i.e., a source to provide coefficients in equations 3.1 and 3.2) for a lack-data cell (i.e., a grid with little or even no historical DPs). Because the maximum number of unknown coefficients in those formulae is 6, linear regressions require at least 6 DPs to derive solutions. Thus, any grids containing less than 6 DPs are lack-data cells.

However, in some cases, the observations are so rare that all adjacent grids of a lack-data cell are also lack-data grids. Thus, the GSESM introduces two guidelines to clarify the definition of a substitutive grid for a lack-data cell:

- The foremost guideline relates to the distance between two grids. Grids associated with the shorter distances have higher priority.
- Among the cells with the same distances to a lack-data grid, the one with largest  $R^2$  is chosen as the substitutive grid.

Figure 3.1 visualizes this concept in an example map. Providing that grid 1 is a lack-data cell, using the first guideline, the cells with highest priority are grids 3, 5, 7, and 9. Grids 2, 4, 6, and 8 will only be considered if all cells with a higher priority are also the lack-data grids. Then the grid with largest  $R^2$  among grids 3, 5, 7, and 9 is selected as the substitutive cell for grid 1.

Furthermore, when comparing various scenarios, on condition that all other factors are similar, an option with smaller number of lack-data grids is clearly a preferable choice. That is because more DPs can be simulated using the coefficients



derived from their own locations, instead of borrowing these coefficients from somewhere else.

### 3.3.3.3 Model limitations (Principle 3)

The mathematical expressions for the track model include some insurmountable problems, so that several DPs have to be excluded from the calculations:

Firstly, DPs where TCs do not move are invalid because at those locations, translation speeds are equal to zero but headings are indefinable. It is called limitation 1.

Secondly, TCs containing 3 or less than 3 DPs are also excluded from the model, due to limitation 3 below. It is called limitation 2.

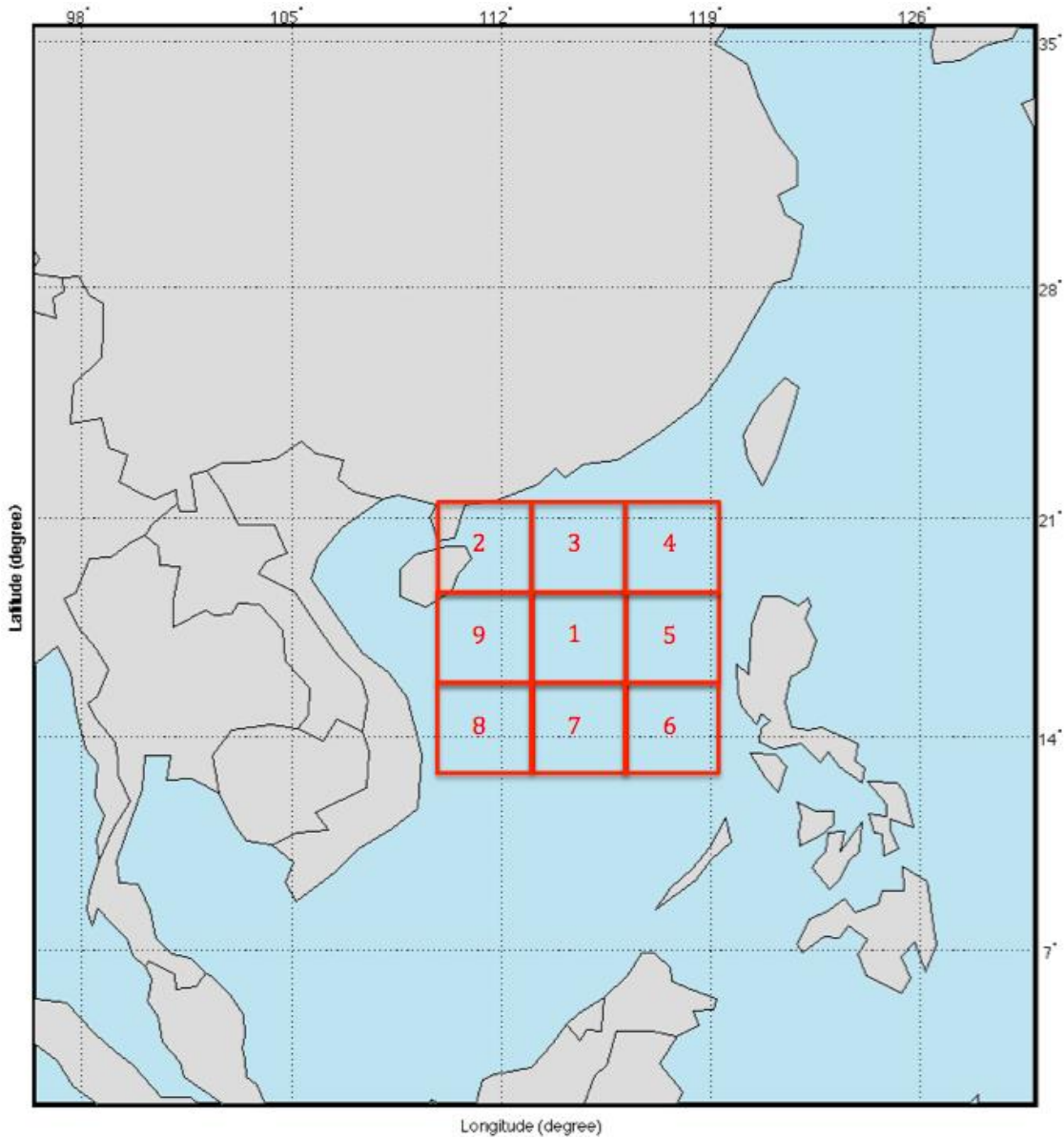
Finally, the translation speed and heading at every DP (e.g., at time step  $i$ ) are defined as the functions of conditions of that DP and the previous one (i.e., at time steps  $i-1$ ). Thus, DPs at time step 1, 2, and the last step of each TC are also removed from the calculations. It is called limitation 3. The reasons for that elimination are:

- At time step 1: latitude and longitude at the previous time step are indefinable, thus translation speed and heading at that time step (in equations 3.1 and 3.2) cannot be obtained.
- At time step 2: heading at the previous time step (i.e., at time step 1) is indefinable, as already indicated above. Such value, therefore, cannot be employed in equation 3.2.
- At the last time step of each TC: geographic coordinates at the next time step are indefinable, hence translation speed and heading at the next time step (in equations 3.3 and 3.4) cannot be acquired.

In Table 3.1, a detailed breakdown of various types of DPs, for the case study of Vietnam in the Western North Pacific (WNP), is given. The table shows that, among 17,935 DPs that were recorded in the WNP since 1977 (the reasons for that choice of time range is presented in subsection 2.1.1), there are 6,141 DPs that are usable for track modelling. Those DPs are obtained after excluding several DPs due to model's limitations, as well as DPs outside the Threat Area (TA) or that have maximum sustained surface wind speed ( $V_{smax}$ ) smaller than 30 kt (supporting arguments for selecting that threshold are provided in subsection 2.2.2).

Because 729 DPs inside the TA are removed due to several model's limitations, an analysis of those DPs is necessary to determine whether that elimination has any significant effects on the model or not. Table 3.2 presents that examination, in which DPs are categorized by different types according to the Japan Meteorological Agency (JMA). The table shows that most of those 729 DPs (i.e., 709 DPs, which account for 97.26% of the total) have the strengths of a tropical depression or weaker (i.e., values of  $V_{smax}$  are equal to or smaller than 30 kt, denoted by types 2 and 6 in the table), and therefore have no impact on the results. The model only loses 20 DPs (denoted by types 3, 4, and 5 in the table), a rather small number (i.e., less than 0.5%) in comparison with the total number of DPs that are usable for the track modelling (i.e., 6,141 DPs).

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015



**Figure 3.1 Example map showing principles to define the substitutive grids**

**Table 3.1 Analysis of historical data (track model)**

Number of DPs									
WNP BTD	Removed due to model limitations				Leftover	Outside the TA	Inside the TA		
	No. 1	No. 2	No. 3	Total			Total	Tropical storm or stronger (winds ≥ 30 kt)	Tropical depression
17935	138	0	1542	1680	16255	7709	8546		
				Outside the TA			Inside the TA		
				951			729	6141	2405

**Table 3.2 Exclusion of DPs in the TA (track model)**

DP type	Number of DPs	Percentage
1	0	0.00
2	689	94.51
3	15	2.06
4	2	0.27
5	3	0.41
6	20	2.74
7	0	0.00
8	0	0.00
9	0	0.00
Total	729	100

### 3.3.4 Possible improvements

In this subsection, a step-by-step process is implemented. At each stage, two or more scenarios are evaluated to find out the optimal one. The options are different in only one parameter (or technique), other factors are identical. The optimal solution determined in a step will be used as the definite technique in the next steps. Through this procedure, one can gradually improve the ETM's approach (Vickery, Skerlj, and Twisdale 2000) to establish the GSESM's methodology.

The improvements are visualized in various tables, in which the starting condition of the corresponding stage is marked in blue colour, and the optimal solution defined at the end of that step is in boldface and marked in red colour. Note that, the coefficients of a lack-data grid (i.e., a grid containing less than 6 DPs) are assigned as the corresponding coefficients of the substitutive cell. Therefore, in these tables,  $R^2$  of a lack-data grid is taken equal to  $R^2$  of the corresponding substitutive cell.

#### 3.3.4.1 Dependence of coefficients on headings (Stage 1)

Because there is no convincing evidence to prove any differences between the characteristics of the eastbound and westbound DPs, it is questionable whether two distinct sets of coefficients are applied to eastbound and westbound DPs. Thus, two scenarios are examined. The first alternative is the ETM's technique, in which DPs are separated by their headings and two different groups of coefficients are obtained. In the second option, all DPs (regardless of eastbound or westbound) are evaluated at once to estimate only one set of coefficients. All other details are identical to the ones employed in the ETM, including type of linear regression solution that assumed to be Ordinary Least Squares (OLS), compositions of variables, and shape and size of computational grids (i.e.,  $5^\circ \times 5^\circ$  square grids). As presented in Table 3.3, for the case study of Vietnam, the second scenario performs better (i.e., larger average  $R^2$ , computed by equation 3.5, see principle 1 described in paragraph 3.3.3.1). This option also has a smaller number of lack-data grids. While a general mean (for both equations 3.1 and 3.2) of the average  $R^2$  and

number of lack-data cells in the first choice are 0.1480 and 15, respectively, those values for the second option are 0.1524 and 7.

#### 3.3.4.2 Linear regression solution (Stage 2)

Although the OLS technique is employed in the previous stage, the type of linear regression was not described in the ETM. Thus, different regression solutions are considered in the GSESM, as shown in Table 3.4 for the case study of Vietnam, including the OLS and various robust regression approaches. The optimal option defined in the previous step (i.e., all DPs, regardless of eastbound or westbound, are evaluated at once) is used; others are kept identical to the one employed in the ETM. As can be clearly seen from the table, while the solution that provides the best performance (i.e., largest average  $R^2$ , computed by equation 3.5, see principle 1 described in paragraph 3.3.3.1) for equation 3.1 is the "talwar" robust regression, the optimal one for equation 3.2 is the OLS.

The rationale behind the invention of the robust fitting methods relates to the validity of a linear regression, which must be based on certain assumptions, such as a normal distribution of errors in the observed responses. If the distribution of errors is asymmetric or prone to outliers, model assumptions are invalidated, and therefore parameter estimates, confidence intervals, and other computed statistics become unreliable. Using a special method, the robust regression is not much affected by outliers and less sensitive than the OLS to the large changes in small parts of the data. Robust regressions work by assigning a weight to each DP. Weighting is done automatically and iteratively using a process called "iteratively reweighted least squares". In the first iteration, each DP is assigned equal weight and model coefficients are estimated using the OLS. At subsequent iterations, weights are recomputed so that points farther from model predictions in the previous iteration are given lower weights. Model coefficients are then recomputed using weighted least squares. The process continues until the values of the coefficient estimates converge within a specified tolerance.

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

**Table 3.3 Scenarios based on headings (track model)**

Scenario	Westbound/Eastbound DPs								All DPs					
	3.1				3.2				Substitute for lack-data grids	3.1		3.2		Substitute for lack-data grids
	Westbound		Eastbound		Westbound		Eastbound			All DPs		All DPs		
Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>
1	19	0.0383	4	0.4645	19	0.1544	4	0.3764	3		23	0.2091	23	0.1302
2	31	0.1652	3	0.4645	31	0.1418	3	0.3764	3		34	0.1660	34	0.1107
3	46	0.2383	19	0.4645	46	0.3977	19	0.3764			65	0.1807	65	0.3246
4	14	0.5528	0	0.4645	14	0.6680	0	0.3764	3		14	0.5528	14	0.6680
5	13	0.3775	0	0.6937	13	0.7771	0	0.6235	10		13	0.3775	13	0.7771
6	5	0.2984	0	0.5763	5	0.2869	0	0.3679	11 & 12		5	0.2984	5	0.2869
7	65	0.0918	0	0.5763	65	0.1231	0	0.3679	12		65	0.0918	65	0.1231
8	291	0.1005	34	0.2166	291	0.1154	34	0.3698			325	0.1313	325	0.1735
9	310	0.1821	27	0.1382	310	0.1345	27	0.1095			337	0.1256	337	0.0862
10	336	0.0729	11	0.6937	336	0.0872	11	0.6214			347	0.1259	347	0.0573
11	24	0.2984	0	0.5763	24	0.2869	0	0.3679	12		24	0.2984	24	0.2869
12	340	0.1436	13	0.5763	340	0.1508	13	0.3679			353	0.1616	353	0.2329
13	584	0.1557	81	0.2330	584	0.1202	81	0.1373			665	0.1582	665	0.1212
14	584	0.1151	212	0.2009	584	0.1270	212	0.1449			796	0.1441	796	0.1351
15	543	0.0850	135	0.1658	543	0.1686	135	0.1978			678	0.1197	678	0.1626
16	8	0.6378	0	0.3859	8	0.5608	0	0.6238	17		8	0.6378	8	0.5608
17	148	0.1089	11	0.3859	148	0.1732	11	0.6238			159	0.1158	159	0.1636
18	271	0.1277	46	0.1332	271	0.2610	46	0.2312			317	0.1233	317	0.2765
19	304	0.1837	165	0.1317	304	0.1128	165	0.1204			469	0.1462	469	0.1682
20	595	0.1201	273	0.1430	595	0.0687	273	0.1852			868	0.1310	868	0.1690
21	0	0.6378	0	0.6238	0	0.5608	0	0.6238	16 & 17		0	0.6378	0	0.5608
22	0	0.1089	0	0.6238	0	0.1732	0	0.6238	17		0	0.1158	0	0.1636
23	2	0.1540	1	0.1818	2	0.2986	1	0.7080	24		3	0.1381	3	0.3380
24	73	0.1540	9	0.1818	73	0.2986	9	0.7080			82	0.1381	82	0.3380
25	290	0.1456	157	0.1432	290	0.1520	157	0.1489			447	0.1141	447	0.1888
26	0	0.6378	0	0.3859	0	0.5608	0	0.6238	16 & 17		0	0.6378	0	0.5608
27	0	0.6378	0	0.3859	0	0.5608	0	0.6238	16 & 17		0	0.1158	0	0.1636
28	0	0.5585	0	0.1818	0	0.8587	0	0.7080	29 & 24		0	0.5585	0	0.8587
29	8	0.5585	0	0.1818	8	0.8587	0	0.7080	24		8	0.5585	8	0.8587
30	16	0.2886	20	0.1407	16	0.3283	20	0.6095			36	0.0702	36	0.2232
Total num of Obs	4920		1221		4920		1221		Num of lack-data grids	6141		6141		Num of lack-data grids
	6141				6141					0.1386		0.1663		
Avg R <sup>2</sup>	0.1423				0.1537				15	0.1524				7

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

**Table 3.4 Options for linear regression solution (track model)**

Method		OLS	OLS	Robust "andrews"		Robust "bisquare"		Robust "cauchy"		Robust "fair"		Robust "huber"		Robust "logistic"		Robust "talwar"		Robust "welsch"		Subs for lack-data grids
Formula		3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	3.1	3.2	
Grid num	Num of Obs	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	
1	23	0.2091	0.1302	0.2014	0.1025	0.2016	0.1026	0.1960	0.0985	0.1390	0.0893	0.2059	0.0965	0.1916	0.0929	0.2091	0.1028	0.2006	0.1019	
2	34	0.1660	0.1107	0.1266	0.0798	0.1268	0.0798	0.1159	0.0747	0.0846	0.0757	0.1389	0.0739	0.1115	0.0745	0.1660	0.1107	0.1230	0.0771	
3	65	0.1807	0.3246	0.1608	0.0141	0.1608	0.0142	0.1579	0.0386	0.1479	0.2922	0.1529	0.1555	0.1572	0.1758	0.1807	0.0114	0.1594	0.0218	
4	14	0.5528	0.6680	0.5090	0.6330	0.5097	0.6332	0.4759	0.6257	0.3686	0.5921	0.5528	0.6337	0.4588	0.6246	0.5528	0.6680	0.5004	0.6298	
5	13	0.3775	0.7771	0.3263	0.7538	0.3269	0.7540	0.2964	0.7473	0.2182	0.7428	0.3775	0.7771	0.2834	0.7502	0.3775	0.7771	0.3175	0.7506	
6	5	0.2984	0.2869	0.1410	0.2390	0.1413	0.2392	0.1094	0.2296	0.1223	0.2164	0.1289	0.2400	0.1208	0.2286	0.1717	0.2869	0.1346	0.2347	11
7	65	0.0918	0.1231	0.0343	0.1492	0.0343	0.1492	0.0401	0.1347	0.0467	0.1290	0.0430	0.1477	0.0450	0.1359	0.0627	0.1455	0.0332	0.1438	
8	325	0.1313	0.1735	0.1461	0.0084	0.1455	0.0088	0.1497	0.0130	0.1525	0.0629	0.1509	0.0433	0.1528	0.0434	0.1908	0.0604	0.1431	0.0090	
9	337	0.1256	0.0862	0.1265	0.0738	0.1258	0.0749	0.1273	0.0674	0.1257	0.0692	0.1240	0.0688	0.1273	0.0703	0.1230	0.1806	0.1266	0.0812	
10	347	0.1259	0.0573	0.2543	0.0549	0.2542	0.0549	0.2258	0.0539	0.1946	0.0614	0.2226	0.0542	0.2120	0.0561	0.2993	0.0776	0.2473	0.0539	
11	24	0.2984	0.2869	0.1410	0.2390	0.1413	0.2392	0.1094	0.2296	0.1223	0.2164	0.1289	0.2400	0.1208	0.2286	0.1717	0.2869	0.1346	0.2347	
12	353	0.1616	0.2329	0.1438	0.0636	0.1442	0.0642	0.1511	0.0747	0.1520	0.1873	0.1587	0.1261	0.1533	0.1363	0.1460	0.0802	0.1457	0.0657	
13	665	0.1582	0.1212	0.1438	0.0645	0.1449	0.0643	0.1558	0.0684	0.1567	0.1002	0.1533	0.0802	0.1579	0.0855	0.1997	0.0828	0.1498	0.0638	
14	796	0.1441	0.1351	0.1270	0.0570	0.1280	0.0570	0.1421	0.0529	0.1484	0.0737	0.1389	0.0619	0.1451	0.0634	0.1655	0.0636	0.1339	0.0561	
15	678	0.1197	0.1626	0.1534	0.1958	0.1533	0.1961	0.1469	0.1983	0.1427	0.1914	0.1372	0.1890	0.1441	0.1901	0.1625	0.2117	0.1514	0.1984	
16	8	0.6378	0.5608	0.6081	0.5214	0.6086	0.5221	0.5871	0.4918	0.4685	0.3299	0.6378	0.5608	0.5758	0.4752	0.6378	0.5608	0.6026	0.5139	
17	159	0.1158	0.1636	0.1416	0.3003	0.1415	0.3003	0.1251	0.0614	0.1155	0.0976	0.1120	0.0763	0.1197	0.0786	0.1358	0.2962	0.1385	0.0545	
18	317	0.1233	0.2765	0.0948	0.1205	0.0952	0.1200	0.1142	0.1250	0.1208	0.3138	0.1214	0.1869	0.1175	0.2116	0.1151	0.1399	0.1026	0.1186	
19	469	0.1462	0.1682	0.1298	0.0241	0.1301	0.0243	0.1440	0.0326	0.1454	0.0807	0.1423	0.0599	0.1464	0.0623	0.1721	0.0306	0.1366	0.0240	
20	868	0.1310	0.1690	0.1065	0.0492	0.1072	0.0493	0.1199	0.0599	0.1243	0.1408	0.1221	0.0990	0.1223	0.1067	0.1228	0.0567	0.1133	0.0488	
21	0	0.6378	0.5608	0.6081	0.5214	0.6086	0.5221	0.5871	0.4918	0.4685	0.3299	0.6378	0.5608	0.5758	0.4752	0.6378	0.5608	0.6026	0.5139	16
22	0	0.1158	0.1636	0.1416	0.3003	0.1415	0.3003	0.1251	0.0614	0.1155	0.0976	0.1120	0.0763	0.1197	0.0786	0.1358	0.2962	0.1385	0.0545	17
23	3	0.1381	0.3380	0.1406	0.0407	0.1416	0.0407	0.1529	0.0450	0.1573	0.1494	0.1497	0.0720	0.1555	0.0780	0.1102	0.0416	0.1472	0.0407	24
24	82	0.1381	0.3380	0.1406	0.0407	0.1416	0.0407	0.1529	0.0450	0.1573	0.1494	0.1497	0.0720	0.1555	0.0780	0.1102	0.0416	0.1472	0.0407	
25	447	0.1141	0.1888	0.0921	0.2550	0.0920	0.2551	0.0986	0.2406	0.1028	0.1933	0.1044	0.2199	0.1002	0.2133	0.1134	0.2635	0.0946	0.2536	
26	0	0.6378	0.5608	0.6081	0.5214	0.6086	0.5221	0.5871	0.4918	0.4685	0.3299	0.6378	0.5608	0.5758	0.4752	0.6378	0.5608	0.6026	0.5139	16
27	0	0.1158	0.1636	0.1416	0.3003	0.1415	0.3003	0.1251	0.0614	0.1155	0.0976	0.1120	0.0763	0.1197	0.0786	0.1358	0.2962	0.1385	0.0545	17
28	0	0.5585	0.8587	0.4992	0.8289	0.5002	0.8295	0.4552	0.8027	0.3115	0.5763	0.5585	0.8587	0.4313	0.7862	0.5585	0.8587	0.4878	0.8226	29
29	8	0.5585	0.8587	0.4992	0.8289	0.5002	0.8295	0.4552	0.8027	0.3115	0.5763	0.5585	0.8587	0.4313	0.7862	0.5585	0.8587	0.4878	0.8226	
30	36	0.0702	0.2232	0.0772	0.1550	0.0772	0.1550	0.0798	0.1377	0.0781	0.1087	0.0838	0.1186	0.0803	0.1142	0.0702	0.1626	0.0783	0.1551	
Avg R <sup>2</sup>		<b>0.1386</b>	<b>0.1663</b>	0.1364	0.0986	0.1367	0.0987	0.1418	0.0942	0.1406	0.1348	0.1415	0.1119	0.1422	0.1155	<b>0.1620</b>	0.1177	0.1392	0.0925	

### 3.3.4.3 Compositions of available variables (Stage 3)

To reiterate, the ETM's mathematical expressions are provided as follows:

$$\Delta \ln c = a_1 + a_2 \psi_i + a_3 \lambda_i + a_4 \ln c_i + a_5 \theta_i + \varepsilon \quad (3.1)$$

$$\Delta \theta = b_1 + b_2 \psi_i + b_3 \lambda_i + b_4 c_i + b_5 \theta_i + b_6 \theta_{i-1} + \varepsilon \quad (3.2)$$

To reduce model complexity, introducing equations 3.1 and 3.2 in the linear forms is an acceptable idea. However, the approach to derive specific structures of those formulae is still obscure in all existing ETM-based studies. There are two questions that must be addressed:

- Which parameters are available to construct the expressions of changes in translation speed and heading over the current period?
- Using the above factors, which combinations of those values give the best performance?

The first topic is far beyond the scope of this research. Generally, the answer can only be achieved by exploring all parameters, which possibly contribute to the developments of TCs. This process requires not only certain knowledge of TC's behaviour, but also lots of trials and errors. Therefore, all parameters provided by the ETM are considered as useable values for the GSESM.

As described earlier in paragraph 3.3.3.1, the optimal compositions can be obtained by evaluating the adjusted  $R^2$  of all possible combinations of utilizable parameters. Table 3.5 and Table 3.6 show the analyses for equations to determine the changes in translation speed and heading, respectively, for the case study of Vietnam. In these tables, each term is represented by its coefficient. For instance, " $a_2$ " in Table 3.5 is the representative of " $a_2 \psi_i$ " term, and term " $b_2 \psi_i$ " is denoted by " $b_2$ " in Table 3.6. As a result, the "Original" scenario in Table 3.5 (i.e., equation 3.1) is a combination of terms  $a_1, a_2, a_3, a_4,$  and  $a_5$ . Similarly, the "Original" option in Table 3.6 (i.e., equation 3.2) is a composition of terms  $b_1, b_2, b_3, b_4, b_5,$  and  $b_6$ .

Moreover, because there is no use of two available parameters in equation 3.1 (i.e.,  $\theta_{i-1}$  and  $c_{i-1}$ ), two new terms associated with these unused elements are also considered, including " $a_6 \theta_{i-1}$ " and " $a_7 \ln c_{i-1}$ " (denoted by " $a_6$ " and " $a_7$ " in Table 3.5). Likewise, another term is introduced to the second formula, which is " $b_7 c_{i-1}$ " (denoted by " $b_7$ " in Table 3.6).

The optimal combinations obtained from those tables (i.e., largest average adjusted  $R^2$ , computed by equation 3.5, see principle 1 described in paragraph 3.3.3.1) establish the equations to determine the changes in translation speed and heading in the GSESM. For the specific case study of Vietnam, they are given as follows:

$$\Delta \ln c = a_1 + a_2 \psi_i + a_3 \lambda_i + a_4 \ln c_i + a_5 \theta_i + a_6 \theta_{i-1} + a_7 \ln c_{i-1} + \varepsilon \quad (3.6)$$

$$\Delta \theta = b_1 + b_2 \psi_i + b_3 \lambda_i + b_4 c_i + b_5 \theta_i + b_6 \theta_{i-1} + b_7 c_{i-1} + \varepsilon \quad (3.7)$$

**Table 3.5 Possible combinations of available parameters to determine the changes in translation speed**

Grid num	Num of Obs	Adjusted R <sup>2</sup>																Subs for lack-data grid
		Original	a1; a2	a1; a3	a1; a4	a1; a5	a1; a6	a1; a7	a1; a2; a3	a1; a4; a7	a1; a5; a6	a1; a2; a3; a4; a7	a1; a2; a3; a5; a6	a1; a4; a5; a6; a7	Original add a6	Original add a7	Original add a6; a7	
1	23	0.0333	-0.0476	-0.0120	0.1047	0.0453	-0.0475	0.0643	-0.0626	0.0599	0.1167	-0.0413	0.0621	0.0882	0.0484	-0.0033	-0.0095	
2	34	0.0510	-0.0307	-0.0276	0.0566	-0.0311	0.0567	0.0710	-0.0602	0.0421	0.3053	0.0586	0.2987	0.3150	0.3766	0.0372	0.3588	
3	65	0.1261	0.0530	-0.0144	0.0766	-0.0121	-0.0157	-0.0016	0.0382	0.0861	-0.0247	0.1324	0.0134	0.0741	0.1113	0.1321	0.1172	
4	14	0.3541	0.1363	0.0373	0.4310	-0.0801	0.0093	0.1432	-0.0502	0.4369	-0.0801	0.3870	-0.1970	0.4339	0.3748	0.3947	0.3672	
5	13	0.0663	-0.0474	-0.0641	0.1635	-0.0853	-0.0854	0.0079	-0.1396	0.0916	-0.1591	0.1913	-0.3749	-0.1007	-0.0104	0.0802	-0.0647	
6	5	-0.0027	-0.0104	0.0879	0.0299	0.1121	-0.0352	-0.0378	0.0375	0.2925	0.0795	0.1457	-0.0036	0.0731	-0.0584	0.1090	0.1982	11
7	65	0.0002	0.0011	0.0578	0.0210	-0.0040	0.0038	0.0518	0.0349	0.1975	-0.0133	0.1400	-0.0022	0.1653	-0.0190	0.1218	0.1036	
8	325	0.1806	0.0017	0.0010	0.1730	0.0120	-0.0007	0.0462	0.0021	0.2015	0.0326	0.1892	0.0266	0.2277	0.1794	0.1914	0.2123	
9	337	0.1124	-0.0013	0.0022	0.1247	0.0009	0.0062	0.0334	-0.0009	0.1945	0.0032	0.2406	0.0008	0.1477	0.1171	0.1733	0.2373	
10	347	0.2911	0.0024	0.0013	0.3046	0.1160	0.0803	0.1776	0.0013	0.3167	0.0902	0.3046	0.0903	0.3198	0.2826	0.3084	0.3015	
11	24	-0.0027	-0.0104	0.0879	0.0299	0.1121	-0.0352	-0.0378	0.0375	0.2925	0.0795	0.1457	-0.0036	0.0731	-0.0584	0.1090	0.1982	
12	353	0.1362	0.0054	0.0288	0.1069	0.0054	0.0131	0.0298	0.0259	0.1372	0.0116	0.2031	0.0201	0.1347	0.1334	0.1914	0.1913	
13	665	0.1948	0.0006	0.0014	0.1853	-0.0006	0.0075	0.0460	0.0023	0.2047	0.0071	0.2037	0.0095	0.2056	0.1884	0.1998	0.2003	
14	796	0.1613	0.0045	0.0030	0.1687	0.0023	0.0023	0.0600	0.0045	0.1690	0.0015	0.1775	0.0056	0.1757	0.1636	0.1790	0.1768	
15	678	0.1576	-0.0001	-0.0006	0.1487	0.0066	-0.0013	0.0597	-0.0007	0.1563	0.0090	0.1620	0.0083	0.1564	0.1594	0.1644	0.1607	
16	8	0.1549	-0.0527	-0.1417	-0.1319	-0.0778	0.0204	-0.1107	0.2412	-0.3316	-0.1729	-0.1530	0.0916	-0.0884	0.9425	-0.1165	0.8935	
17	159	0.1134	0.0267	-0.0043	0.1340	0.0431	-0.0027	-0.0006	0.0194	0.2360	-0.0081	0.2189	0.0306	0.2349	0.1528	0.2304	0.2151	
18	317	0.1038	-0.0022	-0.0024	0.1141	0.0004	-0.0005	0.0056	-0.0049	0.1824	-0.0025	0.1728	-0.0089	0.1780	0.0923	0.1458	0.1428	
19	469	0.1649	-0.0015	-0.0005	0.1723	0.0107	-0.0013	0.0475	-0.0027	0.1723	0.0157	0.1667	0.0153	0.1718	0.1533	0.1631	0.1671	
20	868	0.1188	-0.0004	0.0034	0.1014	0.0015	-0.0003	0.0235	0.0024	0.1021	0.0111	0.1151	0.0106	0.1052	0.1133	0.1187	0.1173	
21	0	0.1549	-0.0527	-0.1417	-0.1319	-0.0778	0.0204	-0.1107	0.2412	-0.3316	-0.1729	-0.1530	0.0916	-0.0884	0.9425	-0.1165	0.8935	16
22	0	0.1134	0.0267	-0.0043	0.1340	0.0431	-0.0027	-0.0006	0.0194	0.2360	-0.0081	0.2189	0.0306	0.2349	0.1528	0.2304	0.2151	17
23	3	0.0640	0.0017	0.0701	0.1398	0.0195	-0.0001	-0.0105	0.0569	0.1783	0.0069	0.1313	0.0563	0.1573	0.0499	0.1218	0.1013	24
24	82	0.0640	0.0017	0.0701	0.1398	0.0195	-0.0001	-0.0105	0.0569	0.1783	0.0069	0.1313	0.0563	0.1573	0.0499	0.1218	0.1013	
25	447	0.1054	0.0072	0.0062	0.0736	0.0246	0.0087	0.0069	0.0099	0.0824	0.0348	0.0870	0.0319	0.1070	0.1166	0.1176	0.1200	
26	0	0.1549	-0.0527	-0.1417	-0.1319	-0.0778	0.0204	-0.1107	0.2412	-0.3316	-0.1729	-0.1530	0.0916	-0.0884	0.9425	-0.1165	0.8935	16
27	0	0.1134	0.0267	-0.0043	0.1340	0.0431	-0.0027	-0.0006	0.0194	0.2360	-0.0081	0.2189	0.0306	0.2349	0.1528	0.2304	0.2151	17
28	0	-0.0302	-0.1487	-0.0700	0.0781	0.0463	-0.0060	-0.0222	-0.0778	0.2636	-0.0780	0.1109	-0.1182	0.1575	-0.2149	0.1409	-0.3227	29
29	8	-0.0302	-0.1487	-0.0700	0.0781	0.0463	-0.0060	-0.0222	-0.0778	0.2636	-0.0780	0.1109	-0.1182	0.1575	-0.2149	0.1409	-0.3227	
30	36	-0.0498	0.0152	-0.0288	0.0037	-0.0291	-0.0239	-0.0193	-0.0140	0.0581	-0.0519	0.0068	-0.0733	0.0007	-0.0842	-0.0221	-0.0572	
Avg adj R <sup>2</sup>		0.1482	0.0023	0.0042	0.1450	0.0125	0.0065	0.0440	0.0044	0.1663	0.0156	0.1718	0.0168	0.1673	0.1486	0.1684	0.1734	



**Table 3.6 Possible combinations of available parameters to define the changes in heading**

Grid num	Num of Obs	Adjusted R <sup>2</sup>														Substitute for lack-data grids
		Original	b <sub>1</sub> ; b <sub>2</sub>	b <sub>1</sub> ; b <sub>3</sub>	b <sub>1</sub> ; b <sub>4</sub>	b <sub>1</sub> ; b <sub>5</sub>	b <sub>1</sub> ; b <sub>6</sub>	b <sub>1</sub> ; b <sub>7</sub>	b <sub>1</sub> ; b <sub>2</sub> ; b <sub>3</sub>	b <sub>1</sub> ; b <sub>4</sub> ; b <sub>7</sub>	b <sub>1</sub> ; b <sub>5</sub> ; b <sub>6</sub>	b <sub>1</sub> ; b <sub>2</sub> ; b <sub>3</sub> ; b <sub>4</sub> ; b <sub>7</sub>	b <sub>1</sub> ; b <sub>2</sub> ; b <sub>3</sub> ; b <sub>5</sub> ; b <sub>6</sub>	b <sub>1</sub> ; b <sub>4</sub> ; b <sub>5</sub> ; b <sub>6</sub> ; b <sub>7</sub>	Original add b <sub>7</sub>	
1	23	-0.1257	-0.0476	-0.0464	-0.0060	0.0112	-0.0207	-0.0123	-0.0987	-0.0547	-0.0379	-0.1671	-0.1528	-0.0705	-0.1751	
2	34	-0.0481	0.0531	-0.0165	-0.0288	-0.0119	-0.0025	-0.0277	0.0371	-0.0609	-0.0347	-0.0159	-0.0247	-0.1051	-0.0868	
3	65	0.2674	-0.0141	-0.0120	-0.0157	0.1616	-0.0142	-0.0107	-0.0268	-0.0248	0.2873	-0.0530	0.2681	0.2927	0.2744	
4	14	0.4606	-0.0178	-0.0295	0.2041	0.1309	-0.0721	-0.0832	-0.0339	0.4034	0.0791	0.3460	0.3129	0.3549	0.4515	
5	13	0.6179	-0.0658	0.0434	0.1210	0.4995	0.2741	-0.0602	-0.0492	0.0342	0.4587	-0.1083	0.3965	0.6698	0.5670	
6	5	0.0888	-0.0443	0.0415	0.0541	0.0832	-0.0171	-0.0424	-0.0024	0.0413	0.0408	0.0933	-0.0236	0.0449	0.0354	11
7	65	0.0488	-0.0138	-0.0155	-0.0152	0.0967	0.0745	0.0941	-0.0295	0.2017	0.0883	0.1816	0.0633	0.2332	0.2277	
8	325	0.1606	-0.0026	-0.0017	-0.0005	0.1525	0.0916	-0.0029	-0.0039	-0.0008	0.1513	-0.0048	0.1467	0.1624	0.1580	
9	337	0.0724	-0.0030	-0.0009	0.0008	0.0687	0.0392	0.0019	-0.0038	-0.0010	0.0667	-0.0053	0.0615	0.0820	0.0765	
10	347	0.0435	-0.0027	-0.0002	0.0084	0.0210	0.0007	0.0108	-0.0029	0.0085	0.0276	0.0054	0.0247	0.0490	0.0461	
11	24	0.0888	-0.0443	0.0415	0.0541	0.0832	-0.0171	-0.0424	-0.0024	0.0413	0.0408	0.0933	-0.0236	0.0449	0.0354	
12	353	0.2218	0.0028	-0.0021	-0.0027	0.2056	0.0089	-0.0028	0.0002	-0.0055	0.2105	-0.0054	0.2130	0.2175	0.2201	
13	665	0.1145	-0.0014	-0.0014	-0.0011	0.1006	0.0208	-0.0012	-0.0028	-0.0026	0.1116	-0.0054	0.1114	0.1138	0.1142	
14	796	0.1296	-0.0012	-0.0002	-0.0011	0.1070	0.0201	-0.0006	-0.0014	-0.0017	0.1264	-0.0032	0.1243	0.1354	0.1332	
15	678	0.1564	-0.0007	-0.0013	0.0008	0.1064	0.0071	-0.0006	-0.0020	-0.0005	0.1413	-0.0027	0.1434	0.1538	0.1554	
16	8	-0.5372	-0.1572	-0.0994	-0.1553	-0.0098	0.3584	0.0270	-0.2895	0.1910	0.2333	0.5282	-0.2382	0.2045	-0.1997	
17	159	0.1362	-0.0004	-0.0014	-0.0064	0.1164	0.0090	-0.0047	-0.0002	-0.0091	0.1352	-0.0069	0.1305	0.1423	0.1424	
18	317	0.2649	0.0032	-0.0019	-0.0029	0.1078	-0.0023	0.0044	0.0005	0.0083	0.2380	0.0102	0.2529	0.2521	0.2746	
19	469	0.1592	-0.0020	0.0014	-0.0021	0.1578	0.0600	-0.0008	-0.0008	-0.0018	0.1582	-0.0025	0.1593	0.1572	0.1575	
20	868	0.1641	-0.0012	-0.0008	-0.0008	0.1256	0.0147	-0.0009	-0.0020	-0.0001	0.1659	-0.0022	0.1649	0.1658	0.1647	
21	0	-0.5372	-0.1572	-0.0994	-0.1553	-0.0098	0.3584	0.0270	-0.2895	0.1910	0.2333	0.5282	-0.2382	0.2045	-0.1997	16
22	0	0.1362	-0.0004	-0.0014	-0.0064	0.1164	0.0090	-0.0047	-0.0002	-0.0091	0.1352	-0.0069	0.1305	0.1423	0.1424	17
23	3	0.2944	0.0018	-0.0073	-0.0103	0.0914	0.0218	0.0165	-0.0059	0.0055	0.2672	0.0029	0.2894	0.3153	0.3274	24
24	82	0.2944	0.0018	-0.0073	-0.0103	0.0914	0.0218	0.0165	-0.0059	0.0055	0.2672	0.0029	0.2894	0.3153	0.3274	
25	447	0.1796	0.0011	-0.0017	-0.0017	0.1580	0.0230	-0.0017	-0.0004	-0.0038	0.1738	-0.0049	0.1785	0.1747	0.1777	
26	0	-0.5372	-0.1572	-0.0994	-0.1553	-0.0098	0.3584	0.0270	-0.2895	0.1910	0.2333	0.5282	-0.2382	0.2045	-0.1997	16
27	0	0.1362	-0.0004	-0.0014	-0.0064	0.1164	0.0090	-0.0047	-0.0002	-0.0091	0.1352	-0.0069	0.1305	0.1423	0.1424	17
28	0	0.5053	0.1418	-0.0946	-0.0444	0.1005	-0.1590	-0.1510	0.0931	-0.0307	-0.0770	-0.0196	0.4560	-0.0123	0.2097	29
29	8	0.5053	0.1418	-0.0946	-0.0444	0.1005	-0.1590	-0.1510	0.0931	-0.0307	-0.0770	-0.0196	0.4560	-0.0123	0.2097	
30	36	0.0937	-0.0274	0.0147	0.0272	0.0601	0.0041	0.0225	-0.0132	0.0083	0.0356	0.0406	0.0263	0.0357	0.0776	
Avg adj R <sup>2</sup>		<b>0.1498</b>	-0.0015	-0.0015	-0.0003	0.1174	0.0231	0.0003	-0.0030	0.0021	0.1425	0.0002	0.1424	0.1520	<b>0.1523</b>	

### 3.3.4.4 Computational grids (Stage 4)

As discussed in subsection 2.3.1, the hexagon is used in this study, as it is the best compromise solution for the shape of the computational grids. Thus, for the case study of Vietnam, the system of grids used in the ETM (i.e., 5° x 5° squares) is first compared with a system of hexagon grids with relatively equal resolution (i.e., hexagons that have 7° longest diagonals, result in the same total number of 30 grids over the whole domain).

**Table 3.7 Comparison between 5° squares and 7° hexagons (track model)**

Hexagons (7° longest diagonal)					Squares (5° side length)						
Formula	3.6		3.7		Substitute for lack-data grid	Formula	3.6		3.7		Substitute for lack-data grids
Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>		Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	
1	1	0.4904	1	0.6851	7	1	23	0.2658	23	0.1454	
2	27	0.6394	27	0.1728		2	34	0.4754	34	0.1108	
3	0	0.5868	0	0.6209	6	3	65	0.1999	65	0.3424	
4	30	0.2762	30	0.1726		4	14	0.6593	14	0.7046	
5	49	0.3470	49	0.4431		5	13	0.4677	13	0.7835	
6	15	0.5868	15	0.6209		6	5	0.4074	5	0.2870	11
7	11	0.4904	11	0.6851		7	65	0.1876	65	0.3001	
8	233	0.1691	233	0.1744		8	325	0.2269	325	0.1736	
9	285	0.2550	285	0.0743		9	337	0.2509	337	0.0930	
10	32	0.3181	32	0.4014		10	347	0.3136	347	0.0626	
11	611	0.1786	611	0.1266		11	24	0.4074	24	0.2870	
12	187	0.3490	187	0.2390		12	353	0.2051	353	0.2334	
13	1	0.4904	1	0.6851	7	13	665	0.2075	665	0.1222	
14	670	0.2217	670	0.1405		14	796	0.1830	796	0.1398	
15	796	0.1847	796	0.1262		15	678	0.1682	678	0.1629	
16	225	0.1218	225	0.2272		16	8	0.9848	8	0.8286	
17	916	0.1943	916	0.1834		17	159	0.2449	159	0.1750	
18	347	0.1668	347	0.1555		18	317	0.1591	317	0.2883	
19	1	0.1218	1	0.2272	16	19	469	0.1778	469	0.1683	
20	241	0.2104	241	0.1064		20	868	0.0534	868	0.1705	
21	863	0.1551	863	0.1802		21	0	0.9848	0	0.8286	16
22	0	0.1218	0	0.2272	16	22	0	0.2449	0	0.1750	17
23	106	0.2390	106	0.2925		23	3	0.1679	3	0.3773	24
24	271	0.0900	271	0.1180		24	82	0.1679	82	0.3773	
25	0	0.1218	0	0.2272	16	25	447	0.1319	447	0.1888	
26	0	0.2390	0	0.2925	23	26	0	0.9848	0	0.8286	16
27	199	0.1448	199	0.1818		27	0	0.2449	0	0.1750	17
28	0	0.2390	0	0.2925	23	28	0	0.8110	0	0.8871	29
29	2	0.2390	2	0.2925	23	29	8	0.8110	8	0.8871	
30	22	0.2958	22	0.2059		30	36	0.1240	36	0.2357	
Total num of Obs	6141		6141		Num of lack-data grids	Total num of Obs	6141		6141		Num of lack-data grids
Avg R <sup>2</sup>	0.1927		0.1628			Avg R <sup>2</sup>	0.1829		0.1721		
	<b>0.1778</b>				<b>9</b>		<b>0.1775</b>				<b>7</b>

As can be seen from Table 3.7, in terms of linear regression, a system of hexagons only gives a slightly better performance than the squares. However, when considered together with other important factors as provided in subsection 2.3.1, this confirms the lucid decision of utilizing hexagon grids in this research.

Furthermore, because the grid size also has a remarkable influence on track modelling, other dimensions are also examined. Table 3.8 and Table 3.9 present the evaluations for 6° and 8° hexagons, respectively. All results are then

summarized in Table 3.10. The table shows that, 7° hexagons are optimal computational grids for the case study of Vietnam. This is because the system not only gives the best performance (i.e., largest average  $R^2$ , computed by equation 3.5, see principle 1 described in paragraph 3.3.3.1), but also has a moderate number of lack-data grids. Figure 3.2 presents the optimal option along with model domain.

**Table 3.8 Analysis for 6° hexagons (track model)**

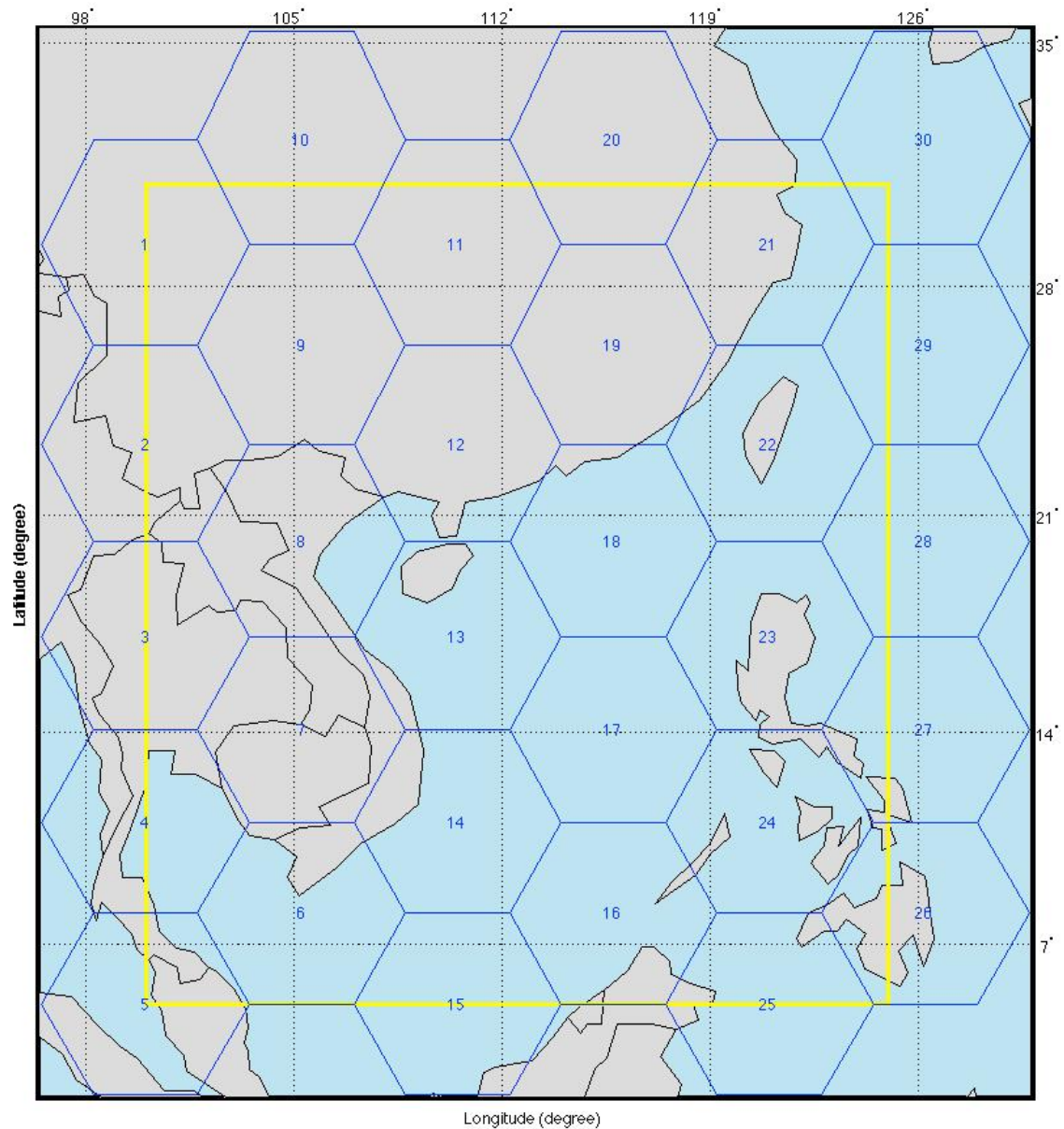
Formula	3.6		3.7		Substitute for lack-data grid
Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	
1	0	0.4945	0	0.6712	8
2	13	0.9449	13	0.4456	
3	0	0.6734	0	0.7653	7
4	0	0.6734	0	0.7653	7
5	23	0.3143	23	0.2069	
6	54	0.2067	54	0.3795	
7	10	0.6734	10	0.7653	
8	12	0.4945	12	0.6712	
9	82	0.1348	82	0.0571	
10	129	0.1388	129	0.1379	
11	15	0.4962	15	0.7752	
12	4	0.4945	4	0.6712	8
13	332	0.2494	332	0.1684	
14	338	0.2930	338	0.0975	
15	0	0.4945	0	0.6712	8
16	255	0.1574	255	0.3411	
17	548	0.2968	548	0.1163	
18	51	0.3225	51	0.1163	
19	105	0.2152	105	0.2600	
20	708	0.0959	708	0.0801	
21	648	0.1890	648	0.1746	
22	1	0.2672	1	0.6282	26
23	414	0.3382	414	0.1573	
24	640	0.1523	640	0.1446	
25	81	0.1768	81	0.3679	
26	15	0.2672	15	0.6282	
27	251	0.1070	251	0.2995	
28	806	0.1094	806	0.0995	
29	0	0.2672	0	0.6282	26
30	6	0.2672	6	0.6282	26
31	207	0.1961	207	0.2058	
32	80	0.0882	80	0.2332	
33	0	0.2672	0	0.6282	26
34	2	0.5612	2	0.3261	38
35	277	0.1453	277	0.1673	
36	0	0.2672	0	0.6282	26
37	0	0.2672	0	0.6282	26
38	19	0.5612	19	0.3261	
39	13	0.7327	13	0.1735	
40	0	0.2672	0	0.6282	26
41	0	0.5612	0	0.3261	38
42	2	0.7327	2	0.1735	39
Total num of Obs	6141		6141		Num of lack-data grids
Avg R <sup>2</sup>	0.1920		0.1628		
	0.1774				15

**Table 3.9 Analysis for 8° hexagons (track model)**

Formula	3.6		3.7		Substitute for lack-data grid
	Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	
1	5	0.9407	5	0.9719	6
2	35	0.3639	35	0.2940	
3	1	0.2207	1	0.0868	5
4	45	0.5250	45	0.2229	
5	85	0.2207	85	0.0868	
6	9	0.9407	9	0.9719	
7	408	0.2491	408	0.1615	
8	334	0.3075	334	0.0789	
9	178	0.1318	178	0.3287	
10	960	0.2166	960	0.1087	
11	4	0.9407	4	0.9719	6
12	1073	0.1923	1073	0.1308	
13	919	0.1530	919	0.1791	
14	217	0.2505	217	0.1938	
15	855	0.1609	855	0.1683	
16	0	0.2505	0	0.1938	14
17	55	0.1511	55	0.4241	
18	806	0.1243	806	0.1784	
19	0	0.1511	0	0.4241	17
20	103	0.1627	103	0.2197	
21	0	0.1511	0	0.4241	17
22	0	0.1511	0	0.4241	17
23	49	0.1856	49	0.2981	
Total num of Obs	6141		6141		Num of lack-data grids
Avg R <sup>2</sup>	0.1924		0.1621		
	0.1772				7

**Table 3.10 Summary of scenarios for computational grids (track model)**

Grid type	Grid size (degrees)	Average R <sup>2</sup>			Number of grids		
		Formula 3.6	Formula 3.7	Mean	Lack-data	Total	% of lack-data
Square	5	0.1829	0.1721	0.1775	7	30	23
Hexagon	6	0.1920	0.1628	0.1774	15	42	36
	<b>7</b>	<b>0.1927</b>	<b>0.1628</b>	<b>0.1778</b>	<b>9</b>	<b>30</b>	<b>30</b>
	8	0.1924	0.1621	0.1773	7	23	30



**Figure 3.2 Computational grids and domain for the case study of Vietnam (track model)**

#### 3.3.4.5 The complete GSESM's approach for the track modelling

Through a step-by-step calibration procedure (see flow chart in Figure 3.3), the comprehensive track modelling technique, which contains all optimal solutions defined in the above stages, is presented as follows:

- First, some DPs are removed from the observed BTD due to several model's limitations (see principle 3 described in paragraph 3.3.3.3).
- The entire research area (i.e., the TA) is then divided into various grids corresponding to the optimal option defined in stage 4 (provided in paragraph 3.3.4.4). In the case of Vietnam, they are 7° hexagons.

- For each grid, all historical DPs inside this cell are collected. According to the result of stage 1 (presented in paragraph 3.3.4.1), if the model performs better with a separation of westbound and eastbound DPs, the observed data will further be divided into two subsets based on their headings. In the case of Vietnam, stage 1 concludes that there is no need for such separation of DPs. Key parameters are then derived from the observed data of each grid.
- Findings of stage 3 (described in paragraph 3.3.4.3) are optimal mathematical expressions to estimate changes in translation speed and heading. These formulae are denoted by equations 3.6 and 3.7.
- The multiple linear regression analyses are carried out to determine all required coefficients, using the optimal linear regression solutions defined in stage 2 (provided in paragraph 3.3.4.2). In the case of Vietnam, these are "talwar" robust and OLS approaches for equations 3.6 and 3.7, respectively.
- The regressions are repeated for every cell in the model domain. For the grids with little or even no historical DPs, the coefficients are assigned as the corresponding coefficients of the substitutive grids. These substitutive cells are defined by guidelines provided in principle 2 (paragraph 3.3.3.2).
- Given the "known" conditions at time step  $i-1$  and  $i$ , the changes in translation speed and heading over the current period are computed by equations 3.6 and 3.7, respectively. Then the "unknown" situation at the next time step (i.e., time step  $i+1$ ) can be estimated from equations 3.3 and 3.4.
- Since the procedure is repeated from the initial points until TCs come to their lysis, the entire tracks of all synthetic TCs are recorded and constitute a full model's database.

Table 3.11 shows a comparison between the track modelling approach employed in the ETM (Vickery, Skerlj, and Twisdale 2000) and the method used in the model of this study (i.e., the GSESM) for the case study of Vietnam. As can be seen from the table, GSESM's technique not only clarifies all remaining ambiguousness of the ETM, but also shows a significant improvement over the original method. While the general mean of average  $R^2$  increases from 0.1480 to 0.1778 (i.e., more than 20% rise), the total number of lack-data grids decreased from 15 to 9 (i.e., 40% reduction).

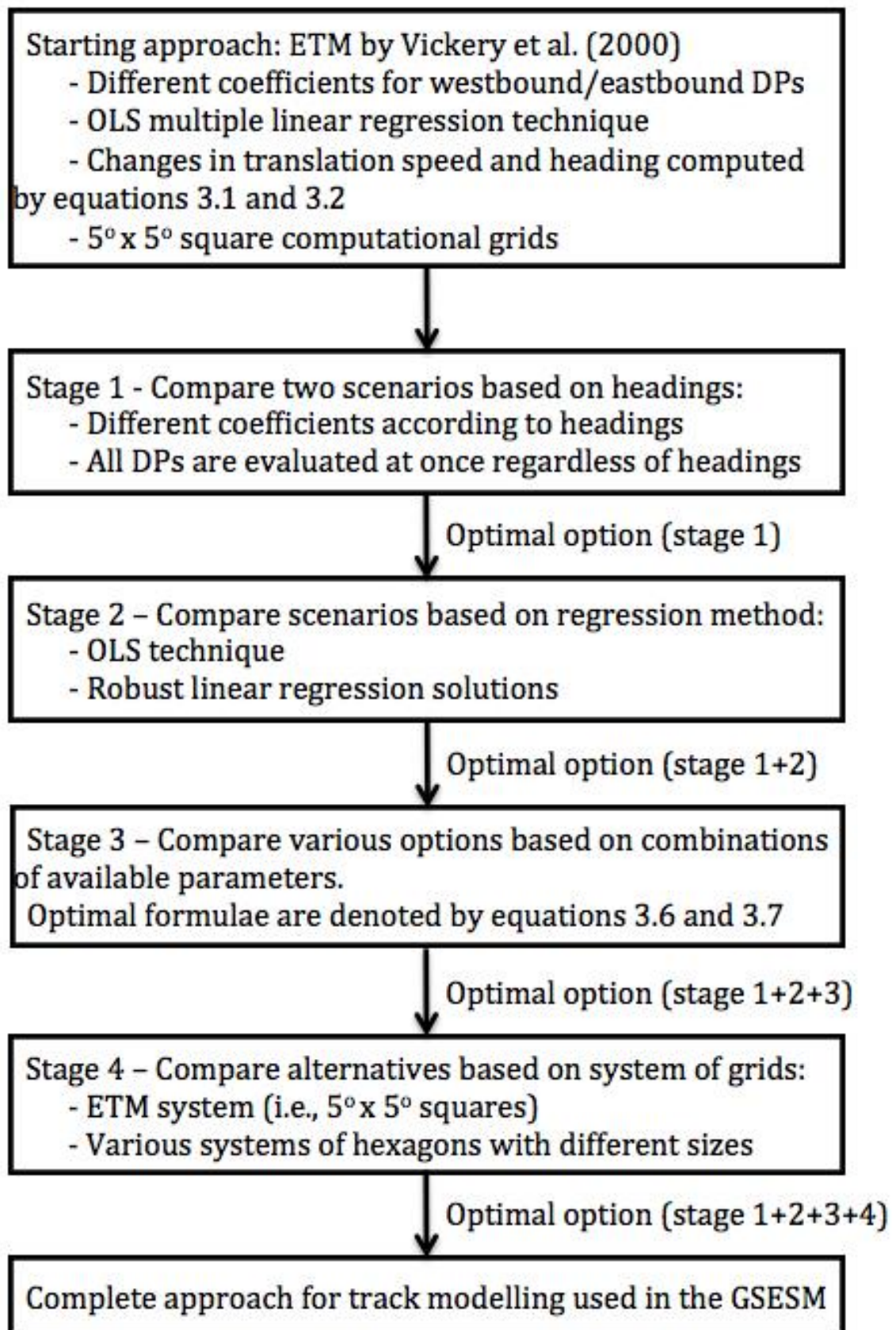


Figure 3.3 Optimization flow chart (track model)

**Table 3.11 Comparison between approaches utilized in the ETM and the GSESM for the case study of Vietnam (track model)**

Method	Empirical Track Modelling (ETM)								Generalized Stochastic Empirical Storm Modelling (GSESM)									
Grid type	Squares (5° side length)								Grid type	Hexagons (7° longest diagonal)								
Formula	3.1				3.2				Substitute for lack-data grids	Formula	3.6				3.7			
Regression	OLS				OLS					Regression	Robust "talwar"				OLS			
DP type	Westbound		Eastbound		Westbound		Eastbound			DP type	All DPs				All DPs			
Grid number	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>		Grid number	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>		
1	19	0.0383	4	0.4645	19	0.1544	4	0.3764	3	1	0.4904	1	0.6851	7				
2	31	0.1652	3	0.4645	31	0.1418	3	0.3764	3	2	0.6394	27	0.1728					
3	46	0.2383	19	0.4645	46	0.3977	19	0.3764		3	0	0.5868	0	0.6209	6			
4	14	0.5528	0	0.4645	14	0.6680	0	0.3764	3	4	30	0.2762	30	0.1726				
5	13	0.3775	0	0.6937	13	0.7771	0	0.6214	10	5	49	0.3470	49	0.4431				
6	5	0.2984	0	0.5763	5	0.2869	0	0.3679	11 & 12	6	15	0.5868	15	0.6209				
7	65	0.0918	0	0.5763	65	0.1231	0	0.3679	12	7	11	0.4904	11	0.6851				
8	291	0.1005	34	0.2166	291	0.1154	34	0.3698		8	233	0.1691	233	0.1744				
9	310	0.1821	27	0.1382	310	0.1345	27	0.1095		9	285	0.2550	285	0.0743				
10	336	0.0729	11	0.6937	336	0.0872	11	0.6214		10	32	0.3181	32	0.4014				
11	24	0.2984	0	0.5763	24	0.2869	0	0.3679	12	11	611	0.1786	611	0.1266				
12	340	0.1436	13	0.5763	340	0.1508	13	0.3679		12	187	0.3490	187	0.2390				
13	584	0.1557	81	0.2330	584	0.1202	81	0.1373		13	1	0.4904	1	0.6851	7			
14	584	0.1151	212	0.2009	584	0.1270	212	0.1449		14	670	0.2217	670	0.1405				
15	543	0.0850	135	0.1658	543	0.1686	135	0.1978		15	796	0.1847	796	0.1262				
16	8	0.6378	0	0.3859	8	0.5608	0	0.6238	17	16	225	0.1218	225	0.2272				
17	148	0.1089	11	0.3859	148	0.1732	11	0.6238		17	916	0.1943	916	0.1834				
18	271	0.1277	46	0.1332	271	0.2610	46	0.2312		18	347	0.1668	347	0.1555				
19	304	0.1837	165	0.1317	304	0.1128	165	0.1204		19	1	0.1218	1	0.2272	16			
20	595	0.1201	273	0.1430	595	0.0687	273	0.1852		20	241	0.2104	241	0.1064				
21	0	0.6378	0	0.6238	0	0.5608	0	0.6238	16 & 17	21	863	0.1551	863	0.1802				
22	0	0.1089	0	0.6238	0	0.1732	0	0.6238	17	22	0	0.1218	0	0.2272	16			
23	2	0.1540	1	0.1818	2	0.2986	1	0.7080	24	23	106	0.2390	106	0.2925				
24	73	0.1540	9	0.1818	73	0.2986	9	0.7080		24	271	0.0900	271	0.1180				
25	290	0.1456	157	0.1432	290	0.1520	157	0.1489		25	0	0.1218	0	0.2272	16			
26	0	0.6378	0	0.3859	0	0.5608	0	0.6238	16 & 17	26	0	0.2390	0	0.2925	23			
27	0	0.6378	0	0.3859	0	0.5608	0	0.6238	16 & 17	27	199	0.1448	199	0.1818				
28	0	0.5585	0	0.1818	0	0.8587	0	0.7080	29 & 24	28	0	0.2390	0	0.2925	23			
29	8	0.5585	0	0.1818	8	0.8587	0	0.7080		29	2	0.2390	2	0.2925	23			
30	16	0.2886	20	0.14074	16	0.3283	20	0.6095		30	22	0.2958	22	0.2059				
Total num of Obs	4920		1221		4920		1221		Num of lack-data grids	Total num of Obs	6141		6141		Num of lack-data grids			
	6141				6141						0.1927		0.1628					
Avg R <sup>2</sup>	0.1423				0.1537				15	Avg R <sup>2</sup>	0.1778				9			
	0.1480																	



### 3.4 Intensity simulation over water

Another part of an ETM-based study is the intensity simulation, which can be combined with the track modelling to provide a complete picture of TC's activity and to derive required statistics for any long-term analyses. Thus, although the developments of TC track and intensity are treated separately by using two distinct processes, most studies evaluate them simultaneously.

#### 3.4.1 Representative of intensity

Generally, TC's strength is measured as an important component in all available databases (Brettschneider 2006). At any given moment, the intensity is represented by the value of the maximum sustained surface wind speed ( $V_{smax}$ ) or by the minimum central pressure ( $p_c$ ), which can be obtained in the TC. Basically, a TC observation should consist of both  $V_{smax}$  and  $p_c$ . However, due to various difficulties, total numbers of measurements for these parameters are not equal. For instance, in the JMA's BTD, while observations of  $p_c$  have been available from 1951,  $V_{smax}$  have only been recorded since 1977.

Thus, it is wise to utilize only one factor to represent intensity. If necessary the other element can be derived from the representative by using a relationship. Between the two parameters, although the potential damage due to TCs is closely related to  $V_{smax}$  (Knaff and Zehr 2007),  $p_c$  is still chosen as the representative of a TC's strength. There are several reasons for this selection.

Firstly, observations of  $p_c$  are more accurate and reliable than the ones for  $V_{smax}$  (Atkinson and Holliday 1977; Knaff and Zehr 2007; Rosendal and Shaw 1982). Generally, there are two sources for directly obtaining intensity: from aircraft reconnaissance flights and from surface equipment. For aircraft measurements, while  $p_c$  can easily and trustfully be defined by releasing dropsondes into the TC's eyes (Rosendal and Shaw 1982),  $V_{smax}$  at the surface level is determined indirectly from wind speed at flight level. Estimates of  $V_{smax}$  in these cases, therefore, are often very subjective and contain a certain amount of uncertainty (Atkinson and Holliday 1977). In case of surface observations, data are collected by weather stations or ships. At coastal stations, the anemometers are usually broken or damaged under the influence of the intense winds of the strong TCs, whereas the barometers still get through (Atkinson and Holliday 1977). Moreover, the narrow strips that experience the TC's maximum winds (i.e., the eyewalls) also lead to doubts whether the vessels actually pass through these bands or not (Rosendal and Shaw 1982).

Secondly, in the BTD, while  $V_{smax}$  is rounded to the nearest 5 kt,  $p_c$  is only approximated, at most, by 1 hPa. When coupled with the smaller absolute values of  $V_{smax}$ , this results in higher accuracy of  $p_c$  records than  $V_{smax}$  measurements.

Thirdly,  $V_{smax}$  and  $p_c$  are not consistent. In actual situations, the observed  $V_{smax}$  varies for a given  $p_c$  (Vickery and Wadhwa 2008). This fact leads to the invention of the Holland parameter [ $B$  (Holland 1980)]. Furthermore, because the winds are, in fact, estimated by the pressure gradient (Landsea 2010),  $p_c$  is a distinctive feature that can distinguish the strength of one TC from another.

Finally, while a TC's eye is the area where  $p_c$  is found, it does not always correlate with the location where  $V_{smax}$  occurs (Brettschneider 2006). That is because quick movement (i.e., high translation speed) and outer elements may result in an eccentricity of maximum winds from the actual centre (Jarvinen 1984). A typical example is Hurricane Carla (1961), in which the TC's eye measured by radar was about 100 miles (i.e., 161 km) south of the centre of ring of maximum winds.

### 3.4.2 Relative intensity

#### 3.4.2.1 Original technique

In the ETM's approach, when a synthetic TC propagates over the ocean, instead of directly simulating central pressure, Vickery et al. (2000) modelled the evolution of Relative Intensity (RI). Introduced by Darling (1991), the concept of RI was an effort to incorporate more physical principles into the simulation, and therefore produce more realistic results. The fundamental idea of this technique is that, the efficiency of a TC is compared with a Carnot cycle heat engine (Vickery, Skerlj, and Twisdale 2000). Using this approach, RI is defined by comparing actual intensity to the theoretical Maximum Potential Intensity (MPI) that a TC can possibly attain at the same location. The determination of MPI, therefore relies on local conditions and employs the method presented by Emanuel (1988). As indicated by Vickery et al. (2009), the key benefit of involving MPI into the model is that the simulated  $p_c$  derived from RI is automatically restricted to physical constraints (i.e., the realistic limits). Thus, an artificial truncation for  $p_c$  of synthetic TCs, which certainly introduces arbitrariness, is not needed.

The procedure for estimating RI is provided by Darling (1991) as follows:

Calculate the values:

$$\epsilon = \frac{T_s - T_0}{T_s}$$

$$L_v = 2.5 \cdot 10^6 - 2320 (T_s - 273)$$

$$e_s = 6.112 \exp \left[ \frac{17.67 (T_s - 273)}{T_s - 29.5} \right]$$

Where:  $\epsilon$  - efficiency of the TC as a heat engine;  $e_s$  - saturation vapour pressure (hPa);  $L_v$  - latent heat of vaporization ( $\text{J kg}^{-1}$ );  $T_s$  and  $T_0$  - air temperature (degrees K) of the sea surface and of the top of the troposphere at TC's centre, respectively.

Compute the quantities:

$$p_{da} = p_n - (RH e_s)$$

$$A = \frac{\epsilon L_v e_s}{(1 - \epsilon) R_v T_s p_{da}}$$

$$B = RH \left[ 1 + \frac{e_s \ln(RH)}{p_{da} A} \right]$$

Where:  $p_{da}$  - surface value of the partial pressure of ambient dry air (hPa);  $p_n$  - environmental pressure, taken to be 1013 (hPa);  $RH$  - relative humidity of ambient air, taken as 0.75;  $R_v$  - gas constant of water vapour, taken to be 461.5 (J kg<sup>-1</sup> K<sup>-1</sup>).

Then define the value of  $x$  in a nonlinear formula by using iteration:

$$\ln(x) = -A \left( \frac{1}{x} - B \right)$$

Other values are obtained from:

$$p_{dc} = x p_{da}$$

$$p_d = p_c - e_s$$

$$I = \frac{p_{da} - p_d}{p_{da} - p_{dc}}$$

Where:  $I$  - relative intensity;  $p_{dc}$  - minimum sustainable surface value of central pressure of dry air (hPa);  $p_d$  - partial pressure for dry air (hPa);  $p_c$  - central pressure (hPa)

Thus, the required parameters for these equations are:

- The temperature at the top of the stratosphere ( $T_0$ ), taken as seasonal mean interpolated from the data provided by Newell (1973), or using a fixed number of 203°K (Emanuel 1988).
- The temperature of sea surface ( $T_s$ ), taken to be the mean seasonal value (Darling 1991), or the mean monthly value (Vickery, Skerlj, and Twisdale 2000) at the TC's centre.

Given these parameters, any values of  $p_c$  can be expressed in terms of RI, and vice versa, whenever the TC's centre is over water (i.e.,  $T_s$  is definable).

#### 3.4.2.2 Approach improvements

Although the above technique is repeated without any changes in many other studies [e.g., (Lee and Rosowsky 2007; Vickery, Skerlj, and Twisdale 2000; Vickery, Wadhwa, Twisdale, et al. 2009; Wang and Rosowsky 2012)], it still contains some drawbacks, which can be overcome by using modern techniques and data. Two major disadvantages are:

- The use of obsolete expressions to estimate latent heat of vaporization ( $L_v$ ) and saturation vapour pressure ( $e_s$ ).
- Assumptions of the values of various important parameters, including  $T_0$  [taken as 203°K (Emanuel 1988)],  $p_n$  [assumed of 1013 hPa (Darling 1991)], and  $RH$  [taken to be 0.75 (Vickery, Skerlj, and Twisdale 2000)].

Thus, the GSESM's approach can improve the original method (Darling 1991) by utilizing state-of-the-art equations and measurements.

Firstly, the empirical, typical Magnus-type formulae to compute  $L_v$  and  $e_s$  are replaced by ones derived from the theoretical relationship (i.e., Clausius-Clapeyron equations). The significance of this method is that, it provides a combination of both theoretical consistency and accuracy, while still maintaining the simplification of the formulae. The detailed methodology given by Koutsoyiannis

(2012), produces negligible errors when compared with outcomes from the laboratory's experiments. The resulting expressions are presented as follows:

$$L_v = 3.139 \cdot 10^6 - 2336 T_s$$

$$e_s = 6.11657 \exp \left[ 24.921 \left( 1 - \frac{273.16}{T_s} \right) \right] \left( \frac{273.16}{T_s} \right)^{5.06}$$

Secondly, the actual values of  $T_o$ ,  $p_n$ , and  $RH$  associated with real TC conditions are employed instead of the fixed and supposed quantities used in other researches. The data are derived from the Twentieth Century Reanalysis Project (20CR (Compo et al. 2011)). This data source gives observations in  $2^\circ \times 2^\circ$  global square grids and is updated regularly. For the uses in the GSESM, three types of data are acquired from the 20CR, including:

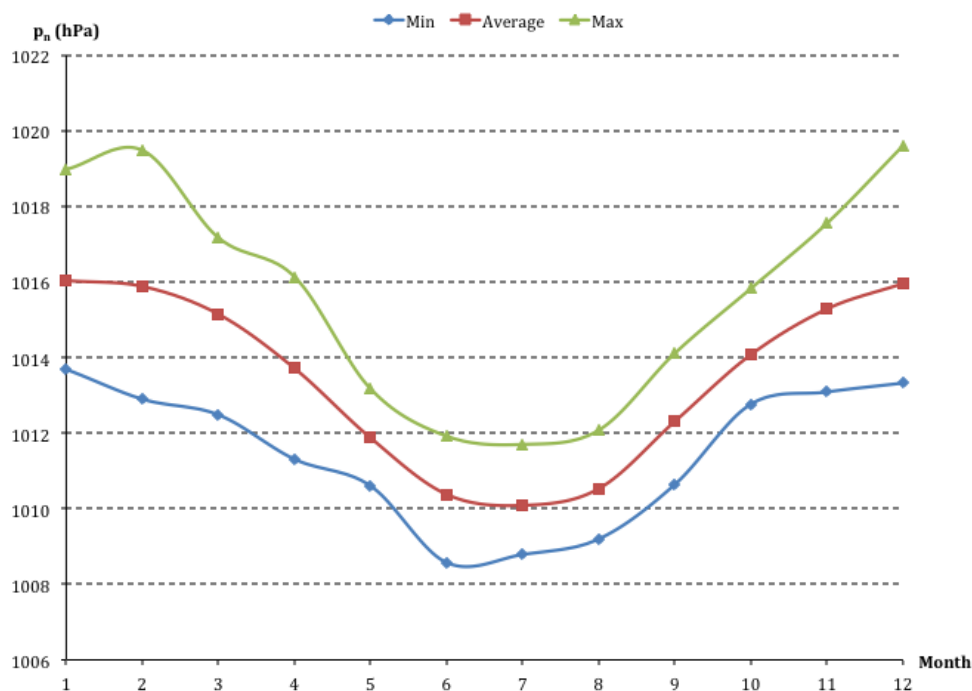
- The monthly mean values of temperature at the top of the stratosphere ( $T_o$ ), which are given for each node of the  $2^\circ \times 2^\circ$  squares. To compute  $T_o$  at TC's centres for uses in the above formulae, the linear two-dimensional interpolations are carried out from values at the nodes of squares.
- The monthly mean values of relative humidity ( $RH$ ), which are provided for each grid of the  $2^\circ \times 2^\circ$  squares (i.e., using one common value for all DPs in the same cell). The appearance of  $RH$  in the above formulae is represented for a characteristic of the ambient air (i.e., the surrounding, normal condition that contrasts with the extreme, abnormal situation, occurs in the vicinity of a TC). Thus, an average of all observed values at every single site, over the entire WNP is a truly acceptable representative for the physical environmental condition. Using this idea, the measurements of relative humidity from the 20CR are first obtained for every grid of the  $2^\circ \times 2^\circ$  squares. Then the site-specific data are averaged over the whole WNP to derive  $RH$  for use in the equations. Note that, there is no available surface relative humidity data, therefore "near surface" observations are utilized. According to Compo et al. (2011), the "near surface" elevation is the "sigma level 0.995", which corresponds to an altitude that has 99.5% of the surface pressure (i.e., equivalent to approximately 42 m above ground level for the standard atmospheric condition).
- The monthly mean values of sea level pressure, which are provided for each grid of the  $2^\circ \times 2^\circ$  squares and are utilized to define the environmental pressure ( $p_n$ ). Historically, although  $p_n$  is an important parameter since a wind field is actually driven by a central pressure deficit  $\Delta p$  (i.e., the difference between  $p_n$  and central pressure  $p_c$ ), the predetermined values are employed in all existing TC climatology research. This is because while in the definition of  $p_n$ , it is clearly stated that  $p_n$  is the pressure corresponding to the outermost closed isobar or gale force winds (Holland 1997), the applications of the concept are problematic and complicated. Various values with large differences are utilized in the same region [e.g., range from 1017 hPa (Chu and Wang 1998) to 1010 hPa (Atkinson and Holliday 1977; Knaff and Zehr 2007) for the WNP; 1016 hPa (Atkinson and Holliday 1977), 1015 hPa (Holland 2008; Holland et al. 2010), or 1013 hPa (Darling 1991; Wang and Rosowsky 2012) for the Atlantic basin].

Furthermore,  $p_n$ , in fact, not only depends on geographic position but also changes over time (Figure 3.4). Because there are also seasonal variations in both TC's strength (Figure 3.5) and occurrence rate (Figure 3.6), the fixed value of  $p_n$  should be replaced by the variable one, if it is readily obtainable (Holland 2008). As shown in Figure 3.6, the most active period of TC's activity is from July to September, which corresponds to the relatively low  $p_n$ . If an average is used instead of the actual  $p_n$  during this period,  $\Delta p$  will be larger and TCs are unreasonably intensified. The process for defining  $p_n$  is similar to the one applied to  $RH$ . First, the observations of sea level pressure from the 20CR are acquired for every cell of the  $2^\circ \times 2^\circ$  squares. Then site-specific data are averaged over the entire WNP to derive  $p_n$  for use in the formulae.

Finally, the most recent record of Sea Surface Temperature (SST, denoted by  $T_s$ ) is utilized. The so-called Extended Reconstructed SST V3b database (NOAA 2014a), provides monthly means of SST for each grid of the  $2^\circ \times 2^\circ$  squares. The  $T_s$  used in the equations is taken to be the value of the cell that contains the TC's centre.

One point to keep in mind is that, the method to define the above parameters is only applicable for historical BTD. That means for each observed DP, location of TC's centre and time (i.e., year and month) is given to search for the completely matched values of  $T_0$ ,  $RH$ ,  $p_n$ , and  $T_s$ .

In case of synthetic TCs, because there is no matching year, the observed sources of parameters (i.e., monthly means) are first averaged for all years in the historical record, to produce the monthly long-term means (e.g., January long-term data contains average values of all January observations in all years). Then, the position of a simulated TC's centre and its time of occurrence (i.e., month) are used to estimate the corresponding quantities of the required parameters.



**Figure 3.4 Seasonal variation in the environmental pressure**

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

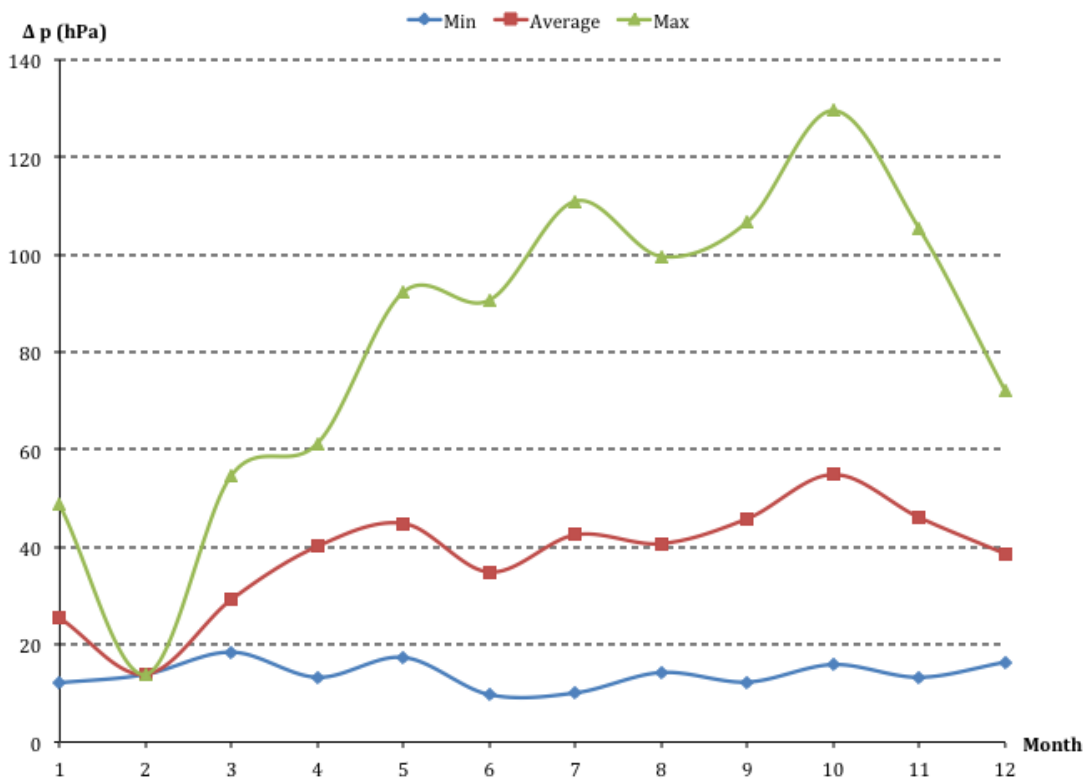


Figure 3.5 Seasonal variation in the central pressure deficit

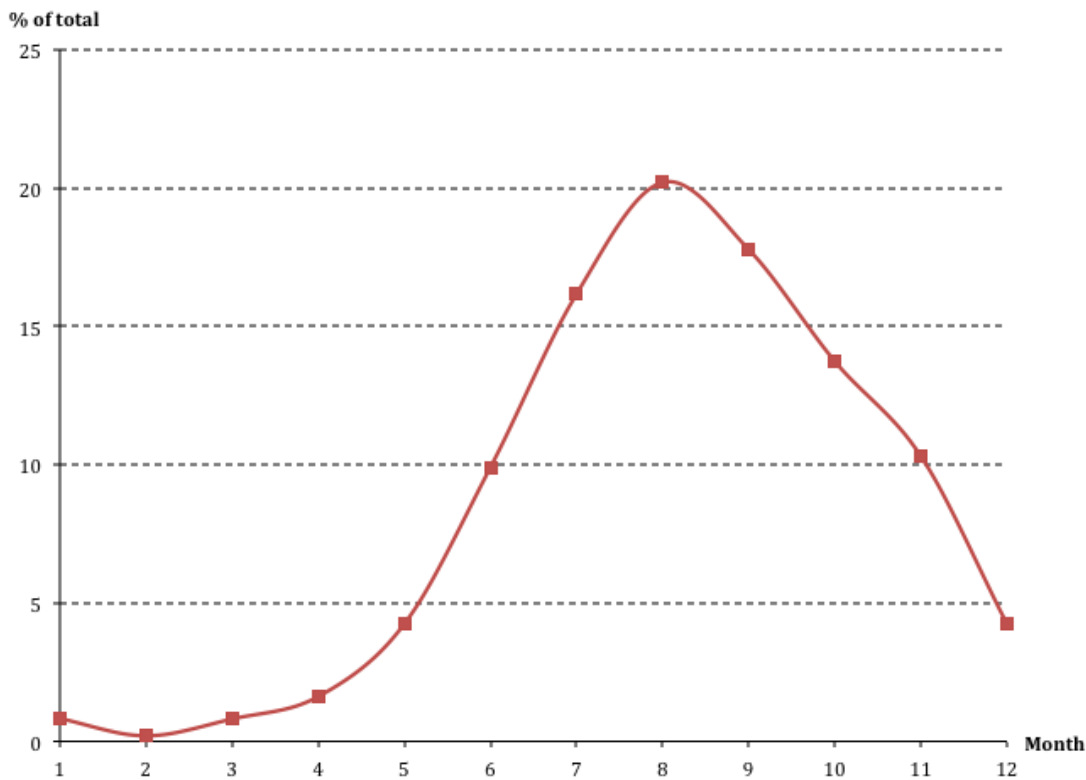


Figure 3.6 Seasonal variation in the TC occurrence

### 3.4.2.3 Significances of the modifications

Besides the clarity in identifying the latest data sources and using formulae with a high accuracy, the key advantage of the GSESM's technique is that for the first time, the realistic values of various critical parameters associated with real TC conditions are employed. Although the importance of utilizing actual data instead of a predetermined one, especially in case of the environmental pressure, has been stated in many studies for years [e.g., (Holland 2008; Vickery 2005)], this study is the first attempt to introduce this practical idea to a TC model.

### 3.4.3 Modelling intensity over the sea

#### 3.4.3.1 Original method

Similar to the approach for simulating tracks, in the ETM (Vickery, Skerlj, and Twisdale 2000), the "unknown" RI at any time step is computed as a linear function of that parameter at the last three steps and an external element (i.e., the SST) as follows:

$$\ln I_{i+1} = c_1 + c_2 \ln I_i + c_3 \ln I_{i-1} + c_4 \ln I_{i-2} + c_5 T_{si} + c_6 (T_{si+1} - T_{si}) + \varepsilon \quad (3.8)$$

Where:  $I$  - relative intensity;  $T_s$  - sea surface temperature;  $c_1, c_2$ , etc. - coefficients, depended on condition at current location.

Vickery et al. (2000) expressed that, although SST is not the only factor affecting RI, and therefore such sole representative of physical processes in the development of intensity is clearly a simplification, it still has some easily recognizable benefits. First, the inclusion of SST partly reduces the uncertainty in intensity simulation, which can occur when excluding this factor and to exclusively use the intensity as both dependent and explanatory variables in the regression (Vickery, Skerlj, and Twisdale 2000). Secondly, SST is a popular parameter, which is readily available in many global archives, and therefore its utilization enables a more easily applicable technique.

The identical method was utilized in numerous researches, without any changes [e.g., (Lee and Rosowsky 2007; Wang and Rosowsky 2012)], while others involved certain modifications. For instance, Vickery et al. (2009) excluded the SST and introduced another element to the model [i.e., the scaled vertical wind shear defined by DeMaria and Kaplan (1999)]. However, because this parameter is unavailable in many regions, a general approach including such a factor, is not feasible. Furthermore, since no direct comparison with the ETM's method is given, there is no evidence of improvements due to this change. Thus, the identical form of equation 3.8 is employed in this study for modelling intensity.

#### 3.4.3.2 Clarifications and possible improvements

Basically, most principles of clarification and stages of improvements that have been applied to track simulation are unchanged in the intensity modelling. The only exception is the combination of available variables. Because all required variables must be present in the mathematical expression, there is no argument about the definite form of equation 3.8. Thus, the detailed descriptions of the approach, which have already been given in section 3.3, are not repeated here.

Figure 3.7 shows a step-by-step process to improve the ETM's technique (Vickery, Skerlj, and Twisdale 2000) and to form a comprehensive methodology for modelling intensity, which contains all optimal solutions in the GSESM.

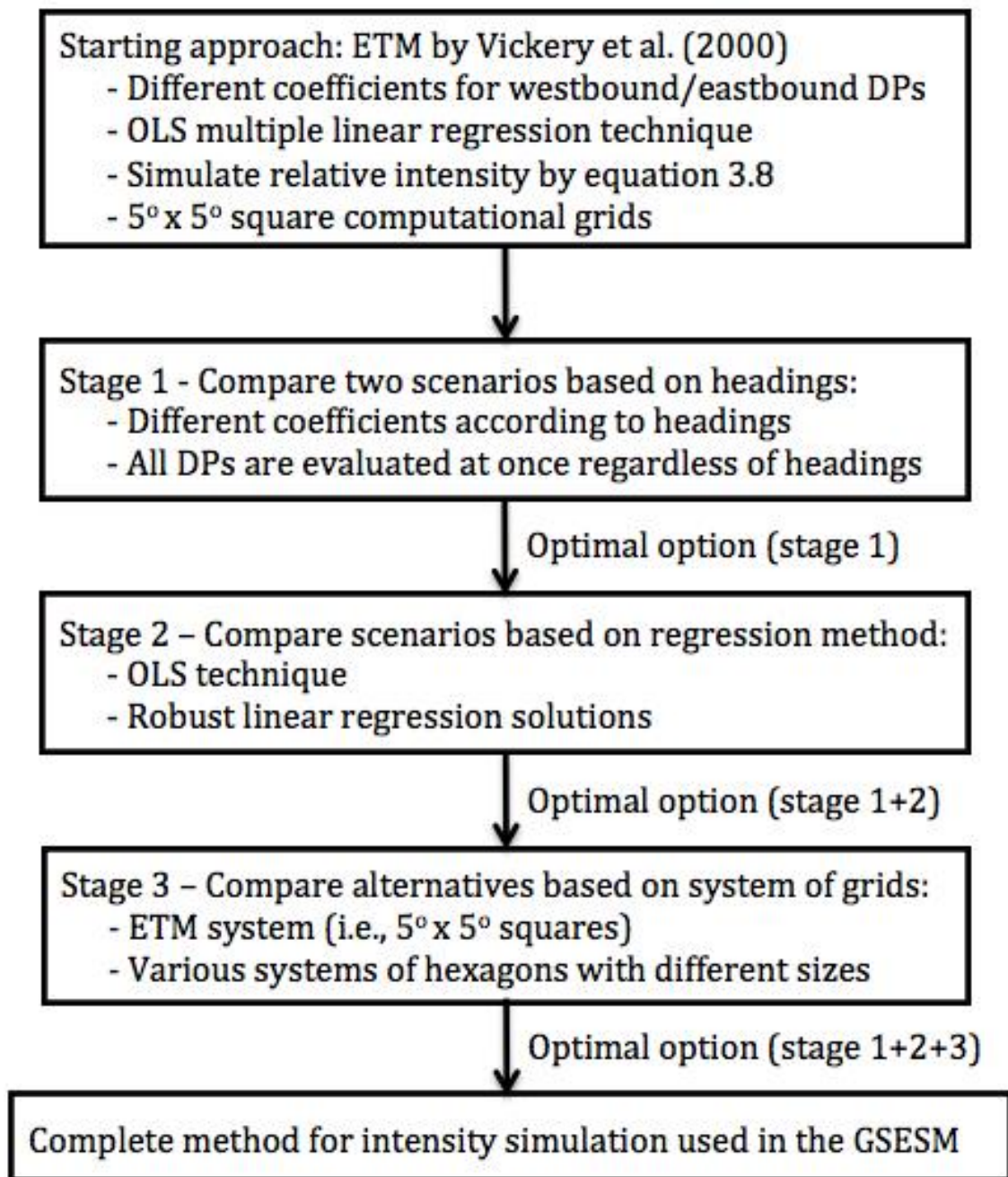
#### 3.4.3.3 Method background and clarifications

Firstly, in preparation for further analyses, various DPs are removed from the JMA's BTD that have been recorded since 1977 in the WNP (see subsection 2.1.1). Those data include all DPs outside the TA or that have a maximum sustained surface wind speed ( $V_{smax}$ ) smaller than 30 kt (see subsection 2.2.2), as well as the ones that violate several model validities (see paragraph 3.3.3.3). For the case study of Vietnam, Table 3.12 shows that among 17,935 DPs of the BTD in the WNP, there are 5,459 DPs that are usable for the intensity analyses over the ocean.

Because there are 731 DPs inside the TA that are excluded from the calculations due to model restrictions, Table 3.13 presents an examination to determine whether this elimination has any considerable impact on the model or not. As can be seen from the table in which DPs are categorized by different types according to the JMA, a large number of the 731 DPs (i.e., 516 DPs, which account for 70.59% of the total) have the strengths of a tropical depression or weaker (i.e.,  $V_{smax}$  smaller than 30 kt, denoted by types 2 and 6 in the table). Thus, they have no effect on the result. The model loses 215 DPs (denoted by types 3, 4, and 5 in the table), and only 96 of them (i.e., types 4 and 5) are significant ones, which are small (less than 2%) in comparison with the total number of DPs that are usable for the intensity model (i.e., 5,459 DPs).

Another point to note is that, since the approach is only applicable for oceanic DPs, any grids containing no DP and entirely located in the mainland, are not considered as lack-data cells, and therefore do not require substitutive grids. That is, because such grids are not taken into account in both the historical analysis (to find linear regression coefficients) and the intensity simulation.





**Figure 3.7 Optimization flow chart (intensity model)**

#### 3.4.3.4 Approach's improvements

As provided in Figure 3.7, the foremost optimization is related to soundness of the data separation due to headings. For the case study of Vietnam, Table 3.14 indicates that if all DPs (regardless of eastbound or westbound) are evaluated at once instead of dividing them by headings, benefits will be gained (i.e., the average  $R^2$ , computed by equation 3.5, rises from 0.9350 to 0.9358 and number of lack-data grids decreases from 6 to 2).

The second stage tests the abilities of various linear regression solutions for equation 3.8, as shown in Table 3.15. The optimal option (i.e., the "talwar" robust technique) provides a small improvement with a slight increase in the average  $R^2$ , when compared with the conventional OLS method.

Finally, to obtain the optimal computational grids, the ones used in the ETM (i.e.,  $5^\circ \times 5^\circ$  squares) are first compared with a system of hexagons that has a relatively equal resolution (i.e., hexagons that have  $7^\circ$  longest diagonals, result in the same total number of 30 grids over the whole domain). With the better performance in the linear model (Table 3.16) and several prominent features (see subsection 2.3.1), hexagons are clearly a preferable choice in this study. Moreover, other cell sizes are also evaluated in Table 3.17 ( $6^\circ$  hexagons) and Table 3.18 ( $8^\circ$  hexagons). All findings are summarized in Table 3.19, which indicates that  $7^\circ$  hexagons (Figure 3.8) are the optimal computational grids for the case study of Vietnam.

Note that, the coefficients of a lack-data grid (i.e., a grid containing less than 6 DPs) are assigned as the corresponding coefficients of the substitutive cell. Therefore, in these tables,  $R^2$  of a lack-data grid is taken equal to  $R^2$  of the corresponding substitutive cell.

**Table 3.12 Analysis of historical data (intensity model)**

Number of DPs												
WNP BTD	Removed (No. 1)		Inland	Over water	Removed (model limitations)			Leftover	Outside the TA	Inside the TA		
					No. 2	No. 3	Total			Total		
17935	138		1723	16074	0	1542	1655		14419	7354	7065	
	Inland	Over water					Outside the TA	Inside the TA			Tropical storm or stronger (wind $\geq$ 30 kt)	Tropical depression
	25	113					924	731			5459	1606

**Table 3.13 Exclusion of data points in the TA (intensity model)**

DP type	Number of DPs	Percentage
1	0	0.00
2	510	69.77
3	119	16.28
4	66	9.03
5	30	4.10
6	6	0.82
7	0	0.00
8	0	0.00
9	0	0.00
Total	731	100

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

**Table 3.14 Scenarios based on headings (intensity model)**

Scenario	Westbound/Eastbound DPs					All DPs		
DP type	Westbound		Eastbound		Substitute for lack-data grid	All DPs		Substitute for lack-data grid
Grid num	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>		Num of Obs	R <sup>2</sup>	
1	19	0.8737	4	0.8881	3	23	0.8698	
2	30	0.8753	3	0.8881	3	33	0.8685	
3	46	0.9003	19	0.8881		65	0.9112	
4	14	0.9065	0	0.9228	9	14	0.9065	
5	13	0.9704	0	0.9024	10	13	0.9704	
6	5	0.8863	0	0.9672	7 & 12	5	0.8863	7
7	13	0.8863	0	0.9672	12	13	0.8863	
8	287	0.9474	34	0.9284		321	0.9470	
9	310	0.9512	27	0.9228		337	0.9489	
10	336	0.9424	11	0.9024		347	0.9394	
11	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
12	171	0.8376	13	0.9672		184	0.8415	
13	584	0.9634	81	0.9615		665	0.9628	
14	584	0.9576	212	0.9522		796	0.9554	
15	442	0.9198	115	0.9237		557	0.9409	
16	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
17	60	0.9132	9	0.8954		69	0.9104	
18	233	0.8905	37	0.8839		270	0.8853	
19	261	0.8794	157	0.9467		418	0.9009	
20	595	0.9593	273	0.9726		868	0.9636	
21	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
22	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
23	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
24	10	0.8745	4	0.9561	25	14	0.8694	
25	262	0.8711	156	0.9561		418	0.8970	
26	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
27	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
28	0	0.0000	0	0.0000	Inland	0	0.0000	Inland
29	0	0.0000	0	0.0000	Inland	0	0.9640	30
30	12	0.9531	17	0.9847		29	0.9640	
Total num of Obs	4287		1172		Num of lack-data grids	5459		Num of lack-data grids
	5459							
Avg R <sup>2</sup>	0.9350				6	0.9358		2

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

**Table 3.15 Options for linear regression solution (intensity model)**

Method		OLS	Robust "andrews"	Robust "bisquare"	Robust "cauchy"	Robust "fair"	Robust "huber"	Robust "logistic"	Robust "talwar"	Robust "welsch"	Subs for lack-data grid
Grid num	Num of Obs	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup>	
1	23	0.8698	0.8563	0.8565	0.8470	0.8050	0.8630	0.8419	0.8698	0.8539	
2	33	0.8685	0.8517	0.8517	0.8484	0.8446	0.8471	0.8481	0.8685	0.8502	
3	65	0.9112	0.9107	0.9107	0.9099	0.9026	0.9134	0.9093	0.9200	0.9107	
4	14	0.9065	0.8960	0.8961	0.8897	0.8508	0.9065	0.8866	0.9065	0.8942	
5	13	0.9704	0.9707	0.9706	0.9718	0.9630	0.9723	0.9718	0.9704	0.9712	
6	5	0.8863	0.8649	0.8652	0.8536	0.7871	0.8863	0.8491	0.8863	0.8614	7
7	13	0.8863	0.8649	0.8652	0.8536	0.7871	0.8863	0.8491	0.8863	0.8614	
8	321	0.9470	0.9583	0.9584	0.9571	0.9537	0.9582	0.9563	0.9624	0.9581	
9	337	0.9489	0.9605	0.9604	0.9611	0.9620	0.9585	0.9611	0.9662	0.9607	
10	347	0.9394	0.9541	0.9541	0.9548	0.9544	0.9541	0.9545	0.9611	0.9547	
11	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
12	184	0.8415	0.9405	0.9401	0.9390	0.9332	0.9376	0.9363	0.9460	0.9401	
13	665	0.9628	0.9665	0.9665	0.9670	0.9675	0.9665	0.9671	0.9694	0.9667	
14	796	0.9554	0.9614	0.9615	0.9617	0.9617	0.9607	0.9616	0.9659	0.9617	
15	557	0.9409	0.9529	0.9529	0.9496	0.9451	0.9485	0.9476	0.9393	0.9526	
16	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
17	69	0.9104	0.9093	0.9093	0.9112	0.9126	0.9117	0.9119	0.9190	0.9102	
18	270	0.8853	0.9026	0.9024	0.9021	0.8971	0.9047	0.9011	0.9107	0.9027	
19	418	0.9009	0.9483	0.9485	0.9476	0.9427	0.9452	0.9453	0.9519	0.9487	
20	868	0.9636	0.9764	0.9763	0.9762	0.9753	0.9745	0.9758	0.9810	0.9764	
21	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
22	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
23	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
24	14	0.8694	0.8523	0.8526	0.8407	0.7769	0.8694	0.8347	0.8694	0.8492	
25	418	0.8970	0.9657	0.9654	0.9641	0.9588	0.9627	0.9619	0.9681	0.9653	
26	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
27	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
28	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Inland
29	0	0.9640	0.9616	0.9617	0.9597	0.9479	0.9651	0.9584	0.9640	0.9612	30
30	29	0.9640	0.9616	0.9617	0.9597	0.9479	0.9651	0.9584	0.9640	0.9612	
Avg R <sup>2</sup>		<b>0.9358</b>	0.9554	0.9554	0.9548	0.9520	0.9542	0.9539	<b>0.9584</b>	0.9555	

**Table 3.16 5° squares compare with 7° hexagons (intensity model)**

<b>Hexagons (7° longest diagonal)</b>				<b>Squares (5° side length)</b>			
Grid num	Num of Obs	R <sup>2</sup>	Substitute for lack-data grid	Grid num	Num of Obs	R <sup>2</sup>	Substitute for lack-data grid
1	1	0.9427	7	1	23	0.8698	
2	27	0.9083		2	33	0.8685	
3	0	0.9775	6	3	65	0.9200	
4	28	0.8367		4	14	0.9065	
5	49	0.9322		5	13	0.9704	
6	15	0.9775		6	5	0.8863	7
7	11	0.9427		7	13	0.8863	
8	200	0.9413		8	321	0.9624	
9	285	0.9506		9	337	0.9662	
10	0	0.9663	14	10	347	0.9611	
11	611	0.9660		11	0	0.0000	Inland
12	187	0.9665		12	184	0.9460	
13	0	0.9427	7	13	665	0.9694	
14	607	0.9663		14	796	0.9659	
15	675	0.9553		15	557	0.9393	
16	32	0.8233		16	0	0.0000	Inland
17	896	0.9656		17	69	0.9190	
18	347	0.9826		18	270	0.9107	
19	0	0.0000	Inland	19	418	0.9519	
20	193	0.9159		20	868	0.9810	
21	845	0.9676		21	0	0.0000	Inland
22	0	0.0000	Inland	22	0	0.0000	Inland
23	16	0.3179		23	0	0.0000	Inland
24	271	0.9936		24	14	0.8694	
25	0	0.0000	Inland	25	418	0.9681	
26	0	0.0000	Inland	26	0	0.0000	Inland
27	141	0.9383		27	0	0.0000	Inland
28	0	0.0000	Inland	28	0	0.0000	Inland
29	0	0.0000	Inland	29	0	0.9640	30
30	22	0.9601		30	29	0.9640	
Total num of Obs	5459		Num of lack-data grids	Total num of Obs	5459		Num of lack-data grids
Avg R <sup>2</sup>	<b>0.9591</b>		<b>4</b>	Avg R <sup>2</sup>	<b>0.9584</b>		<b>2</b>

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

**Table 3.17 Analysis for 6° hexagons (intensity model)**

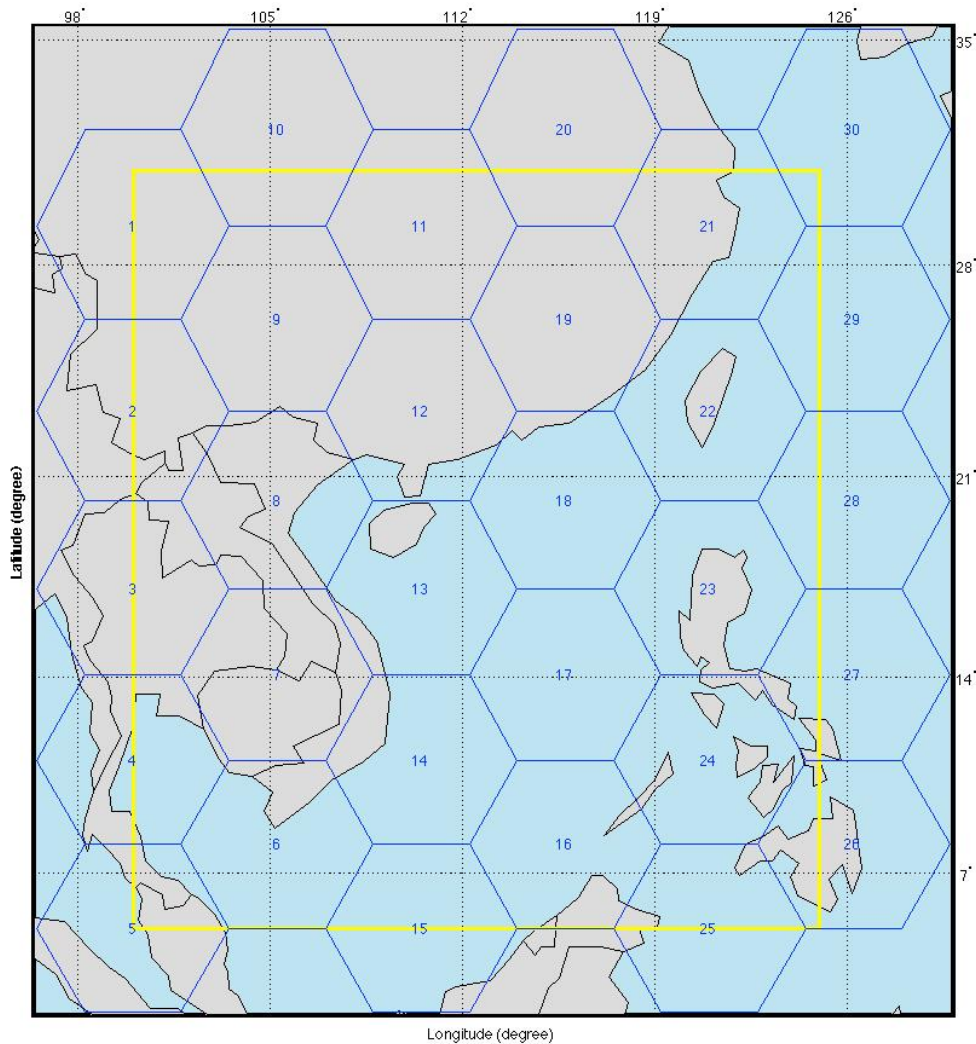
Grid number	Num of Obs	R <sup>2</sup>	Substitute for lack-data grids
1	0	0.9491	8
2	13	0.9368	
3	0	0.9677	10
4	0	0.9887	11
5	23	0.8653	
6	54	0.9253	
7	10	0.9653	
8	12	0.9491	
9	60	0.8579	
10	129	0.9677	
11	15	0.9887	
12	0	0.9603	16
13	332	0.9675	
14	338	0.9588	
15	0	0.9491	8
16	164	0.9603	
17	548	0.9584	
18	51	0.9629	
19	0	0.9603	16
20	708	0.9678	
21	527	0.9610	
22	0	0.0000	Inland
23	305	0.9188	
24	632	0.9722	
25	81	0.9851	
26	0	0.0000	Inland
27	191	0.8952	
28	805	0.9791	
29	0	0.0000	Inland
30	0	0.0000	Inland
31	117	0.8466	
32	80	0.9954	
33	0	0.0000	Inland
34	0	0.0000	Inland
35	249	0.9577	
36	0	0.0000	Inland
37	0	0.0000	Inland
38	0	0.0000	Inland
39	13	0.9802	
40	0	0.0000	Inland
41	0	0.0000	Inland
42	2	0.0000	
Total num of Obs	5459		Num of lack-data grids
Avg R <sup>2</sup>	0.9574		6

**Table 3.18 Analysis for 8° hexagons (intensity model)**

Grid number	Num of Obs	R <sup>2</sup>	Substitute for lack-data grid
1	5	0.9621	6
2	35	0.9231	
3	1	0.9705	5
4	42	0.7773	
5	85	0.9705	
6	9	0.9621	
7	376	0.9604	
8	334	0.9595	
9	25	0.8771	
10	960	0.9619	
11	0	0.0000	Inland
12	1069	0.9432	
13	798	0.9755	
14	66	0.8590	
15	777	0.9553	
16	0	0.0000	Inland
17	10	0.6974	
18	799	0.9767	
19	0	0.0000	Inland
20	19	0.8499	
21	0	0.0000	Inland
22	0	0.0000	Inland
23	49	0.9573	
Total num of Obs	5459		Num of lack-data grids
Avg R <sup>2</sup>	0.9571		2

**Table 3.19 Summary of scenarios for computational grids (intensity model)**

Grid type	Grid size (degrees)	Average R <sup>2</sup>	Number of grids		
			Lack-data	Total	% of lack-data
Square	5	0.9584	2	30	7
<b>Hexagon</b>	6	0.9574	6	42	14
	<b>7</b>	<b>0.9591</b>	<b>4</b>	<b>30</b>	<b>13</b>
	8	0.9571	2	23	9



**Figure 3.8 Computational grids and domain for the case study of Vietnam (intensity model)**

#### 3.4.3.5 The complete GSESM's approach for modelling intensity over water

Through a step-by-step calibration procedure (see flow chart in Figure 3.7), the thorough technique for modelling intensity over water, which contains all optimal solutions defined in the above stages, is formulated. Table 3.20 shows a comparison between the approach employed in the ETM and the method used in the model of this study (i.e., the GSESM) for the case study of Vietnam. As can be clearly seen from the table, GSESM's technique not only clarifies all remaining ambiguousness of the ETM, but also provides an improvement over the original method. While the average  $R^2$ , computed by equation 3.5, is increased from 0.9350 to 0.9591 (i.e., nearly 3% rise), the total number of lack-data grids is decreased from 6 to 4 (i.e., more than 30% reduction). Although there is not much of an upgrade in the linear performance, since the original formula is already a good approximation, other important modifications (e.g., using variable environmental pressures instead of a basin-wise fixed one) still prove the distinction between this study and the others.



One point to keep in mind is that, although at first glance, the technique is nearly identical to the one for modelling the track, every single detail is determined separately. It means that if there is any similarity in particulars between the two methods (e.g., 7° hexagon computational grids in both track and intensity simulations for the case study of Vietnam), it only happens by chance. That flexibility ensures the adaption of the general model to any possible conditions.

**Table 3.20 Comparison between approaches utilized in the ETM and the GSESM for the case study of Vietnam (intensity model)**

Method	Empirical Track Modelling (ETM)					Generalized Empirical Storm Modelling (GSESM)					
Grid type	Squares (5° side length)					Grid type	Hexagons (7° longest diagonal)				
Regression	OLS					Regression	Robust "talwar"				
DP type	Westbound		Eastbound		Substitute for lack-data grid	DP type	All tracks			Substitute for lack-data grid	
Grid number	Num of Obs	R <sup>2</sup>	Num of Obs	R <sup>2</sup>		Grid number	Num of Obs	R <sup>2</sup>			
1	19	0.8737	4	0.8881	3	1	1	0.9427	7		
2	30	0.8753	3	0.8881	3	2	27	0.9083			
3	46	0.9003	19	0.8881		3	0	0.9775	6		
4	14	0.9065	0	0.9228	9	4	28	0.8367			
5	13	0.9704	0	0.9024	10	5	49	0.9322			
6	5	0.8863	0	0.9672	7 & 12	6	15	0.9775			
7	13	0.8863	0	0.9672	12	7	11	0.9427			
8	287	0.9474	34	0.9284		8	200	0.9413			
9	310	0.9512	27	0.9228		9	285	0.9506			
10	336	0.9424	11	0.9024		10	0	0.9663	14		
11	0	0.0000	0	0.0000	Inland	11	611	0.9660			
12	171	0.8376	13	0.9672		12	187	0.9665			
13	584	0.9634	81	0.9615		13	0	0.9427	7		
14	584	0.9576	212	0.9522		14	607	0.9663			
15	442	0.9198	115	0.9237		15	675	0.9553			
16	0	0.0000	0	0.0000	Inland	16	32	0.8233			
17	60	0.9132	9	0.8954		17	896	0.9656			
18	233	0.8905	37	0.8839		18	347	0.9826			
19	261	0.8794	157	0.9467		19	0	0.0000	Inland		
20	595	0.9593	273	0.9726		20	193	0.9159			
21	0	0.0000	0	0.0000	Inland	21	845	0.9676			
22	0	0.0000	0	0.0000	Inland	22	0	0.0000	Inland		
23	0	0.0000	0	0.0000	Inland	23	16	0.3179			
24	10	0.8745	4	0.9561	25	24	271	0.9936			
25	262	0.8711	156	0.9561		25	0	0.0000	Inland		
26	0	0.0000	0	0.0000	Inland	26	0	0.0000	Inland		
27	0	0.0000	0	0.0000	Inland	27	141	0.9383			
28	0	0.0000	0	0.0000	Inland	28	0	0.0000	Inland		
29	0	0.0000	0	0.0000	Inland	29	0	0.0000	Inland		
30	12	0.9531	17	0.9847		30	22	0.9601			
Total num of Obs	4287		1172		Num of lack-data grids	Total num of Obs	5459			Num of lack-data grids	
	5459										
Avg R <sup>2</sup>	0.9350					6	Avg R <sup>2</sup>	0.9591			4

### 3.5 Decay model

The aforementioned relative intensity approach is only applicable whenever the TC's centre is over the sea. The reason is that it involves the SST (i.e., a factor is only obtainable at an oceanic DP). For locations removed from the coast, which are still heavily affected by TCs [i.e., up to 200 km inland (Vickery 2005)], a decay model is required to simulate TC's decay for use in a risk assessment.

#### 3.5.1 Existing methodology

To define the most suitable feature of the approach for modelling intensity after landfall in the GSESM, various details of the existing methods are discussed.

##### 3.5.1.1 Modelling intensity

Firstly, different types of the representative of intensity were modelled in different techniques described in the literature. While maximum sustained surface wind speed ( $V_{smax}$ ) was simulated by several researchers [e.g., Kaplan and Demaria (1995; 2001)], most models focused on central pressure ( $p_c$ ) [e.g., (Batts et al. 1980; Georgiou 1985; Ho et al. 1987; Vickery 2005; Vickery and Twisdale 1995a). The detailed causes of centring on  $p_c$ , which were already given in subsection 3.4.1, are not repeated here. Thus, in this study,  $p_c$  is estimated after TCs make landfall.

Secondly, the choices of a function to model the decay of TCs are also dissimilar. While almost all studies weakened a TC as a function of time since landfall [e.g., (Batts et al. 1980; Kaplan and DeMaria 1995; Vickery and Twisdale 1995a)], Georgiou (1985) employed the distance from the landfall position. In this research, to ensure the consistency with the majority of existing methods and, therefore, to enable an easy comparison with other approaches, the  $p_c$  after landfall is determined by an exponential decay function of time since landfall, in the form:

$$\Delta p_t = \Delta p_0 \exp(-a t) \quad (3.9)$$

Where:  $\Delta p_t$  - central pressure deficit at  $t$  (hours) after landfall (hPa);  $\Delta p_0$  - central pressure difference at the time of landfall (hPa);  $a$  - decay rate

Finally, as stated by Vickery et al. (2009), since the decay rate significantly depends on both local conditions (e.g., climatology, geography, or topography) and TC's properties, a decay model is only applicable for the specific region that is used to develop it. Thus, if a large research area is considered, it should be split into smaller parts to achieve better results. The separations are inconsistent in various techniques. For instance, while Georgiou (1985) presented decay rates for four distinct subareas of the U.S., Ho et al. (1987) utilized only three subareas for the same region. Thus, another benefit of the use of regional domain in this study is that it does not require such arbitrary division.

##### 3.5.1.2 Analysis of decay rate coefficients

Although the form of equation 3.9 is only a simplification, because many physical factors that govern the TC's decay process are ignored [e.g., land surface water (Shen et al. 2002)], it is still suitable for any long-term simulation (Vickery 2005). The reason is that a relatively simple formula can remarkably reduce the required computational demand, especially when subsequent analyses are needed.

Furthermore, there is a close correlation between the decay sub-model and the parent TC model. Thus, if a more sophisticated equation, including more parameters is employed in decay sub-model, a more complex method, which can evaluate the effects of those additional elements, must also be used in parent model. Because the invention, the development and the application of such kind of models are unfeasible, at least for now, the identical form of equation 3.9 is utilized in this study.

For modelling purposes, the decay rate ( $a$ ) must also be estimated at any time step after landfall. While Vickery and Twisdale (1995a) defined  $a$  by using a linear function of  $\Delta p_0$ , Vickery (2005) investigated the impact of two other parameters, including the radius of maximum wind speed and the translation speed, corresponding to the time of landfall. The modifications, therefore produced in total three different formulae for estimating  $a$ . Since only one equation is usable for any particular area, Vickery (2005) evaluated the performance of all three formulae to determine the optimal one. The equations are provided as follows:

$$a = d_1 + d_2 \Delta p_0 + \varepsilon \quad (3.10)$$

$$a = d_1 + d_3 \Delta p_0 / R_{max0} + \varepsilon \quad (3.11)$$

$$a = d_1 + d_4 \Delta p_0 c_0 / R_{max0} + \varepsilon \quad (3.12)$$

Where:  $R_{max0}$  - radius of maximum wind speed at the time of landfall (km);  $c_0$  - translation speed at the time of landfall ( $m\ s^{-1}$ );  $d_1, d_2$ , etc. - decay rate coefficients

One more point to keep in mind is that, the time of landfall mentioned in the simulation approaches is actually taken to be the closest 6-hour time step *prior* to landfall (Wang and Rosowsky 2012). This is because human knowledge has not advanced to the point of yielding reliable estimates of TC's variations between two contiguous steps. Even if the landfall location can be approximately determined by finding the intersection of an implied line connecting these DPs (i.e., by using the implicit concept of a TC's track) and a smooth coastline (Vickery 2005), there is no way to obtain other parameters (e.g.,  $\Delta p_0$ ,  $R_{max0}$ , or  $c_0$ ) at exactly the position where the TC makes landfall. In short, since people know very little about what really happens between the DPs before and after landfall, any kind of interpolation is not rational.

## 3.5.2 The GSESM's approach

### 3.5.2.1 Estimating decay constant

The method used in this study follows the one introduced by Vickery and Twisdale (1995a) and its most recent updated and improved version by Vickery (2005), which are widely employed in many subsequent researches [e.g., (ASCE 1998, 2003; FEMA 2003; Powell et al. 2005)]. The approach for modelling  $a$  is presented as follows:

- Firstly, DPs measured after 1977, which are inside the TA and have the strength of a tropical storm or even stronger (i.e.,  $V_{smax}$  greater than or equal to 30 kt) are obtained from the JMA's BTD, as a result of several reasons given in subsection 2.1.1 and 2.2.2.

- Secondly, among those DPs, the DPs that have valid wind field calculations (the detailed analyses will be provided later, in chapter 4) are taken into account. This is because one factor (i.e.,  $R_{max0}$ ), which is utilized for estimating decay rate ( $a$ ) in equations 3.11 and 3.12, requires the soundness of these wind field computations.
- For a decay model, obviously, only TCs that make landfall are useable. A DP at the closest 6-hour time step prior to landfall is employed for obtaining required parameters at the time of landfall (i.e.,  $\Delta p_0$ ,  $R_{max0}$ , and  $c_0$ ). Values of  $\Delta p_t$  at all inland DPs after that point are used to determine decay rate ( $a$ ) of the corresponding landfall by using a nonlinear regression in equation 3.9. Note that, in reality, there are several TCs with multiple landfalls [e.g., Hurricane Katrina of the 2005 Atlantic hurricane season (Belanger et al. 2009)]. These situations are included in both historical analysis and the GSESM, in which landfalls are evaluated independently, even if they caused by the same TC. For instance, the 120 landfalls shown in Table 3.22 actually consist of 160 TCs, since 4 of them make a double landfall. However, in some cases, several DPs are unusable and must be excluded from the model. Most of the time, this is because the weakening of TC right after the moment of landfall can be so unnoticeable that the central pressure deficit at the time of landfall ( $\Delta p_0$ ) is equal to the one after that moment ( $\Delta p_t$ ) in the BTD record. That leads to the zero value of decay rate ( $a$ ), which is meaningless for the model. Moreover, in very rare cases (e.g., for the case study of Vietnam that shown in Table 3.22, they only account for less than 2% of total events), the plain form of the exponential function is not sufficient to describe the decay processes.

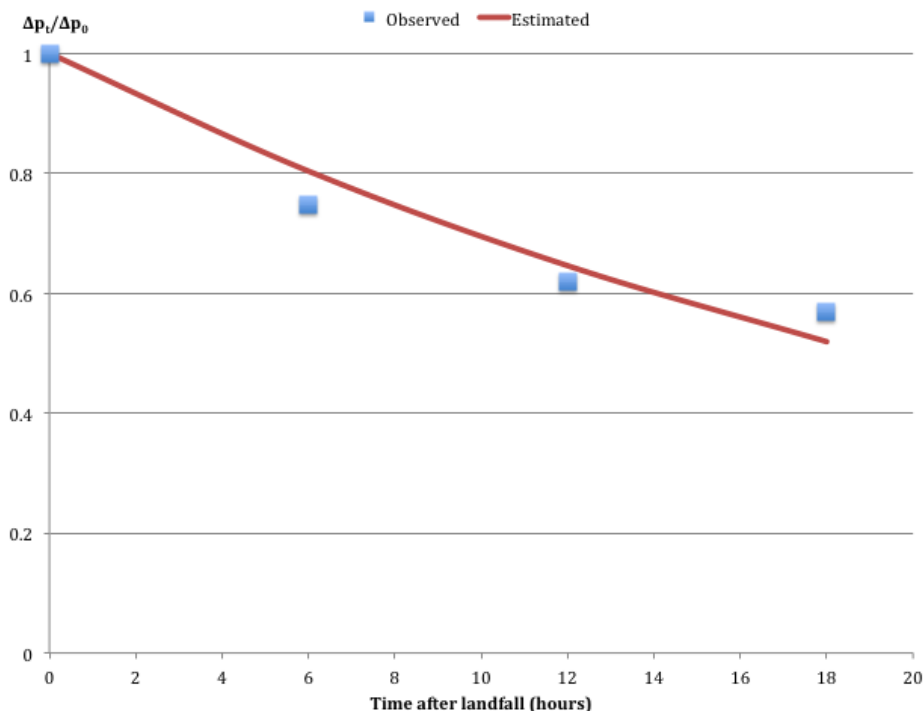
**Table 3.21 Analysis of DPs for use in a decay model**

Number of	All TCs in the WNP BTD (1)	(1) with measurements for wind radius (2)	(2) with at least 1 DP inside the TA (3)	(3) with DPs inside the TA (4)	(4) with DPs have tropical storm or higher intensity (5)	(5) with DPs have valid wind field calculations (6)
TCs	1650	943	514	514	495	464
DPs	60374	35106	17935	9037	6147	5071

**Table 3.22 Analysis of nonlinear regression in the decay model**

Number of	(6) with TCs make landfalls (7)	(7) with DPs are unusable for decay model		(7) with inland DPs fit exponential decay (8)			
		DPs with negligible decays	Decays do not fit with exponential function	Total	(8) with only 1 inland DP (9)	(8) with more than 1 inland DPs (10)	
Landfalls	120	39	2	79	52	27	
DPs	199	80	4	115	52	63	
$R^2$ of equation (3.4)				Mean	0.9533	1	0.8633
				Max	1	1	0.9952
				Min	0.6134	1	0.6134

For the case study of Vietnam, Table 3.21 and Table 3.22 give the analyses of the DPs as well as the nonlinear regression used in the decay model. As can be seen from the tables, the mean  $R^2$  for all exponential fits is 0.9533, ranging from the lowest value of 0.6134 for Typhoon Nancy of the 1982 season (see Figure 3.9) to a high of 1.00. The maximum  $R^2$  values of 1.00 are corresponding to the landfalls that have only two DPs (i.e., one inland DP and one DP at the time of landfall). Regarding only landfalls with three or more DPs (i.e., more than one inland DPs, denoted by type 10 in Table 3.22), the average  $R^2$  is 0.8633. The high  $R^2$  shows that simulating TC's decay using an exponential formula (i.e., equation 3.9) is an appropriate method. Agreements with this conclusion can be found in many research in the literature [e.g., (Vickery 2005; Vickery and Twisdale 1995a)].



**Figure 3.9 Observed and estimated decay after landfall, TC Nancy (1982)**

### 3.5.2.2 Modelling decay rate coefficients

The GSESM's approach improves the one used by Vickery (2005) in two stages:

- First, instead of examining only three equations (3.10, 3.11, and 3.12) to determine the optimal one (i.e., largest adjusted  $R^2$ , see paragraph 3.3.3.1), all possible combinations of available terms are considered. Following the idea provided in paragraph 3.3.4.3, results are given in Table 3.23, where each term is represented by its coefficient. For instance, " $d_2$ " is the representative of the " $d_2 \Delta p_0$ " term. For the case study of Vietnam, the table shows that although the scenario with the largest adjusted  $R^2$  is option 1, it leads scenarios 4, 5, and 7 by very small margins. Moreover, the use of scenario 1 does not include two factors, which may contribute to the observed variations of decay rate ( $a$ ) in the real TCs. The introduction of the first additional element (i.e., the radius of maximum winds at the time of landfall  $R_{max0}$ ) is based on the perception that smaller TCs would tend to weaken more quickly than the larger ones (Kaplan and DeMaria 1995; Malkin 1959). The notion of possible impacts of  $R_{max0}$  originates from the theory that for a large TC, the relatively smaller amount of TC's core energy is lost at a slower pace when it comes onshore, in comparison with the situation happening in a smaller TC (Vickery 2005). Thus, it is logical to include another factor, which can quantify how quickly the TC's energy is extracted from its source. This parameter is the translation speed at the time of landfall ( $c_0$ ). Thus, among the four options with the highest adjusted  $R^2$  (i.e., scenarios 1, 4, 5, and 7) options 5 and 7, which include both  $R_{max0}$  and  $c_0$  are the preferable ones. Option 5 has a higher adjusted  $R^2$  value than option 7, and therefore it is chosen as the definite equation for the modelling decay rate coefficients for the case study of Vietnam, as follows:

$$a = d_1 + d_2 \Delta p_0 + d_4 \Delta p_0 c_0 / R_{max0} + \epsilon_{35} \quad (3.13)$$

- Secondly, different linear regression solutions for equation 3.13 are also considered in the GSESM, as shown in Table 3.24, for the case study of Vietnam, including the OLS and various robust regression approaches. As can be seen in the table, the solution that provides the best performance (i.e., largest  $R^2$ , see paragraph 3.3.3.1) for equation 3.13 is the "cauchy" robust regression.

Table 3.25 shows a comparison between the GSESM's formula and the empirical equations for the modelling decay rate coefficients used in other research, including:

- Formulae 3.14, 3.15, and 3.16 (Vickery and Twisdale 1995a):

$$a = 0.06 + 0.00046 \Delta p_0 \quad (3.14)$$

$$a = 0.035 + 0.00050 \Delta p_0 \quad (3.15)$$

$$a = 0.038 - 0.00029 \Delta p_0 \quad (3.16)$$

Formulae 3.17, 3.18, 3.19, and 3.20 (Vickery 2005):

$$a = 0.0413 + 0.0018 \Delta p_0 c_0 / R_{max0} \quad (3.17)$$

$$a = 0.0225 + 0.0017 \Delta p_0 c_0 / R_{max0} \quad (3.18)$$

$$a = 0.00364 + 0.0016 \Delta p_0 c_0 / R_{\max 0} \quad (3.19)$$

$$a = 0.0034 + 0.0010 \Delta p_0 \quad (3.20)$$

The results shown in the table are various goodness-of-fit measures. They confirm that the GSESM's approach is definitely a considerable improvement over the techniques employed in other studies.

**Table 3.23 Combinations of usable terms (decay rate coefficient model)**

Combination	d <sub>1</sub> ; d <sub>2</sub>	d <sub>1</sub> ; d <sub>3</sub>	d <sub>1</sub> ; d <sub>4</sub>	d <sub>1</sub> ; d <sub>2</sub> ; d <sub>3</sub>	<b>d<sub>1</sub>; d<sub>2</sub>; d<sub>4</sub></b>	d <sub>1</sub> ; d <sub>3</sub> ; d <sub>4</sub>	d <sub>1</sub> ; d <sub>2</sub> ; d <sub>3</sub> ; d <sub>4</sub>
Combi num	1	2	3	4	<b>5</b>	6	7
Adjusted R <sup>2</sup>	0.3048	0.1231	0.1307	0.2961	<b>0.2999</b>	0.1235	0.2944
R <sup>2</sup>	0.3137	0.1343	0.1418	0.3142	<b>0.3179</b>	0.1459	0.3215

**Table 3.24 Options for regression solution (decay rate coefficient model)**

Method	OLS	Robust "andrews"	Robust "bisquare"	<b>Robust "cauchy"</b>	Robust "fair"	Robust "huber"	Robust "logistic"	Robust "talwar"	Robust "welsch"
R <sup>2</sup>	<b>0.3179</b>	0.3322	0.3338	<b>0.3417</b>	0.3276	0.3382	0.3415	0.3138	0.3386

**Table 3.25 Comparison between the GSESM' equation and various empirical formulae for the case study of Vietnam (decay rate coefficient model)**

Goodness-of-fit measure number <sup>20</sup>	<b>GSESM (formula 3.13)</b>	Formula 3.14	Formula 3.15	Formula 3.16	Formula 3.17	Formula 3.18	Formula 3.19	Formula 3.20
1	<b>0.0007</b>	0.0011	0.0008	0.0021	0.0011	0.0020	0.0035	0.0011
2	<b>0.6812</b>	1.1311	0.8222	2.0757	1.0987	1.9619	3.5563	1.0701
3	<b>0.0261</b>	0.0336	0.0286	0.0455	0.0331	0.0442	0.0596	0.0327
4	<b>0.8253</b>	1.0635	0.9068	1.4407	1.0482	1.4007	1.8858	1.0344
5	<b>0.0200</b>	0.0299	0.0212	0.0321	0.0235	0.0327	0.0511	0.0225
6	<b>0.3728</b>	0.7592	0.4024	0.4467	0.3882	0.4632	0.8522	0.3209
7	<b>0.5636</b>	0.5601	0.5601	-0.5601	0.3766	0.3766	0.3766	0.5601
8	<b>0.3417</b>	0.3137	0.3137	0.3137	0.1418	0.1418	0.1418	0.3137
9	<b>0.3101</b>	-0.1456	0.1673	-1.1023	-0.1128	-0.9870	-2.6019	-0.0838
10	<b>0.0969</b>	0.0895	0.1124	0.1520	0.1284	0.1475	0.1666	0.1170
11	<b>1.2691</b>	3.3431	1.8455	0.9677	1.6217	0.8459	0.9614	0.7533

<sup>20</sup> List of goodness-of-fit measures presented in the table (denoted by number):

- 1 - mean squared error
- 2 - normalised mean squared error
- 3 - root mean squared error
- 4 - normalised root mean squared error
- 5 - mean absolute error
- 6 - mean absolute relative error
- 7 - coefficient of correlation
- 8 - coefficient of determination
- 9 - coefficient of efficiency
- 10 - maximum absolute error
- 11 - maximum absolute relative error

## 4 SPATIAL WIND FIELD MODEL

### Research questions:

- Which are the current theoretical frameworks for modelling key parameters? What are their pros and cons? Which one should be chosen as a basis for the Generalized Stochastic Empirical Storm Model (GSESM)?
- What are the limitations of the (chosen) existing method? How can the GSESM overcome these limitations? In comparison with the original technique, what are the GSESM's improvements?

As indicated earlier, a Tropical Cyclone (TC) is an area of low pressure where energy is accumulated when the water vapour changes its state (Brettschneider 2006). In the previous chapter, the real TCs are represented by their centres in order to simplify the actual complexity of the modelling of the developments in track and intensity, while the TCs propagate over time. However, since a TC actually does not concentrate on any single point, a wind field model must be employed to simulate the spread of the TC's effects over a broad region based on conditions at the centre. Such a model can determine both pressure and wind speed at all positions in the vicinity of the TC, and therefore it forms a basis for other related analyses (e.g., risk assessments or storm surge estimates). When combined with the track and intensity model, they can approximate the real TC's behaviour with an acceptable accuracy for practical applications.

This chapter is concerned with the modelling of the general parametric wind field (section 4.1) and its two most important factors, which are the radius of maximum wind and the Holland parameter ( $B$ ), in sections 4.2 and 4.3, respectively.



## 4.1 General parametric wind field model

### 4.1.1 Existing techniques

In most TC analyses, the modelling of the surface wind field is especially important, since it directly drives the storm surge and wave simulations and governs their efficiency and accuracy (Hu, Chen, and Fitzpatrick 2012). Thus, in recent decades, there have been numerous attempts to develop wind field models to be used in the practical applications.

Several of the existing methodologies are relatively complicated, such as:

- The kinematic analysis wind approaches [e.g., the H\*Wind (Powell et al. 1998, 2010)]
- The steady-state slab Planetary Boundary Layer (PBL) models [e.g., (Thompson and Cardone 1996; Vickery, Skerlj, Steckley, et al. 2000)]
- The interactive objective kinematic analysis systems, that are a combination of actual measurements and the PBL models [e.g., the IOKA (Cardone and Cox 2009; Cox et al. 1995)]
- The mesoscale weather models [e.g., the Weather Research and Forecasting Model (WRF), the Geophysical Fluid Dynamics Laboratory Operational Hurricane Prediction System (GFDL), and the Pennsylvania State University/National Center for Atmospheric Research model (MM5)].

In some cases, such methods were utilized to force the surge, the wave and the hydrodynamic systems [e.g., (Li et al. 2006; Xie et al. 2008)]. However, the complexity of these techniques, especially the mesoscale ones [e.g., (Kwun et al. 2009; Mandal et al. 2004)], makes it difficult for them to either simply or completely integrate with the storm surge simulations [e.g., the Advanced Coastal Circulation Model (ADCIRC) (Luettich Jr et al. 1992)] and wave models [e.g., the Simulating Waves Nearshore (SWAN) (Booij et al. 1999)], as stated in other research [e.g., (Hu, Chen, and Fitzpatrick 2012; Hu, Chen, and Kimball 2012)].

Another distinct branch of TC wind estimates is the parametric wind field model. Using this technique, the radial wind profiles are represented by several mathematical expressions, which consist of only a few key inputs (e.g., central pressure, or radius of maximum wind). Because of its reasonable simplicity, extremely low computational demand and flexibility in domain resolution, the parametric method is chosen in many practical studies, particularly the one related to long-term simulations, such as the TC risk assessments or the establishments of extreme wind speed and probability maps [e.g., (Holland et al. 2010; Vickery and Twisdale 1995a)]. Although the formulae used in a parametric wind model are intentionally kept simple, they still reveal the fundamental principles of a TC structure, in which pressure drops exponentially towards the centre until it attains its minimum level at the eye, while wind speed first strengthens exponentially towards the centre, reaches its maximum value at the eyewall, and then weakens to a relatively calm condition at the eye. The equations that describe the radial profiles are the modifications of the Rankine combined vortex formulae, which are based on the assumptions of a solid-body rotation inside the eyewall and the

reducing tangential wind due to a radial scaling parameter outside that region (Hu, Chen, and Fitzpatrick 2012; Schloemer et al. 1954)

#### 4.1.2 Parametric wind field methodology

Following the concept of a modified Rankine vortex approach, numerous research has been conducted since the 1970s, with gradual improvements in both clarity and accuracy (Vickery, Masters, et al. 2009).

Generally, a parametric wind field model consists of two major stages:

- The first step is estimating the wind speed at the top of the boundary layer [i.e., at gradient height, usually taken to be 500 - 2000 m (Vickery and Wadhera 2008)]. This value is assumed to be equivalent to a mean wind speed associated with a long averaging time, such as a 10-minute period (Vickery, Masters, et al. 2009).
- Such gradient wind speed is then modified to an average surface value at a specific height (taken to be 10 m in most cases) by using the atmospheric boundary layer notions [e.g., (Vickery, Wadhera, Powell, et al. 2009)] with an assumption of neutral stability (Vickery, Masters, et al. 2009).

##### 4.1.2.1 Gradient wind field modelling

An example of the pioneering works in estimating the gradient wind field was presented by Batts et al. (1980), in which the maximum value of wind speed at gradient height is obtained from:

$$V_{gmax} = K \sqrt{\Delta p - \frac{R_{max} f}{2}} \quad (4.1)$$

Where:  $V_{gmax}$  - maximum gradient wind speed ( $m s^{-1}$ );  $K$  - empirical constant;  $R_{max}$  - radius of maximum wind (m);  $f$  - Coriolis parameter ( $rad s^{-1}$ );  $\Delta p$  - central pressure deficit (hPa).

$$f = 2 \Omega \sin \Psi$$

$$\Delta p = p_n - p_c$$

Where:  $\Omega$  - the rotation rate of the Earth, taken to be  $7.2921 \cdot 10^{-5}$  ( $rad s^{-1}$ ) (Vallis 2006);  $\Psi$  - the latitude of the TC's centre (degrees);  $p_n$  and  $p_c$  are the environmental pressure and central pressure (hPa), respectively.

In the case that the Coriolis effect is ignored [i.e.,  $f \approx 0$  (Holland, 1980)], equation 4.1 becomes:

$$V_{gmax} \approx K \sqrt{\Delta p} \quad (4.2)$$

As can be seen from formula 4.2,  $V_{gmax}$  in Batts et al.'s (1980) model and similar studies [e.g., (Russell 1968; Schwerdt et al. 1979; Tryggvason et al. 1976)] is directly proportional to  $\sqrt{\Delta p}$ . However, the most critical disadvantage of those techniques is that the models cannot successfully describe the variation in observed wind speeds for a given central pressure.

As a result, Holland (1980) introduced an additional factor to the modelling of wind speed, which is often referred to as the Holland parameter ( $B$ ). Because TC winds contain a remarkable fluctuation in the radial profiles, Holland (1980) adjusted the Schloemer et al.'s (1954) formula to approximate the radial pressure variation by using the rectangular hyperbolas, as follows:

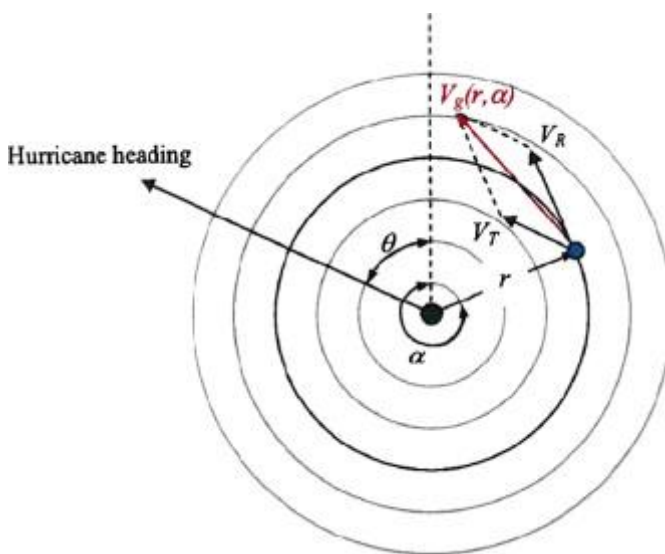
$$p_r = p_c + \Delta p \exp\left(-\frac{A}{r^B}\right) \quad (4.3)$$

Where:  $p_r$  – the surface pressure at a distance  $r$  (m) from the TC's centre (hPa);  $A$  – the location parameter;  $B$  – the pressure profile factor (Holland parameter).

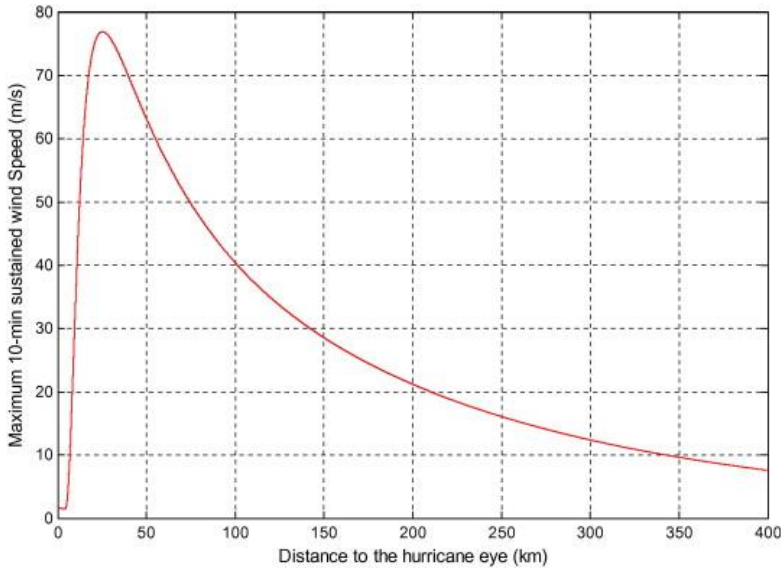
Holland (1980) provided that  $R_{\max} = A^{1/B}$  and, therefore equation 4.3 can be presented as:

$$p_r = p_c + \Delta p \exp\left[-\left(\frac{R_{\max}}{r}\right)^B\right] \quad (4.4)$$

The real basis structure of a gradient wind field can be explored by utilizing the aircraft measurements together with the existing knowledge of translation speed. Note that, although most of the time, the flight-level is about 3000 m, which is higher than the actual gradient-level (i.e., 500 - 2000 m), several researchers [e.g., (Powell 1990; Sparks and Huang 1999)] stated that there is a very little difference in the measured wind speeds at those two levels. Using aircraft observations, it is confirmed that the well-formed gradient wind field can be simulated as a translating vortex (Lee and Rosowsky 2007). Thus, the gradient wind speed can be considered as a combination of translational and rotational components, which enables the expression of such rotative speed as a function of distance from the TC's centre [see Figure 4.1 and Figure 4.2 (Wang and Rosowsky 2012)]. Once the gradient rotational wind speed vortex is assumed to be axially symmetrical about the centre, a TC can simply be considered as a rotational vortex propagating with translation speed.



**Figure 4.1 Decomposition of wind speed at gradient-level [from (Wang and Rosowsky 2012)]**



**Figure 4.2 An example of radial profiles of a TC wind, Hurricane Katrina (2005) [from (Wang and Rosowsky 2012)]**

A Rankine combined vortex, which is the fundamental approximation of parametric radial wind profiles, describes the TC's structure as a solid-body rotation (i.e., constant vorticity) inside the eyewall combines with a circulation with zero vorticity outside the region (Holland et al. 2010). Thus, the wind speed is obtained from:

$$V_{gr} = V_{gmax} \left( \frac{r}{R_{max}} \right) \quad (r < R_{max})$$

$$= V_{gmax} \left( \frac{R_{max}}{r} \right)^x \quad (r \geq R_{max})$$

Where:  $V_{gr}$  – the gradient wind speed at a distance  $r$  (m) from the TC's centre ( $m s^{-1}$ );  $x$  – the scaling parameter that modifies the profile shape (Depperman 1947).

Although in the original Rankine vortex,  $x$  is taken to be 1, smaller values are often employed to include the effect of surface friction on the angular momentum loss that normally observed in a real TC. The use of  $x = 0.5$ , introduced by (Riehl 1954), is a popular choice in most studies.

For a stationary TC, the cyclostrophic wind equation is applied to the gradient height, as follows (Vickery, Skerlj, and Twisdale 2000):

$$\frac{1}{\rho} \frac{\partial p_r}{\partial r} = \frac{V_{gr}^2}{r} + f V_{gr} \quad (4.5)$$

Substituting (4.4) into (4.5), one can determine the gradient balance velocity from:

$$V_{gr} = \left\{ \frac{B}{\rho_a} \left( \frac{R_{max}}{r} \right)^B \Delta p \exp \left[ - \left( \frac{R_{max}}{r} \right)^B \right] + \gamma \left( \frac{r f}{2} \right)^2 \right\}^x - \gamma \left( \frac{r f}{2} \right) \quad (4.6)$$

Where:  $\rho_a$  - density of air ( $kg m^{-3}$ ).

Although  $\rho_a$  can be estimated by using an approximation in Holland's (2008) research, this unnecessarily magnifies the model's complexity, since such a

parameter is not crucial and a fixed value can be employed with no significant influence on the results (Holland et al. 2010). Therefore, following (Holland 1980),  $\rho_a$  is taken to be  $1.15 \text{ (kg m}^{-3}\text{)}$ .

While Holland (1980) neglected the Coriolis effect (i.e.,  $\gamma = 0$  in formula 4.6), Hu et al. (2012) retained this factor (i.e.,  $\gamma = 1$ ) and indicated that the exclusion of Coriolis term results in an up to 20% error in the estimates of  $V_{gmax}$  for weak but large TCs.

Using the recommended values from the previous studies (i.e.,  $x = 0.5$  and  $\gamma = 1$ ), equation 4.6 becomes:

$$V_{gr} = \sqrt{\frac{B}{\rho_a} \left(\frac{R_{max}}{r}\right)^B \Delta p \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right] + \left(\frac{rf}{2}\right)^2} - \left(\frac{rf}{2}\right) \quad (4.7)$$

At the eyewall (i.e.,  $r = R_{max}$ ), the maximum gradient wind speed is:

$$V_{gmax} = \sqrt{\frac{B \Delta p}{\rho_a e} + \left(\frac{R_{max} f}{2}\right)^2} - \left(\frac{R_{max} f}{2}\right) \quad (4.8)$$

In the case that the Coriolis effect is ignored [i.e.,  $f \approx 0$  (Holland 1980)], equation 4.8 becomes:

$$V_{gmax} \approx \sqrt{\frac{B \Delta p}{\rho_a e}} \quad (4.9)$$

Where:  $e$  – the base of the natural logarithm.

Equation 4.9 shows that  $V_{gmax}$  in the approach, based on Holland's (1980) methodology, is directly proportional to  $\sqrt{B\Delta p}$ , instead of  $\sqrt{\Delta p}$  in the pioneering works [e.g., (Batts et al. 1980)] as presented in formula 4.2. Thus, the variation in observed wind speed for a given central pressure in the historical TC records can be described by using  $B$ , which varies from 0.5 to 2.5 (Holland 1980). For instance, a larger value of  $B$  represents the bigger concentration of decrease in the atmospheric pressure near the eyewall and makes the wind profile more "peaked" (Hu, Chen, and Fitzpatrick 2012). Therefore, the technique is a considerable improvement over the modelling of wind fields using a formula in the form of equation 4.2 (Vickery and Wadhera 2008).

Most of subsequent researchers have followed the basic principles provided by Holland's (1980) approach [e.g., (Georgiou 1985; Vickery, Skerlj, Steckley, et al. 2000)]. Although the methods are limited by the two dimensional (2D) nature of the equation (Hu, Chen, and Fitzpatrick 2012), they still give a means to include the effects of various important elements in the models, such as the surface friction or sea-land transition (Vickery, Masters, et al. 2009). Three dimensional (3D) studies, although they have not yet been found in any peer-reviewed publications so far, are likely to be applied in the near future when significant improvements in computational power will be made. The use of 3D models will possibly present a means to better describe the change in the vertical formation of the TC [e.g., (Foster 2005)].

While popular, the Holland's (1980) model still has some drawbacks, such as the inability to accurately represent the double eyewalls or to simultaneously simulate the wind speeds at the eyewall and the outer-core, as has been discussed in several

studies [e.g., (Thompson and Cardone 1996; Vickery, Wadhera, Powell, et al. 2009; Willoughby et al. 2006; Willoughby and Rahn 2004). Moreover, according to Hu et al. (2012), the most serious problem of Holland's (1980) model is related to its assumption of the symmetric vortexes, which are hardly observed in reality, particularly at landfall locations. This has lead to several attempts to modify Holland's (1980) technique, which is to introduce asymmetry into the method. An example of such an approach was provided by Xie et al. (2006) and the subsequent methodology (Mattocks and Forbes 2008) based on that research, which was utilized in the ADCIRC storm surge model. However, the inclusion of an additional factor requires a significant increase in computational demand for long-term simulation and the assumption of symmetric vortexes is generally considered to be acceptable for well-formed TCs (Wang and Rosowsky 2012). Thus, the disadvantages of Holland's (1980) model are ignored and this method is employed in many TC risk studies [e.g., (Georgiou et al. 1983; Harper 1999; Lee and Rosowsky 2007)].

Although the gradient rotational wind speed vortex is assumed to be symmetrical near the centre in the above procedure, the asymmetry in the wind speed due to translation speed can actually be included by utilizing the Blaton's adjusted radius of curvature (Georgiou 1985) based on Shapiro's (1983) concept, which is:

$$r_t = \frac{r}{1 - \frac{c}{V_{gr}} \sin \alpha}$$

Where:  $c$  – the translation speed ( $m\ s^{-1}$ );  $\alpha$  – the angle from the translation direction to the profile location (degrees, positive clockwise).

Substituting (4.4) into (4.5), replacing  $r$  with  $r_t$ , and using the values of  $x$  and  $y$  suggested by equation 4.7, the formula for estimating the gradient wind speed is (Vickery, Skerlj, and Twisdale 2000):

$$V_{gr} = \frac{1}{2}(c \sin \alpha - f r) + \sqrt{\frac{1}{4}(c \sin \alpha - f r)^2 + \frac{B \Delta p}{\rho_a} \left(\frac{R_{max}}{r}\right)^B \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right]} \quad (4.10)$$

At the eyewall (i.e.,  $r = R_{max}$ ), the maximum gradient wind speed is:

$$V_{gmax} = \frac{1}{2}(c \sin \alpha - f R_{max}) + \sqrt{\frac{1}{4}(c \sin \alpha - f R_{max})^2 + \frac{B \Delta p}{\rho_a e}} \quad (4.11)$$

Equations 4.10 and 4.11 indicate that in order to accurately reproduce the asymmetry in the wind field, the single value of  $B$  (or  $R_{max}$ ) must be replaced by the directionally varying quantities  $B_\alpha$  (or  $R_{max\alpha}$ ). In some areas [e.g., the Atlantic basin], meteorological agencies recently provide data that can be a basis for feasible estimates of those directionally varying values, such as the specified radii associated with different wind speeds in four quadrants. However, since that four-quadrant information is not readily available in most of TC records at other locations, the techniques that included the asymmetry in the wind field are restricted only to the basins where the directional data are obtainable.

In the Western North Pacific (WNP), although the Japan Meteorological Agency (JMA) gives the information of maximum and minimum radii of the 30 kt and 50 kt wind speeds, as well as the directions of those maximum wind radii in the Best

Track Data (BTD), such parameters are not really helpful. That is because the main technique for assembling the BTD set, based on the satellite imagery (i.e., the Dvorak method), has many limitations (see subsection 2.1.1), which leads to the lack of required accuracy to determine the asymmetry structure of the TCs. For instance, at lots of Data Points (DPs), the Dvorak technique is not sensitive enough to identify the distinction between the maximum and minimum 30 kt wind radii, which results in equal estimates of these two quantities. That means that two different values of  $\alpha$  yield the same result of  $V_{gr}$  (i.e., the gradient wind speed associated with the surface 30 kt one) with the identical quantity of  $r$ , due to the low-sensibility of the Dvorak technique. In such cases, no definite solution can be derived from equations 4.10 and 4.11.

Therefore, only one radius value is utilized in this study. Due to "the safer the better" approach and the significance of the maximum radius of 30 kt wind ( $R_{max30}$ ), discussed in subsection 2.2.2,  $R_{max30}$  is employed, which corresponds to the profile angle  $\alpha = 90$  degrees (i.e.,  $\sin \alpha = 1$ ). Formula 4.10 becomes:

$$V_{gr} = \frac{1}{2}(c - fr) + \sqrt{\frac{1}{4}(c - fr)^2 + \frac{B \Delta p}{\rho_a} \left(\frac{R_{max}}{r}\right)^B \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right]} \quad (4.12)$$

At the eyewall (i.e.,  $r = R_{max}$ ), the maximum gradient wind speed is:

$$V_{gmax} = \frac{1}{2}(c - f R_{max}) + \sqrt{\frac{1}{4}(c - f R_{max})^2 + \frac{B \Delta p}{\rho_a e}} \quad (4.13)$$

#### 4.1.2.2 Gradient-to-surface adjustment and sea-land reduction in wind speed

Giving a value of mean wind speed at gradient-level ( $V_g$ ) in the previous step, the wind speed at the surface (typically 10 m above water or ground, denoted as  $V_s$ ) is estimated by adjusting the corresponding gradient quantity through the use of a simple wind speed conversion factor ( $V_s/V_g$ ) or, more sophisticatedly, an atmospheric boundary layer model.

The simple wind speed reduction factors, both for over water and overland cases, are widely utilized in the pioneering studies [e.g., (Batts et al. 1980; Georgiou 1985; Schwerdt et al. 1979)] as well as in the most recent research [e.g., (Lee and Rosowsky 2007; Sparks and Huang 2001; Wang and Rosowsky 2012)]. The values for winds over the ocean used in those models vary from a low of 0.650 (Sparks and Huang 2001) to 0.865 (Batts et al. 1980) or even rise as high as 0.950 in the technique presented by Schwerdt et al. (1979). Georgiou (1985) also took into account the profile location that led to two distinct conversion factors (i.e., 0.825 near the eyewall and decreasing to 0.750 away from that region). For the overland cases, the surface winds are further weakened under the influence of land roughness, which results in an immediate reduction in the mean speed as the winds move from sea to land. That reduction is also different in various studies, ranging from 11-22% (Schwerdt et al. 1979), 15% (Batts et al. 1980), 16-25% (Georgiou 1985), to a high of about 30% (Sparks and Huang 2001). The corresponding wind speed ratios  $V_s/V_g$  are 0.845 at the coast reducing to 0.745 at 19 km inland (Schwerdt et al. 1979), 0.740 (Batts et al. 1980), 0.620 (Georgiou 1985), and 0.450 (Sparks and Huang 2001).

In other research [e.g., (Powell et al. 2005, 2003; Vickery, Skerlj, Steckley, et al. 2000; Vickery, Wadhera, Powell, et al. 2009)], the wind speed conversion factors are derived from the atmospheric boundary layer models. However, these models are generally very complex, depend on lots of factors [e.g., wind speed (Powell et al. 2005; Vickery, Skerlj, Steckley, et al. 2000) or TC size (Vickery, Wadhera, Powell, et al. 2009)], and require accurately observed information from the dropsondes, which are very limited in many regions, especially in case of overland profiles. Thus, the simple wind speed reduction values are employed in most long-term simulations [e.g., (Lee and Rosowsky 2007; Wang and Rosowsky 2012)]. Table 4.1 provides a summary of the values of  $V_s/V_g$  and the additional reductions, due to the sea-land transition, for various models available in the literature.

In this study, although the simple wind speed reduction factors are used, they are taken to be the values suggested by the most advanced and recent model of Vickery et al. (2009). While the value of  $V_s/V_g$  is taken to be 0.71 when TCs are over the ocean, the immediate reduction in the mean speed as the winds move from sea to land is 19% (average of 18-20% shown in the table). As a result, the value of  $V_s/V_g$  for overland cases is  $0.71 * (100\% - 19\%) = 0.575$

**Table 4.1 Example values of wind speed reduction factors [adapted from (Vickery, Masters, et al. 2009)]**

Source	$V_s/V_g$ over water (near eyewall)	Sea-land additional reduction
Schwerdt et al. (1979)	0.950	11%, at the coast 22%, 19 km inland
Batts et al. (1980)	0.865	15%, at the coast
Georgiou (1985)	0.825 (near the eyewall)	0%, at the coast
	0.750 (away from the eyewall)	25%, 50 km inland
Vickery et al. (2000)	~ 0.70 - 0.72	14-20%, at the coast
		23-28%, 50 km inland
Sparks and Huang (2001)	0.650	30%, a few km inland
Powell et al. (2005)	~ 0.73	15-20%, at the coast
Powell et al. (2003)	~ 0.71	N/A
Vickery et al. (2009)	~ 0.71 (varies from 0.67 to 0.74)	18-20%, at the coast

#### 4.1.3 The GSESM's approach for modelling the parametric wind field

Using the above methodology, the GSESM's approach for the modelling of the parametric wind field is presented as follows:

Firstly, DPs measured after 1977, which are inside the Threat Area (TA) and have the strength of a tropical storm or stronger [i.e., maximum sustained surface wind speed ( $V_{smax}$ ) greater than or equal to 30 kt] are collected from the JMA's BTD, as explained in subsection 2.1.1 and 2.2.2.

Secondly, several DPs also have to be excluded from the calculations, due to their inability to provide the crucial parameters required in the model, including:

- DPs where TCs do not move are excluded, because at the locations the translation speeds are equal to zero, but headings are indefinable. This



exclusion ensures the consistency in the input data for the modelling of the parametric wind field and other parts of the research (e.g., the simulation of central track and intensity addressed in chapter 3). It is called limitation 1.

- DPs at the first time step of each TC are also removed from the calculations. The reason is that, at the first time step, the latitude and longitude at the previous time step are indefinable. Therefore, the translation speed and heading at that time step, which are needed in various equations, cannot be obtained. It is called limitation 2.

Table 4.2 gives the details of available DPs. As can be seen from the table, among 17,935 DPs that were recorded in the WNP since 1977, 6,147 DPs are usable for the modelling of the parametric wind field. Moreover, because there are 238 DPs inside the TA that are removed due to the two above mentioned limitations, an analysis of those DPs is necessary to determine whether the elimination has any significant effect on the model or not. Table 4.3 presents this examination, in which DPs are categorized according to their types defined by the JMA. The table shows that most of the 238 DPs (i.e., 224 DPs, which account for 94.12% of the total) have the strength of a tropical depression or weaker (i.e.,  $V_{max}$  smaller than 30 kt, denoted by types 2 and 6 in the table) and therefore have no impact on the results. The model only loses 14 DPs (denoted by types 3, 4, and 5 in the table), which are very small (i.e., less than 0.3%) in comparison with the total number of DPs that are usable for the model (i.e., 6147 DPs).

**Table 4.2 Analysis of historical data (wind field model)**

Number of DPs								
WNP BTD	Removed (model limitations)			Leftover	Outside the TA	Inside the TA		
	No. 1	No. 2	Total			Total		
17935	138	514	652		17283	8246	9037	
			Outside the TA	Inside the TA			Tropical storm or stronger (winds $\geq$ 30 kt)	Tropical depression
			414	238				

**Table 4.3 Exclusion of DPs in the TA (wind field model)**

DP type	Number of DPs	Percentage
1	0	0.00
2	222	93.28
3	9	3.78
4	2	0.84
5	3	1.26
6	2	0.84
7	0	0.00
8	0	0.00
9	0	0.00
Total	238	100

Thirdly, at each DP, variously required parameters are obtained or derived from the BTD, including the maximum sustained surface wind speed ( $V_{smax}$ ), the central pressure deficit [ $\Delta p$ , derived from the central pressure ( $p_c$ ) provided by the BTD and the environmental pressure ( $p_n$ ) defined by the technique presented in chapter 3], the maximum radius of 30 kt wind ( $R_{max30}$ ), and the translation speed ( $c$ , derived from the locations of TC's centre at current and previous time steps).

The gradient wind speed corresponding to the 30 kt (i.e.,  $15.433 \text{ m s}^{-1}$ ) and the maximum wind speed at surface-level are estimated by using the wind speed conversion factors depending on the current location (i.e.,  $V_s/V_g$  is taken to be 0.71 and 0.575 for the over water and overland cases, respectively). This means:

$$V_{g30} = 15.433 / (V_s/V_g)$$

$$V_{gmax} = V_{smax} / (V_s/V_g)$$

Where:  $V_{g30}$  - gradient wind speed ( $\text{m s}^{-1}$ ) corresponding to a 30 kt surface wind.

The values of  $B$  and  $R_{max}$  at current DP can then be computed by iterations:

- In the first iteration, a guess (initial value) of  $B$  (i.e.,  $B_1$ ) is first estimated by neglecting the Coriolis effect (i.e.,  $f = 0$ ) in equation 4.13, as follows:

$$V_{gmax} = \frac{1}{2}c + \sqrt{\frac{1}{4}c^2 + \frac{B_1 \Delta p}{\rho_a e}} \rightarrow B_1 = \left[ \left( V_{gmax} - \frac{1}{2}c \right)^2 - \frac{1}{4}c^2 \right] \frac{\rho_a e}{\Delta p} \quad (4.14)$$

- The value of  $R_{max}$  associated with  $B_1$  (i.e.,  $R_{max1}$ ) is estimated by using formula 4.12, providing that  $r = R_{max30} \rightarrow V_{gr} = V_{g30}$ . This means that  $R_{max1}$  is obtained from:

$$V_{g30} = \frac{1}{2}(c - f R_{max30}) + \sqrt{\frac{1}{4}(c - f R_{max30})^2 + \frac{B_1 \Delta p}{\rho_a} \left( \frac{R_{max1}}{R_{max30}} \right)^{B_1} \exp \left[ - \left( \frac{R_{max1}}{R_{max30}} \right)^{B_1} \right]} \quad (4.15)$$

- The new value of  $B$  corresponding to  $R_{max1}$  (i.e.,  $B_2$ ) is determined by using equation 4.13, with  $r = R_{max1} \rightarrow V_{gr} = V_{gmax}$ . This means that  $B_2$  is defined from:

$$V_{gmax} = \frac{1}{2}(c - f R_{max1}) + \sqrt{\frac{1}{4}(c - f R_{max1})^2 + \frac{B_2 \Delta p}{\rho_a e}} \quad (4.16)$$

- In the subsequent iterations, the starting value  $B_1$  of an iteration is set to the ending value  $B_2$  of the previous iteration and the process continues, using the equations 4.15 and 4.16, until the values of  $B$  converge (i.e.,  $B_1 \approx B_2$ ) within a specified tolerance, which is selected to be 0.0001 in this research.

Note that, the above iteration procedure does not always converge to a solution since errors are caused by the inconsistency between the mathematical approximations of TC winds, as could be seen in equations 4.15 and 4.16, and the physical constraints. The realistic limits are obtained by logical inference or stated in various wind field models, as follows:

- Since the DPs are cut off by the 30 kt wind speed threshold, the value of  $R_{max}$  (associated with  $V_{gmax}$ ) must be smaller than  $R_{max30}$  (corresponding to the

30 kt wind). The violation of such a limit (i.e., produce  $R_{\max 1} \geq R_{\max 30}$  in formula 4.15) is called error 1.

- According to various wind field estimates,  $B$  can only vary within a certain range. While in Holland's (1980) approach, a margin of about 0.5 - 2.5 was employed, other studies restricted the value of  $B$  from about 0.7 to 2.2 due to the suggestions derived from the analyses of DPs in their specific research areas [e.g., (Powell et al. 2005; Vickery and Wadhera 2008; Willoughby and Rahn 2004)]. To facilitate the general model that can be applied at any location, a wide range of  $B$  (i.e., from 0.5 to 2.5) is selected. In the case that the equation 4.16 results in  $B$  outside this margin, the calculation falls into the error 2 category.

Table 4.4 shows that among the total of 6,147 DPs, there are 5,071 DPs (i.e., accounting for 82.5%) have valid wind field computations.

**Table 4.4 Wind field calculations**

Number of DPs				
Input	Removed (errors)			Converge to a solution
	No. 1	No. 2	Total	
6147	796	280	1076	5071

Finally, the specifications of the surface wind and atmospheric pressure fields (i.e., wind and pressure at every single point within the vicinity of the TC at the current time step) are defined by using formula 4.4 (for surface pressures) and equation 4.12 coupled with simple wind speed reduction factors  $V_s/V_g$  (for surface wind).

Figure 4.3 provides the flow chart of the step-by-step approach used in this study.

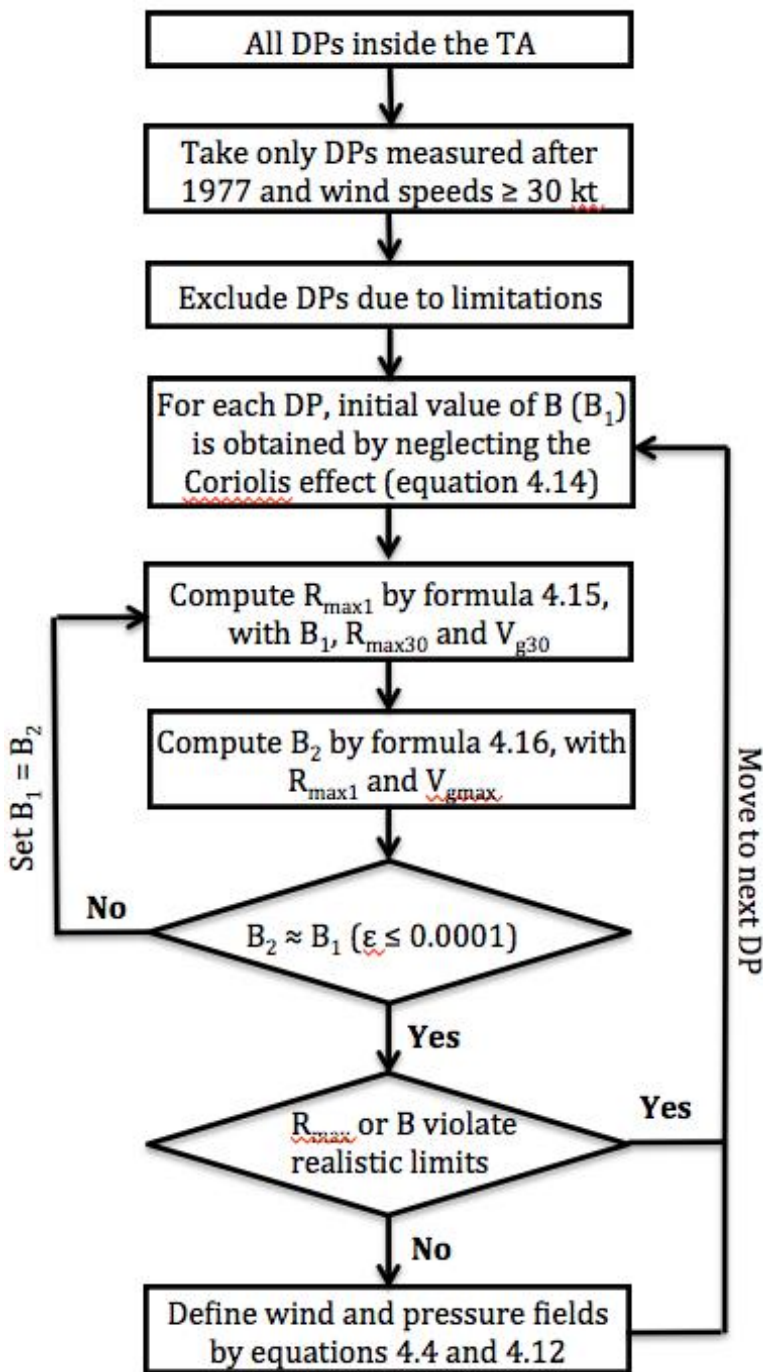


Figure 4.3 Flow chart (wind field model)

In a complete risk-related research, an important stage is the validation to ensure the ability of the parametric wind field model utilized in the simulation to reproduce wind speed directly measured on site. However, as a result of the various difficulties presented in subsection 3.4.1, the verification has not been included or has been carried out rather restrictedly in most studies. The parameters, chosen as the subjects of the comparisons between modelled and observed data, are inconsistent. While some methods provided the validation of

the surface wind speeds and the wind directions [e.g., (Georgiou 1985; Harper 1999; Lee and Rosowsky 2007; Vickery, Skerlj, Steckley, et al. 2000; Vickery and Twisdale 1995a)], others also performed an additional verification of surface pressure [e.g., (McConochie et al. 2004; Vickery, Wadhera, Powell, et al. 2009)] to prove the ability of the model in reproducing both wind and pressure fields. However, in cases where there is significant scatter of field data like the Vietnamese coastline and the nearby ocean zone in this study, the parametric wind field model, which already proved to be a reasonable approach for practical applications in many areas, is employed "as is" without any detailed validations. Obviously, if reliable field measurements are available as a result of future research, it will be preferable to utilize these data to carry out the verifications of this parametric method.

## 4.2 Radius of maximum wind modelling

### 4.2.1 Existing methodologies

Considering the above approach for the modelling of the parametric wind field, as demonstrated in equation 4.13, the radius of maximum wind ( $R_{max}$ ) has very little or even no influence (in case of neglecting the Coriolis factor) on the estimation of the maximum wind speed. However, it has a remarkable effect on the probability calculations in cases of near misses. Since a TC spreads its effect over a broad region, a specific location, which is not directly hit by this TC (i.e., is not crossed by the track connecting the central points), still experiences strong winds that can cause significant losses. Furthermore,  $R_{max}$  is also an important input for the storm surge and wave modelling (Vickery, Masters, et al. 2009). Thus, such a parameter must be defined in any risk-related studies, such as the long-term probability calculations for risk assessments or the estimations of potential damage for insurance purposes.

Based on the knowledge derived from past research in the literature, the value of  $R_{max}$  is generally considered to correlate with two factors (Vickery, Masters, et al. 2009). The first element is the central pressure deficit ( $\Delta p$ ). The negative correlation between  $R_{max}$  and  $\Delta p$  means that a weaker TC (i.e., smaller  $\Delta p$ ) has a larger  $R_{max}$  than the stronger one.  $R_{max}$  also decreases when the latitude  $\psi$  drops. These relationships are proven and quantified by Vickery and Twisdale (1995b), in which the observations of  $R_{max}$ ,  $\Delta p$ , and  $\psi$  yield a positive correlation coefficient of 0.47 between  $R_{max}$  and  $\psi$ , and a negative correlation coefficient of -0.23 between  $R_{max}$  and  $\Delta p$ . Unfortunately, the measurements of  $R_{max}$  used in the Vickery and Twisdale's (1995b) study are obtained from Ho et al. (1987), which are only available for the Atlantic basin, and normally are not included in TC records at other locations [e.g., the Australian coast (James and Mason 2005)]. Thus, in most studies, the above notion is employed as general knowledge and  $R_{max}$  is computed by using an equation consisting of  $\psi$  and/or  $\Delta p$ .

One point to keep in mind is that, in a model where a large domain is considered, the modelling of  $R_{max}$  often separates data into different groups based on latitude, which requires the arbitrary definition of levels for each subset [e.g., 30° North in

Vickery et al.'s (2000) technique]. Therefore, another benefit of the use of the regional domain in this study is that it does not require such subjective division.

#### 4.2.2 The GSESM's method for modelling the radius of maximum wind

Following the suggestions provided in other research [e.g., (Powell et al. 2005; Vickery, Skerlj, and Twisdale 2000; Vickery and Twisdale 1995b; Vickery and Wadhwa 2008)],  $R_{max}$  is modelled as log-normally distributed. The equation for estimating the median value of the distribution is a combination of different terms used in the literature, which may have effects on  $R_{max}$ . Thus, the general form of a formula for modelling  $R_{max}$  is:

$$\ln R_{max} = e_1 + \{e_2 \psi\} + \{e_3 \psi^2\} + \{e_4 \Delta p\} + \{e_5 \Delta p^2\} + \varepsilon$$

Where:  $e_1, e_2$ , etc. – the coefficients;  $\varepsilon$  – the random error term. The brackets in each term indicate whether that term is included in the final equation or not, depending on the subsequent analysis.

The GSESM's approach improves the ones employed in other research as follows:

- First, all possible combinations of available terms are considered to determine the optimal one (i.e., largest adjusted  $R^2$ , see paragraph 3.3.3.1). Following the idea provided in paragraph 3.3.4.3, results are given in Table 4.5, where each term is represented by its coefficient. For instance, " $e_2$ " is the representative of the " $e_2 \psi$ " term. For the case study of Vietnam, the table shows that the scenario with the largest adjusted  $R^2$  is option 10 and therefore the equation for modelling  $R_{max}$  is provided as follows:

$$\ln R_{max} = e_1 + e_2 \psi + e_3 \psi^2 + e_4 \Delta p + e_5 \Delta p^2 + \varepsilon \quad (4.17)$$

**Table 4.5 Combinations of available terms (radius of maximum wind model)**

Combi	$e_1; e_2$	$e_1; e_3$	$e_1; e_4$	$e_1; e_5$	$e_1; e_2; e_4$	$e_1; e_2; e_5$	$e_1; e_2; e_4; e_5$	$e_1; e_2; e_3; e_4$	$e_1; e_2; e_3; e_5$	$e_1; e_2; e_3; e_4; e_5$	$e_1; e_3; e_4; e_5$	$e_1; e_3; e_4$	$e_1; e_3; e_5$	$e_1; e_4; e_5$
Combi num	1	2	3	4	5	6	7	8	9	<b>10</b>	11	12	13	14
Adj $R^2$	0.0837	0.0796	0.0200	0.0084	0.0902	0.0851	0.1120	0.0904	0.0854	<b>0.1123</b>	0.1086	0.0866	0.0812	0.0498
$R^2$	0.0839	0.0797	0.0202	0.0086	0.0906	0.0855	0.1125	0.0910	0.0859	<b>0.1130</b>	0.1091	0.0869	0.0816	0.0502

- Secondly, different linear regression solutions for equation 4.17 are also considered in the GSESM, as shown in Table 4.6 for the case study of Vietnam, including the Ordinary Least Squares (OLS) and various robust regression approaches. As can be seen in the table, the solution that provides the best performance (i.e., largest  $R^2$ , see paragraph 3.3.3.1) for equation 4.17 is the "talwar" robust regression.

**Table 4.6 Options for regression solution (radius of maximum wind model)**

Method	OLS	Robust 'andrews'	Robust 'bisquare'	Robust 'cauchy'	Robust 'fair'	Robust 'huber'	Robust 'logistic'	<b>Robust 'talwar'</b>	Robust 'welsch'
$R^2$	<b>0.1130</b>	0.1098	0.1098	0.1089	0.1075	0.1100	0.1085	<b>0.1230</b>	0.1095

Table 4.7 shows a comparison between the GSESM's formula and the empirical equations for the modelling of  $R_{max}$  used in other research, including:

Formulae 4.18 and 4.19 (Vickery and Twisdale 1995b):

$$\ln R_{max} = 3.853 - 0.0061 \Delta p \quad (4.18)$$

$$\ln R_{max} = 2.395 + 0.0426 \psi \quad (4.19)$$

Formulae 4.20, 4.21, 4.22, 4.23, and 4.24 (Vickery, Skerlj, and Twisdale 2000):

$$\ln R_{max} = 3.919 - 0.00737 \Delta p \quad (4.20)$$

$$\ln R_{max} = 2.569 + 0.037842 \psi \quad (4.21)$$

$$\ln R_{max} = 2.636 - 0.00005086 \Delta p^2 + 0.0394899 \psi \quad (4.22)$$

$$\ln R_{max} = 2.097 + 0.0187793 \Delta p - 0.00018672 \Delta p^2 + 0.0381328 \psi \quad (4.23)$$

$$\ln R_{max} = 2.713 - 0.0056748 \Delta p + 0.0416289 \psi \quad (4.24)$$

Formula 4.25 (Powell et al. 2005):

$$\ln R_{max} = 2.0633 + 0.0182 \Delta p - 0.00019008 \Delta p^2 + 0.0007336 \psi^2 \quad (4.25)$$

Formulae 4.26, 4.27, and 4.28 (Vickery and Wadhera 2008):

$$\ln R_{max} = 3.015 - 6.29 \cdot 10^{-5} \Delta p^2 + 0.0337 \psi \quad (4.26)$$

$$\ln R_{max} = 3.858 - 7.7 \cdot 10^{-5} \Delta p^2 \quad (4.27)$$

$$\ln R_{max} = 3.421 - 4.6 \cdot 10^{-5} \Delta p^2 + 0.00062 \psi^2 \quad (4.28)$$

The results shown in the table are various goodness-of-fit measures. They confirm that the GSESM's approach is definitely a considerable improvement over the techniques employed in other studies.

**Table 4.7 Comparison between the GSESM' approach and various empirical formulae for the case study of Vietnam (radius of maximum wind model)**

Goodness-of-fit measure <sup>21</sup>	<b>GSESM (formula 4.17)</b>	Formula 4.18	Formula 4.19	Formula 4.20	Formula 4.21	Formula 4.22	Formula 4.23	Formula 4.24	Formula 4.25	Formula 4.26	Formula 4.27	Formula 4.28
1	0.2411	0.5038	1.0611	0.4957	0.9123	0.9026	1.0748	0.9168	2.2379	0.5681	0.4287	0.5280
2	0.8871	1.8541	3.9047	1.8239	3.3572	3.3215	3.9551	3.3738	8.2353	2.0905	1.5777	1.9428
3	0.4910	0.7098	1.0301	0.7040	0.9551	0.9501	1.0367	0.9575	1.4960	0.7537	0.6548	0.7266
4	0.9419	1.3616	1.9760	1.3505	1.8323	1.8225	1.9887	1.8368	2.8697	1.4459	1.2561	1.3939
5	0.3924	0.5720	0.9213	0.5650	0.8421	0.8341	0.9301	0.8402	1.4150	0.6278	0.5200	0.5994
6	0.1000	0.1342	0.2166	0.1330	0.1973	0.1955	0.2189	0.1970	0.3374	0.1470	0.1234	0.1401
7	0.3507	-0.1422	0.2896	-0.1422	0.2896	0.2308	0.3319	0.1993	0.3196	0.1910	-0.0926	0.1763
8	0.1230	0.0202	0.0839	0.0202	0.0839	0.0533	0.1102	0.0397	0.1021	0.0365	0.0086	0.0311
9	0.1127	-0.8544	-2.9054	-0.8243	-2.3579	-2.3222	-2.9559	-2.3745	-7.2369	-1.0910	-0.5780	-0.9432
10	2.2424	2.1179	2.5867	2.1044	2.4950	2.4865	2.5048	2.5201	3.0083	2.2282	2.0858	2.1908
11	1.1965	0.9539	0.6893	0.9679	0.7362	0.7612	0.7032	0.7548	0.6376	0.9012	1.0182	0.9096

## 4.3 Holland parameter model

### 4.3.1 Existing methods

As can be seen in formula 4.9, the Holland parameter ( $B$ ) is an important factor in the calculation of the maximum wind speed. Thus, there is an obvious demand for a technique that is able to estimate the critical element based on the obtainable parameters. One of the first attempts for modelling  $B$  as a function of other elements was introduced by Harper and Holland (1999), where  $B$  was derived as:

$$B = 2.0 - (p_c - 900)/160$$

This means that  $B$  decreases when the central pressure ( $p_c$ ) increases [i.e., central pressure deficit ( $\Delta p$ ) decreases].

Vickery et al. (2000) confirmed the above relationship, so that as  $\Delta p$  decreases,  $B$  decreases. Moreover, the negative correlation between  $B$  and radius of maximum wind ( $R_{max}$ ) was also revealed.

In contrast, using the values of  $B$  estimated by Willoughby and Rahn (2004), and the flight-level measurements, Powell et al. (2005) found no relationship between

<sup>21</sup> List of goodness-of-fit measures presented in the table (denoted by number):

- 1 - mean squared error
- 2 - normalised mean squared error
- 3 - root mean squared error
- 4 - normalised root mean squared error
- 5 - mean absolute error
- 6 - mean absolute relative error
- 7 - coefficient of correlation
- 8 - coefficient of determination
- 9 - coefficient of efficiency
- 10 - maximum absolute error
- 11 - maximum absolute relative error



$B$  and  $\Delta p$ . However, the negative correlation with  $R_{max}$  still remained unchanged, and  $B$  was modelled as a linear function of  $R_{max}$  and the latitude  $\psi$ .

One most recent research was carried out by Vickery and Wadhwa (2008), in which  $B$  was obtained from the same aircraft observations as used by Willoughby and Rahn (2004) and Willoughby et al. (2006). Another factor was introduced, which represented the effects of  $R_{max}$ ,  $\Delta p$  as well as  $\psi$  [through the use of the Coriolis parameter ( $f$ )] and the sea surface temperature ( $T_s$ ). Parameter  $A$  is defined from:

$$A = \frac{R_{max} f}{\sqrt{2 R_d T_s \ln\left(1 + \frac{\Delta p}{p_c e}\right)}}$$

Where:  $R_d$  - gas constant for dry air, taken to be 287.058 (J kg<sup>-1</sup> K<sup>-1</sup>)

Since both the denominator and numerator of  $A$  have units of velocity,  $A$  is non-dimensional, and  $B$  was estimated by a function of  $A$ .

According to Vickery et al. (2009), the modelling of variable  $B$  is a remarkable improvement over the approach used in the research of the earlier days [e.g., (Georgiou 1985; Vickery and Twisdale 1995a; b)], which constrained  $B$  to have a constant value of unity (i.e.,  $B = 1$ ). However, at each time step, this modelling technique can only produce a single value of  $B$ , which is incapable of reproducing the actual wind fields in some cases, as already discussed in the recent studies [e.g., (Thompson and Cardone 1996; Willoughby et al. 2006; Willoughby and Rahn 2004)]. Luckily, such an inability of a single parameter model only appears in very few circumstances [e.g., Hurricane Wilma (2005)], and the model still provides a reasonable approximation, which can be included in the simulation process. Thus, the simple and widely acceptable linear equation is employed in this study for the modelling of  $B$ .

#### 4.3.2 The GSESM's method for the modelling of the Holland parameter

Following the concept for the estimation of  $R_{max}$  provided in the subsection 4.2.2, the equation for the modelling of  $B$  is a combination of different terms, which may have effects on  $B$ , as indicated in the previous subsection, including  $\Delta p$ ,  $R_{max}$ ,  $\psi$ , and  $A$ . Thus, the general form of a formula for the modelling of  $B$  is:

$$B = f_1 + \{f_2 R_{max}\} + \{f_3 \psi\} + \{f_4 \Delta p\} + \{f_5 A\} + \varepsilon$$

Where:  $f_1, f_2$ , etc. - the coefficients;  $\varepsilon$  - the random error term. The brackets enclosing each term mean that the term can be included or excluded in the final equation, depending on the subsequent analysis.

The GSESM's approach improves the approaches employed in previous research as follows:

- First, all possible combinations of available terms are considered to determine the optimal one (i.e., largest adjusted  $R^2$ , see paragraph 3.3.3.1). Following the idea provided in paragraph 3.3.4.3, results are given in Table 4.8, where each term is represented by its coefficient. For instance, " $f_2$ " is the representative of the " $f_2 R_{max}$ " term. For the case study of Vietnam, the

table shows that the scenario with the largest adjusted  $R^2$  is option 8. Therefore, the equation for the modelling of B is provided as follows:

$$B = f_1 + f_2 R_{\max} + f_3 \psi + f_4 \Delta p + \varepsilon \quad (4.29)$$

**Table 4.8 Combinations of available terms (Holland parameter model)**

Combination	$f_1; f_2$	$f_1; f_3$	$f_1; f_4$	$f_1; f_5$	$f_1; f_2; f_3$	$f_1; f_2; f_4$	$f_1; f_3; f_4$	$f_1; f_2; f_3; f_4$
Combi num	1	2	3	4	5	6	7	<b>8</b>
Adjusted $R^2$	0.2954	0.1675	0.0996	0.2622	0.3647	0.3569	0.2203	<b>0.4048</b>
$R^2$	0.2955	0.1676	0.0997	0.2624	0.3649	0.3572	0.2207	<b>0.4051</b>

- Secondly, different linear regression solutions for equation 4.29 are also considered in the GSESM, as shown in Table 4.9 for the case study of Vietnam, including the OLS and various robust regression approaches. As can be seen from the table, the solution that provides the best performance (i.e., largest  $R^2$ , see paragraph 3.3.3.1) for equation 4.29 is the "talwar" robust regression.

**Table 4.9 Options for a regression solution (Holland parameter model)**

Method	OLS	Robust 'andrews'	Robust 'bisquare'	Robust 'cauchy'	Robust 'fair'	Robust 'huber'	Robust 'logistic'	<b>Robust 'talwar'</b>	Robust 'welsch'
$R^2$	<b>0.3654</b>	0.3915	0.3917	0.3931	0.3919	0.3918	0.3928	<b>0.4299</b>	0.3928

Table 4.10 shows a comparison between the GSESM's formula and the empirical equations for the modelling of B as used in other research, including:

Formulae 4.30 and 4.31 (Vickery, Skerlj, and Twisdale 2000):

$$B = 1.34 + 0.00328 \Delta p - 0.00522 R_{\max} \quad (4.30)$$

$$B = 1.38 + 0.00184 \Delta p - 0.00309 R_{\max} \quad (4.31)$$

Formula 4.32 (Powell et al. 2005):

$$B = 1.881093 - 0.010917 \psi - 0.005567 R_{\max} \quad (4.32)$$

Formulae 4.33, 4.34, and 4.35 (Vickery and Wadhwa 2008):

$$B = 1.881 - 0.01295 \psi - 0.00557 R_{\max} \quad (4.33)$$

$$B = 1.7642 - 1.2098 \sqrt{A} \quad (4.34)$$

$$B = 1.833 - 0.326 \sqrt{R_{\max} f} \quad (4.35)$$

The results shown in the table are various goodness-of-fit measures. They confirm that the GSESM's approach is definitely a considerable improvement over the techniques employed in other studies.

**Table 4.10 Comparison between the GSESM' approach and various empirical formulae for the case study of Vietnam (Holland parameter model)**

Goodness-of-fit measure <sup>22</sup>	<b>GSESM (formula 4.29)</b>	Formula 4.30	Formula 4.31	Formula 4.32	Formula 4.33	Formula 4.34	Formula 4.35
1	<b>0.0706</b>	0.5596	0.3722	0.4037	0.4398	0.6866	0.6362
2	<b>0.6372</b>	5.0495	3.3586	3.6435	3.9688	6.5258	6.0462
3	<b>0.2657</b>	0.7480	0.6101	0.6354	0.6632	0.8286	0.7976
4	<b>0.7982</b>	2.2471	1.8326	1.9088	1.9922	2.5546	2.4589
5	<b>0.2042</b>	0.6211	0.5026	0.4996	0.5248	0.6981	0.6576
6	<b>0.1269</b>	0.3380	0.2736	0.2767	0.2891	0.3863	0.3629
7	<b>0.6557</b>	-0.4311	-0.4369	-0.5646	-0.5692	-0.5122	-0.6106
8	<b>0.4299</b>	0.1859	0.1909	0.3187	0.3240	0.2624	0.3729
9	<b>0.3627</b>	-4.0505	-2.3592	-2.6442	-2.9696	-5.5271	-5.0475
10	<b>1.0719</b>	2.4591	1.8300	2.3481	2.3843	2.2982	2.4381
11	<b>1.0507</b>	1.0486	1.0310	1.4821	1.4477	1.2043	1.3486

<sup>22</sup> List of goodness-of-fit measures presented in the table (denoted by number):

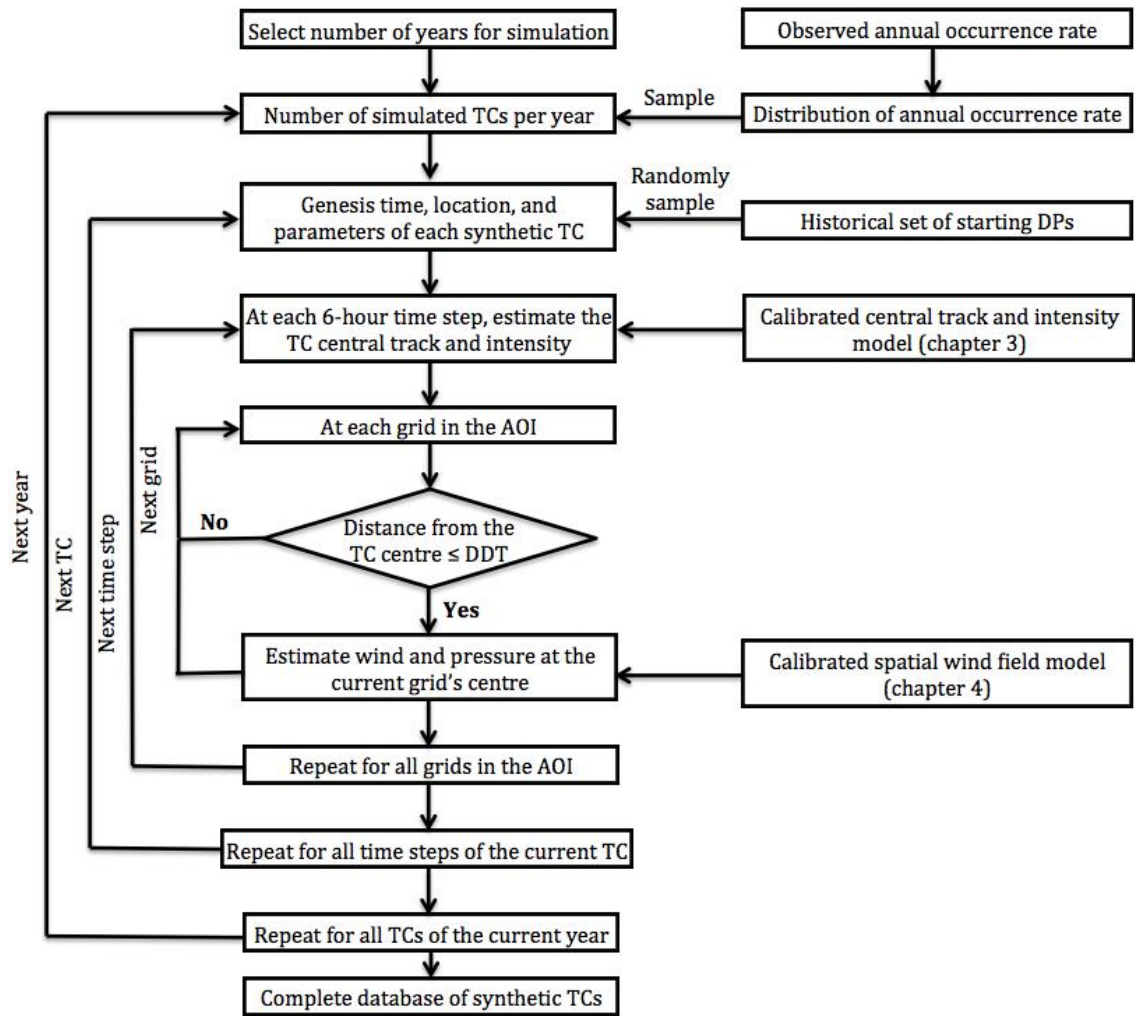
- 1 - mean squared error
- 2 - normalised mean squared error
- 3 - root mean squared error
- 4 - normalised root mean squared error
- 5 - mean absolute error
- 6 - mean absolute relative error
- 7 - coefficient of correlation
- 8 - coefficient of determination
- 9 - coefficient of efficiency
- 10 - maximum absolute error
- 11 - maximum absolute relative error

# 5 SIMULATION AND RESULTS

## Research questions:

- How long should a reasonable length of synthetic Tropical Cyclone (TC) database be?
- Which basic discrete distribution should be employed to approximate the TC annual occurrence rate?
- How to define a set of initial points for TCs in the simulation?
- What is the Damage Distance Threshold (DDT)? How to define the DDT?
- Which physical boundary conditions should be included to introduce the realistic limits of the parameters as well as to defined the lysis of a TC?
- How to validate the model?
- What are the possible applications of the model outcomes?

The final product of this research is a long-term database of synthetic TCs that is specifically developed for the local conditions of the user-defined Area Of Interest (AOI). After specifying the critical model components through the step-by-step optimization processes presented in the previous chapters, the simulation is carried out as described in Figure 5.1. This chapter gives the basic aspects of the operational procedure (section 5.1) and the detailed evaluations of the model results (section 5.2) to prove its ability to reproduce the actual TC characteristics and to generate a useable long-term database with an acceptable accuracy for practical projects. Finally, the design wind speed maps and the annual exceedance probability maps are provided as the possible applications of the model outcomes (section 5.3).



**Figure 5.1 Flow chart (simulation)**

## 5.1 Simulation procedure

### 5.1.1 Number of years

Generally, the methodology benefits from a large number of years used in the model. The reason is that a long-term simulation not only provides sufficient data to approximate the "true" TC statistics, but also increases the possibility of TC occurrence in some places, which may be threatened by extremely infrequent but highly destructive TCs. Although the approach is theoretically capable of implementing a model with unlimited length, a prolonged simulation significantly increases computational demand, which can be a crucial limitation, especially when other computationally demanding studies must be carried out based on model results, or simulations must be repeated many times with various alternatives. The issue of a reasonable model timescale has not been addressed in the literature, which led to the arbitrary and subjective choices of number of years

used in the past researches, ranging from 10,000 years (Wang and Rosowsky 2012); 15,000 years (Lee and Rosowsky 2007); 20,000 years (Vickery, Skerlj, and Twisdale 2000) to a high of 100,000 years (Vickery, Wadhera, Twisdale, et al. 2009).

In most cases, the main interest centres on the design levels (i.e., TC characteristics) with acceptable exceedance probabilities within the typical lifespans of practical engineering projects (e.g., 20, 50, or 100 years). Therefore, the questions are:

- Whether the relatively short-term evaluations really benefit from a simulation with a much longer period?
- What is the optimal timescale that balances the computational demand and the model performance?

Those questions will be dealt later in section 5.2, where a period of 5,000 years will be defined to be the reasonable length of TC simulation in this study.

### 5.1.2 Annual occurrence rate

The next step is to determine the number of TCs to be simulated in any year. Basically, this quantity is obtained by sampling a distribution, which is defined based on the historical annual occurrence rate of TCs in the research area. The selection of that distribution was ambiguous and inconsistent among different studies. For instance, while some researchers employed the negative binomial distribution [e.g., (Vickery, Skerlj, and Twisdale 2000)], others utilized the binomial [e.g., (Vickery, Wadhera, Twisdale, et al. 2009)] and the Poisson processes [e.g., (Wang and Rosowsky 2012)]. Powell et al. (2005) examined both the negative binomial and the Poisson distribution to decide which one would best fit the observed data set, based on the results from a goodness-of-fit test (i.e., chi-square test).

In this research, all of the three basic discrete distributions (i.e., the negative binomial, binomial, and Poisson distributions) are tested to the historical annual occurrence rate. Instead of relying only on the ordinary chi-square test, several fit indices are used to determine the best-fit distribution. These criteria are the major ones, currently employed in the literature, including:

- The Bayesian Information Criterion (BIC)
- The Akaike Information Criterion (AIC)
- The AIC with a correction for finite sample sizes (AICc)
- The Negative of the Log Likelihood (NLogL)

The reason for using the alternative measures of fit is the considerable disadvantage of the chi-square test (McDonald 2009). Firstly, its high sensitivity to the sample size makes the test reasonable only for models with less than about 200 cases. For models with more cases, since the chi-square statistics automatically increase with the increase of the sample size, the test is almost always statistically significant. Secondly, if a small expected frequency is observed

in any category, which is exactly the case of the historical annual occurrence rate, the chi-square test may also give unreliable outcomes.

Furthermore, a chi-square test is an absolute fit index, which does not give much information about the strength of the relationship between the observed and modelled data. On the other hand, other criteria (e.g., the BIC or AIC) are comparative indices that provide a means to directly compare two estimated fits. Using these techniques, the distribution with the lower value of BIC (or AIC) is the one to be preferred, and the strength of the evidence against the estimates with the higher value can be summarized in Table 5.1 and Table 5.2 for the BIC and AIC difference criteria, respectively.

**Table 5.1 BIC difference criterion [adapted from (Kass and Raftery 1995)]**

BIC difference	Evidence against higher BIC
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
> 10	Very strong

**Table 5.2 AIC difference criterion [adapted from (Burnham and Anderson 2013)]**

AIC difference	Support for equivalency of fits
0 to 2	Substantial
4 to 7	Weak
> 10	None

For the case study of Vietnam, an analysis of the observed annual occurrence rate can be summarized as follows:

- Because the data has a variance (i.e., 6.3 TCs/year) smaller than its mean (i.e., 11.9 TCs/year), the negative binominal distribution is unusable.
- The binominal distribution requires a predefined parameter, which is the number of trials ( $N$ ). For a primary analysis,  $N$  is estimated by the Method of Moments and the corresponding binominal fit is called "Binominal 1".
- Figure 5.2 and Figure 5.3 give the comparisons of the Probability Mass Function (PMF) and the Cumulative Distribution Function (CDF) between the "Binominal 1" and the Poisson fits. As can be seen from the figures, the Poisson fit provides a worse approximation.

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015

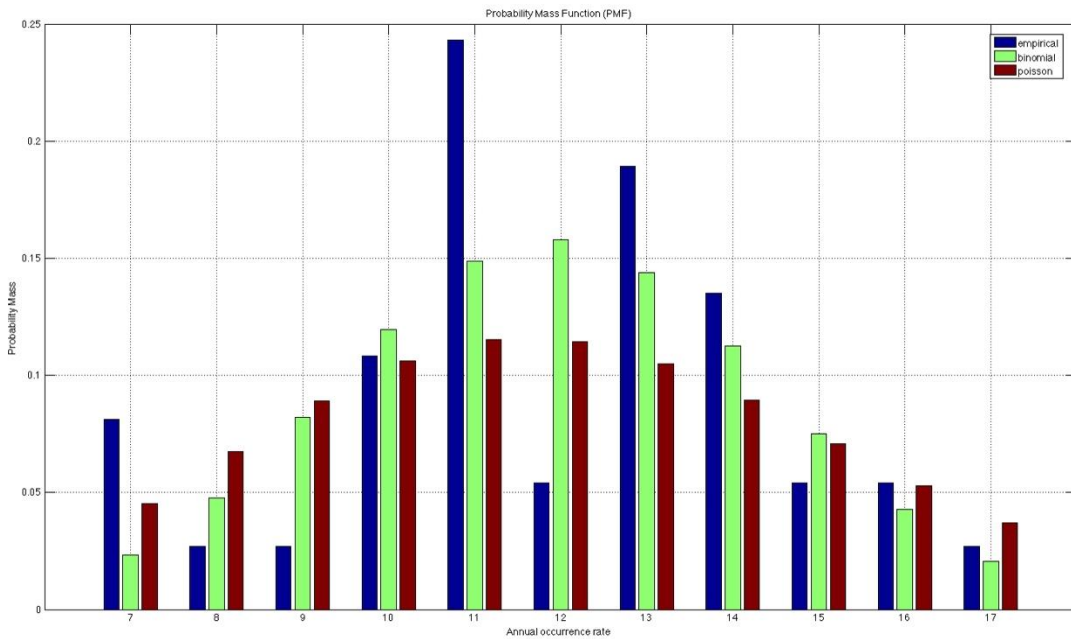


Figure 5.2 Comparison of PMF (annual occurrence rate) <sup>23</sup>

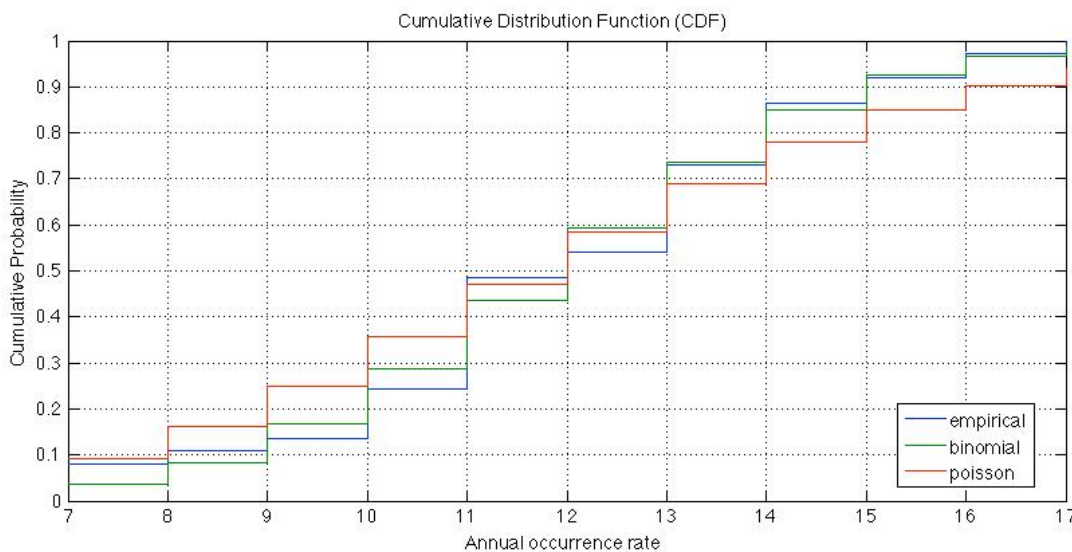


Figure 5.3 Comparison of CDF (annual occurrence rate)<sup>23</sup>

Although the "Binominal 1" fit performs better than the Poisson one, it has a critical limitation [i.e., its restricted number of trials ( $N$ )] that makes it unsuitable for this research. While the Poisson distribution allows an infinite  $N$ , this quantity is limited to a predefined value in the binominal distribution. Therefore, another binomial fit, which contains the advantages of both the binominal (i.e., giving

<sup>23</sup> The number of trials of the binominal fit shown in the figure (i.e., the "Binominal 1" distribution) is estimated by the Method of Moments.



better approximation) and the Poisson distributions (i.e., allowing for a large  $N$ ), should be formulated. This fit is called "Binominal 2" and is determined as follows:

Firstly, the maximum possible value of the number of years for simulation ( $n$ ) is defined. For current and future research, this quantity can only reach a high of 100,000 years. The reason is that, the model does not benefit much for value of  $n$  is greater than 5,000 years, as will be proven in section 5.2. Therefore,  $n_{max}$  is taken to be 100,000 years, the maximum value in the literature.

Secondly, the number of trials corresponding to  $n_{max}$  (i.e.,  $N_{max}$ ) is defined using the Chernoff bound argument (Mitzenmacher and Upfal 2005) of the Poisson distribution that fits to the data. Using this approach, the upper bound for the tail probabilities of a Poisson random variable  $N \sim Pois(\lambda)$  is:

$$P(N \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}, \text{ for } x > \lambda \quad (5.1)$$

Where:  $e$  - the base of the natural logarithm;  $\lambda$  is the parameter of the Poisson fit and equal to the mean value of the data.

In this study, the data is the historical annual occurrence rate. Thus:

- $\lambda = 11.9$  TCs/years.
- $N_{max}$  is the maximum possible value of the number of TCs that may occur in any given year when sampling for  $n_{max}$  years.

Providing that  $x = N_{max}$ , the left-hand side of equation 5.1 is the annual exceedance probability ( $P_a$ ) of an event with the return period of  $n_{max}$ . In this case, the formula (5.1) becomes:

$$P_a = \frac{e^{-\lambda}(e\lambda)^{N_{max}}}{N_{max}^{N_{max}}} \quad (5.2)$$

Furthermore, if the return period of an event is  $n_{max}$  years, the probability of an event occurring, ( $p$ ) can be obtained from:

$$p = 1/n_{max}$$

The probability of occurrence of this event for  $r$  times in a period of  $t$  years is:

$$P_t = (t/r) p^r (1-p)^{t-r}$$

Therefore, the annual exceedance probability (i.e., when  $r = 1$  and  $t = 1$ ) is:

$$P_a = p = 1/n_{max} \quad (5.3)$$

Substituting (5.3) into (5.2), one can determine  $N_{max}$  from:

$$1/n_{max} = \frac{e^{-\lambda}(e\lambda)^{N_{max}}}{N_{max}^{N_{max}}} \quad (5.4)$$

For  $n_{max} = 100,000$  years and  $\lambda = 11.9$  TCs/years, (5.4) results in  $N_{max} = 32$ .

Finally, a binominal distribution with number of trials equal to  $N_{max}$  is fitted to the observed annual occurrence rate. This fit is called "Binominal 2".

Table 5.3 gives a comparison of the "Binominal 2" and the Poisson fits. The table shows that the "Binominal 2" fit not only provides a considerably better performance than the Poisson one, but also still allows for a large number of TCs per year, which may occur in the simulation.

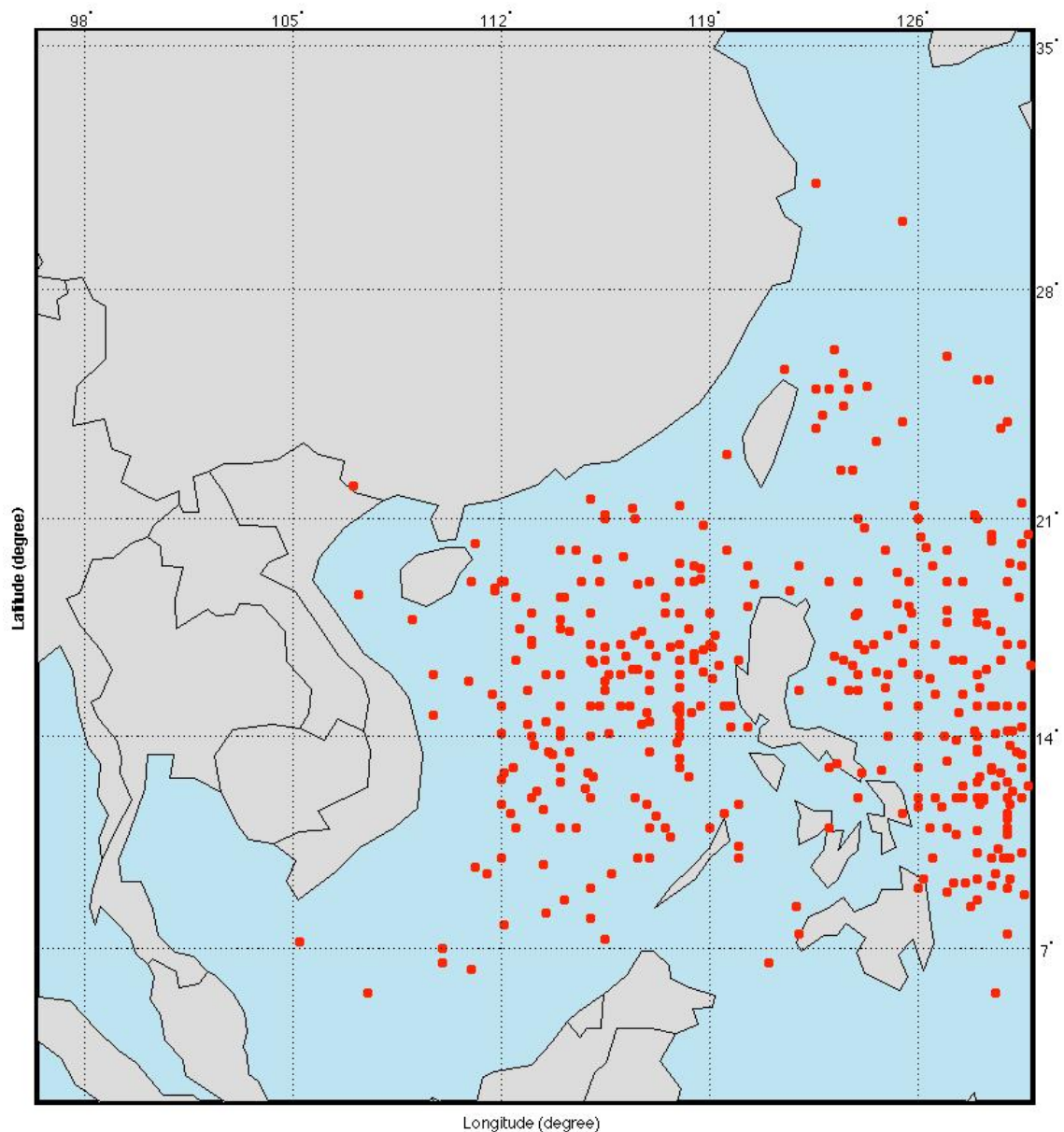
**Table 5.3 Comparison of fit indices (annual occurrence rate)**

Fit	Poisson	Binominal 2 N = 32 (defined by formula 5.4)	BIC (see table 5.1)		AIC (see table 5.2)	
			Difference	Evidence against higher BIC	Difference	Support for equivalency of fits
BIC	182.63	180.12	2.51	Positive	4.12	Weak
AIC	181.02	176.90				
AICc	181.13	177.25				
NLogL	89.51	86.45				

### 5.1.3 Model initialization

For each synthetic TC, the initial specifications (i.e., location, time, and all other key parameters) of the first time step are sampled from a set of starting conditions. Generally, there are two approaches to obtain this set. The first method is a purely empirical one, in which the initial specifications of the historical TCs are used directly. The significance of this approach is that it retains the TC characteristics corresponding to any seasonal preferences of the starting point (Vickery, Skerlj, and Twisdale 2000). Although the technique is simple and was widely employed in many studies [e.g., (Lee and Rosowsky 2007; Vickery, Wadhera, Twisdale, et al. 2009; Wang and Rosowsky 2012)], it also restricts the initial specifications of the simulated TCs to only a minority of starting conditions that could arise. This limitation results in some attempts [e.g., (James and Mason 2005; Powell et al. 2005)] on developing an extended set of initial specifications to generate TCs in the simulation. However, the theoretical framework for enlarging the observed data is so complex that it is only applicable to an oversimplified model, especially when the correlation among the TC key parameters must be preserved. For instance, although a quite sophisticated method (Scheffner et al. 1996) was employed in a long-term model by James and Mason (2005), it can extend the starting conditions with only 6 TC characteristics. In this study, the initial specifications contain nearly 20 parameters, making the establishment of an expanded database impossible.

Therefore, a purely empirical approach must be utilized in this study. However, a set of Data Points (DPs) when the historical TCs first entered the Threat Area (TA), instead of the actual first DPs, is used to start the synthetic TCs in the model. That is, because the majority of observed TCs originating outside the TA did not enter this region and only the DPs inside the TA can be expected to have effects on the AOI, as already indicated earlier in subsection 2.2.2. Thus, the exclusive origination within the TA not only significantly reduces the computing time and power, but also considerably has a better control over the starting conditions of the simulated TCs. Figure 5.4 shows the locations of the historical DPs, which are useable for the model initialization in this research.



**Figure 5.4 Starting positions of observed TCs**

#### 5.1.4 Track and central intensity development

Once a synthetic TC is given the starting conditions, it then propagates in space along a trajectory defined by the track model. All the details of this model are calibrated with the specific case study to achieve the most suitable approach described in section 3.3.

Regarding the central intensity, this specification also changes over time when a TC moves along its track. Depending on the eye's location, the central intensity is estimated by two different models that are also adjusted through the stepwise calibration procedures. When the TC is over water, the relative intensity method provided in section 3.4 is employed to simulate the intensity development. Once the TC makes landfall, it gradually loses its source of energy (Miller 1964),

resulting in a decreasing intensity, associated with an increase in central pressure (Powell et al. 2005). Therefore, a separated model given in section 3.5 must be used to describe the decay of the inland DPs, after a TC makes landfall. The outcome of the track and central intensity models is a full trajectory of the simulated TC, together with the specification of its central location and strength.

One point to keep in mind is that, at every 6-hour time step, the deterministic part of each formula used to simulate the evolution of the track (i.e., equations 3.6 and 3.7) and the central intensity for an over water case (i.e., equation 3.8) is computed first. The error term is then independently and randomly sampled from a normal distribution, which is assumed to fit the residuals of the linear regression using historical data in the corresponding formula. Given conditions at the current time step, these two quantities (i.e., the deterministic part and the random error term) are added up to define the position and central intensity at the subsequent interval. For inland locations, where the TCs progressively weaken due to losses of the energy source, a slightly different method is employed. Using equation 3.13, all required parameters are obtained at the time of landfall and the error term is also randomly sampled only once at the same step to calculate the filling rate coefficient. After that, this coefficient is utilized throughout the entire decayed period to estimate the central intensity at all successive time steps of the same TC. Other details of the methodology given in chapter 3 are not repeated here.

### 5.1.5 Spatial wind field

As a TC propagates along its track, it also spreads the destructive effects over the nearby region. Thus, an analysis of TC's impacts on any particular area must include the wind fields produced by the TCs that directly hit this area, as well as the extreme "near miss" ones. Using this method, one can properly simulate the real TC's influences in many cases, which cannot be estimated alone by the aforementioned track and central intensity model. For example, a TC that has a track (i.e., path of the central location) that parallels or does not cross the coastline, is still considered in the landfall analysis, as long as it generates winds that are strong enough to cause damage in the coastal zone.

However, the evaluation of TC's effects on a specific region will unnecessarily increase the computational demand if the TC is too far away. Therefore, the induced wind field is not examined unless a TC passes within a distance at which the TC can cause losses in the concerned area. Such space is called Damage Distance Threshold (DDT) and must depend on a typical TC characteristic. The reason is that, TCs vary in both shape and size, so a constant threshold is not applicable to all cases as this standard may be too small for large TCs and vice versa. In other research [e.g., (Powell et al. 2005)], the DDT was defined as a function of the radius of maximum wind ( $R_{max}$ ), in which this value was multiplied by an empirical factor. However, no satisfactory explanation was given in the literature for the selection of this multiplication factor.

Fortunately, one parameter, which can be a very useful indicator of TC effects on a given region, is the maximum radius of 30 kt wind speed ( $R_{max30}$ ). As already mentioned in subsection 2.2.2, this value is a limit of the extreme winds, which most possibly can cause structural damage to the exposed properties in the storm-

prone region. Thus, in this study the DDT is taken equal to  $R_{max30}$ . After computing the central location and intensity at each time step,  $R_{max30}$  is first estimated by using equation 4.15. The values of  $R_{max}$  and the Holland parameter ( $B$ ) in this formula are calibrated with the specific case study to achieve the most suitable approach. The details of this method described in subsection 4.2.2 (for  $R_{max}$ ) and 4.3.2 (for  $B$ ) are not repeated here. The distance from the TC's centre to the centre of every grid (cell) in the AOI ( $D$ ) is then calculated and compared with  $R_{max30}$ . If  $D$  is smaller than  $R_{max30}$ , the TC surely has an impact on the concerned cell. The surface wind speed as well as the atmospheric pressure at a grid's centre is recorded in the time series for this cell, using the general parametric wind field technique provided in subsection 4.1.3.

Note that, unlike the track and central intensity models, the error term in the estimates of  $R_{max}$  and  $B$  (i.e., equations 4.17 and 4.29, respectively) is sampled only once, prior to the beginning of each synthetic TC. This random factor is employed throughout the whole TC's lifetime. Therefore, if a simulated TC with the starting quantity of  $R_{max}$  (or  $B$ ) is smaller or larger than average, the values of these parameters will also be kept smaller or larger at all other subsequent time steps.

### 5.1.6 Boundary conditions

As the model uses a combination of different mathematical expressions to approximate the real TC development, it is necessary to include some boundary conditions in the simulation procedure. The introduction of these limits represents the physical constraints of key TC characteristics, which leads to the more reasonable results. Generally, there are two types of boundary conditions for a TC model that are the ending conditions of the lysis and the realistic limits of simulated parameters.

For each synthetic TC, the simulation should be stopped and then moved to the next one, when the current TC is not strong enough to bring about losses at any site within the AOI. There are two criteria to indicate the lysis of a TC. Firstly, the model will be ended if the TC moves outside the TA and thus cannot be expected to have any effect on the AOI. Secondly, a TC is also considered to be completely dispersed when its central intensity drops below a certain degree. In this case, this level is obviously the maximum sustained surface wind speed ( $V_{smax}$ ) of 30 kt (10-min average), as described in subsection 2.2.2.

Furthermore, the simulated parameters derived from the model are restricted by their physical constraints.

Firstly, as suggested by Powell et al. (2005),  $R_{max}$  in the model is bound by its observed maximum and minimum levels, and  $B$  is limited by its theoretical range [i.e., between 0.5 and 2.5 (Holland 1980)].

Secondly,  $V_{smax}$  is not only kept above a lower level (i.e., 30 kt) but also censored to remain below a theoretical maximum value. According to Vickery and Wadhwa (2008), this upper limit can be defined by an equation introduced by Emanuel (1988) as follows:

$$V = \sqrt{2 R_d T_s \ln \left( 1 + \frac{\Delta p}{p_c e} \right)}$$

Where:  $V$  - upper limit of  $V_{max}$  ( $\text{m s}^{-1}$ );  $R_d$  - gas constant for dry air, taken to be  $287.058 \text{ (J kg}^{-1} \text{ K}^{-1})$ ;  $T_s$  - sea surface temperature (K);  $\Delta p$  - central pressure deficit (hPa);  $p_c$  - central pressure (hPa).

Finally, an approach should be utilized to prevent the unreasonable large TCs generated by the simulation. The method is first introduced by this study, in which the ratio  $R_{max30}/R_{max}$  is used to remove unrealistic cases from the final database, when this ratio is greater than or equal to 10. This technique is inspired by Powell et al.'s (2005) idea that a TC spreads out its effects at most  $\sim 10 R_{max}$  from the centre location.

A breakdown of DPs that are removed from the model outcome due the upper limit of  $V_{max}$  and the ratio  $R_{max30}/R_{max}$  is carried out. For a 5,000-year simulation in the case study of Vietnam, the analysis shows that there are 90 DPs where the TC intensities go beyond their theoretical maximum values, while the TCs are considered to be unreasonable large at 673 time steps. However, those 763 rejectable DPs are only a rather small fraction (i.e. less than 0.1%) of the total simulated data (i.e., 924,659 DPs).

## 5.2 Model validation

The model is evaluated through comparisons of historical and simulated TC statistics in the AOI. This validation not only explicitly proves the model's ability to reproduce the actual TC characteristics, but also tests the assumption that the observed values come from the same statistical population as the simulated ones. The modelled parameters are derived from a 5,000-year simulation of TCs originating within the TA.

One important aspect to remember is that, the observed values are also used for the model development and for the calibration process. Therefore, the validation can only be properly carried out when two completely independent historical data sets are utilized for model estimation and for model testing. Basically, it can be done by splitting the entire observed record into two subsets with approximately equal number of DPs. Although the approach ensures that there is absolutely no overlap between the data for model calibration and validation procedures, it also remarkably reduces the sample size of the historical measurements used in these processes. This data separation results in two critical issues.

Firstly, in the model development stage, the use of a limited database severely magnifies the data scarcity problem that reduces the accuracy of various calculations. This is especially the case when coefficients must be determined for each grid in the track model and in the central intensity modelling presented in chapter 3. At some sites, the most precious observations are even missing as they are accidentally moved to the testing data, which makes the coefficient estimation impossible. Although in these cases, values of the substitutive grids (see paragraph 2 in subsection 3.3.3) are employed, it is obviously not the preferable situation and thus the model should minimize the number of lack-data sites. Secondly, according to James and Mason (2005), the uses of limited observed data in the testing process compromises the power of the tests, particularly when the evaluation is carried out for a subregion in the model domain.

In other studies, the historical database was not divided and all observations were employed for both the model estimation and the validation procedure. The reason given in other research is that, the measurements used as testing data hardly overlap the ones utilized to develop the model, due to the nature of the model's validation. In this study, since the tests are performed exclusively within the AOI (or in some cases, the subareas of the AOI), while the model coefficients are determined based on weather conditions over a much larger region (i.e., the TA, see Figure 2.12 in subsection 2.2.2), the validation data account for a small fraction of the model input. Furthermore, for each historical TC, whereas the whole track (i.e., from the initial point to the ending position) is employed to estimate the model's specifications, the approach is tested only at the point of maximum intensity (i.e., maximum central pressure deficit) within the part of the track, which is located inside the AOI. Therefore, the ratio of number of DPs used in the model validation to the ones utilized for the model development is indeed very small. For example, in the case study of Vietnam, even when the model is validated over the entire AOI allowing more observed data to be evaluated, although this proportion varies depending on the sub-model (see Table 5.4), the ratio of number of DPs is always smaller than 5%.

**Table 5.4 Number of DPs used in model validation and model development**

Model validation	Model development					
	Track model		Central intensity model		Wind field model	
Num of DPs	Num of DPs	Ratio to model validation (%)	Num of DPs	Ratio to model validation (%)	Num of DPs	Ratio to model validation (%)
204	6141	3.32	5459	3.74	5071	4.02

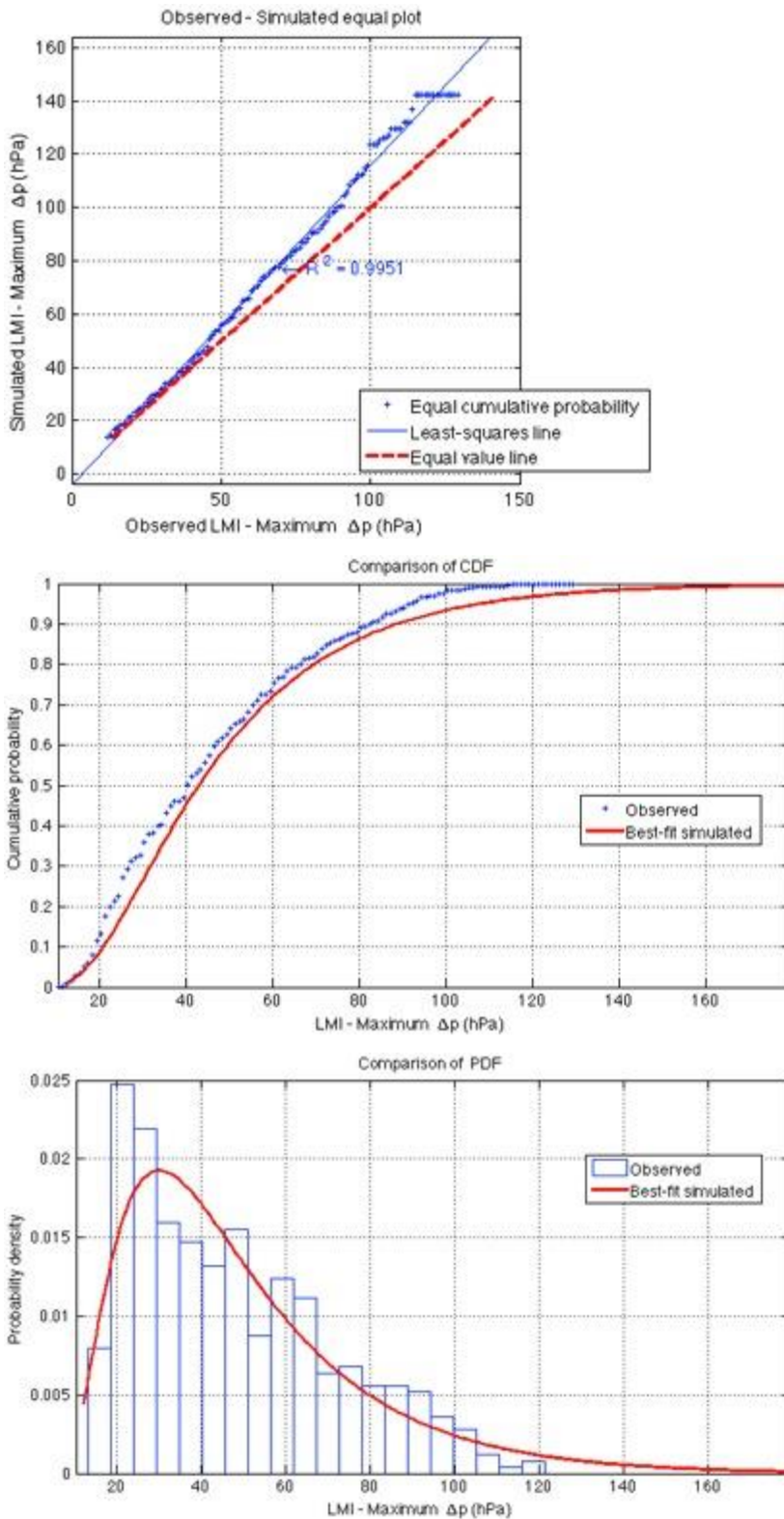
In this research, two models are developed. The first model is the main one, which provides the outcomes for practical applications. Using the identical testing technique as in other studies, a same set of observations (i.e., 37-year historical database) is employed for both the model estimation and the validation procedure. As discussed earlier, although the approach does not completely satisfy the condition of no data overlap, this violation does not much compromise the model's performance. The second model splits the observed record into two roughly equal subsets (i.e., 19-year and 18-year databases), using one for model development and the other for model validation. This model is established only for testing purposes and is solely utilized to validate model results over the entire AOI (i.e., not for testing at each grid within the AOI), when there is still sufficient number of DPs for a meaningful comparison.

### 5.2.1 Testing over the entire AOI

The overall model's performance is first evaluated through comparisons of historical and simulated TC statistics in the AOI. Two parameters, which represent the general TC's behaviour, are examined.

Figure 5.5 gives the test for the Lifetime Maximum Intensity (LMI), which is the maximum value of the central pressure deficit. As can be seen from the figure,

statistical characteristics derived from the synthetic TCs are proven to imitate those of the population of observed TCs with a very high value of coefficient of determination (i.e.,  $R^2 = 0.9951$ ). The TCs produced by the model tend to be more intense than the observed ones, which is the same situation found in other studies [e.g., (Barcikowska 2012)] at the same location (i.e., the Western North Pacific).



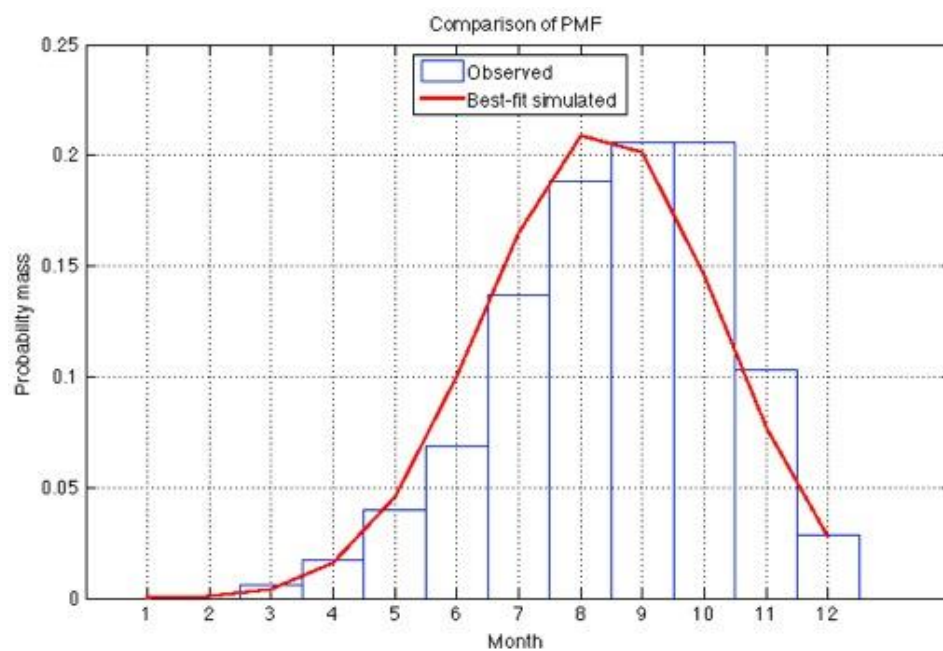
**Figure 5.5 Comparison of LMI over the entire AOI**



Figure 5.6 and Figure 5.7 present another means to validate the model by showing the monthly and annual occurrence rates of the Intense TCs (ITCs) in both the historical and simulated databases. Since ITCs are relatively infrequent but highly destructive, they are certainly the TCs of the most concern in any study. The ITCs are defined as the TCs that have the strength of a category 3 hurricane on the Saffir-Simpson Hurricane Scale (see Table 5.5) or stronger. As can be seen from Figure 5.6, the model retains the important seasonal variation in occurrence of the ITCs.

**Table 5.5 Saffir-Simpson Hurricane Scale [from (Vickery, Skerlj, and Twisdale 2000)]**

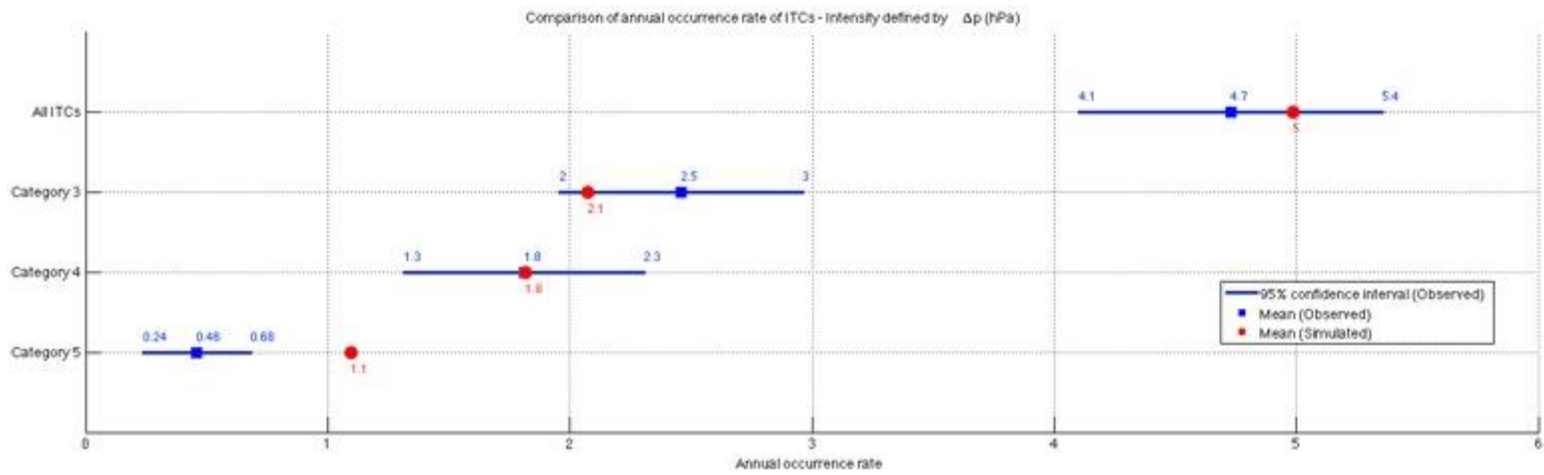
Saffir-Simpson category (1)	Minimum central pressure (mbar) (2)	Maximum sustained wind speed (over water) [m/s (mph)] (3)	Maximum gust speed (over water) [m/s (mph)] (4)	Maximum gust speed (over land, $z_s = 0.03$ m) [m/s (mph)] (5)
1	$\geq 980$	33.1–42.0 (74–94)	40.6–51.9 (91–116)	36.8–48.1 (82–108)
2	979–965	42.0–49.6 (94–110)	51.9–61.7 (116–140)	48.1–58.1 (108–130)
3	964–945	49.6–58.1 (110–130)	61.7–72.7 (140–165)	58.1–69.7 (130–156)
4	944–920	58.1–69.3 (130–155)	72.7–87.3 (165–195)	69.7–85.5 (156–191)
5	$< 920$	$> 69.3$ ( $> 155$ )	$> 87.3$ ( $> 195$ )	$> 85.5$ ( $> 191$ )



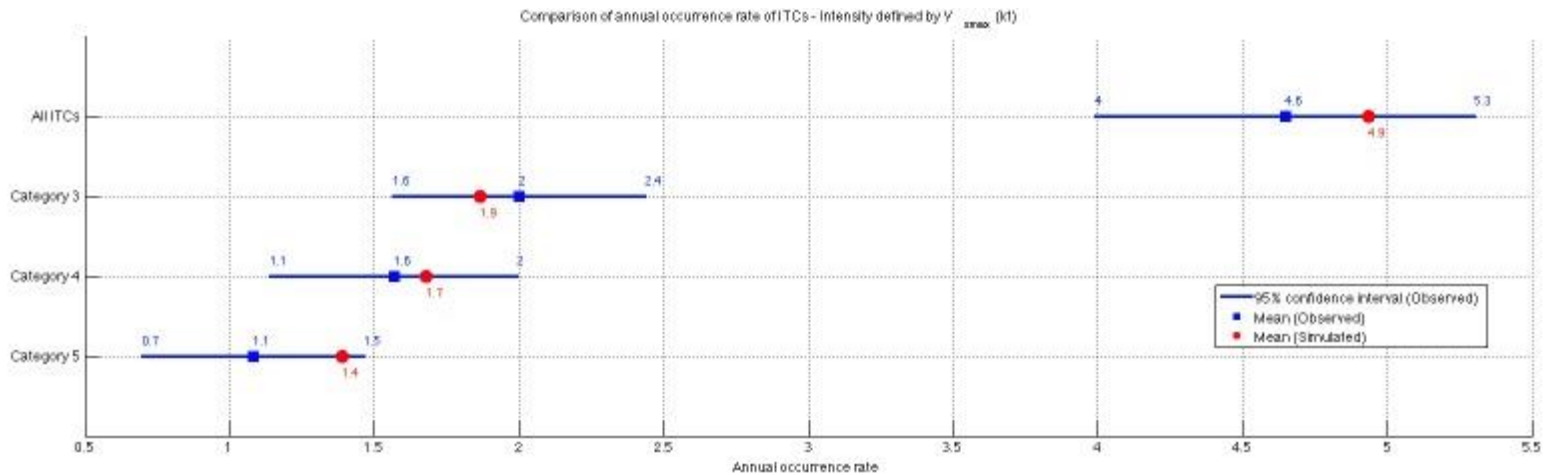
**Figure 5.6 Comparison of monthly occurrence rate of the ITCs over the AOI**

Figure 5.7 compares the annual occurrence rate of the ITCs categorized by the central pressure deficit (Figure 5.7a) as well as by the maximum sustained wind speed (Figure 5.7b). The reason for using both pressure and wind speed to categorize TCs is that, although the classification according to pressure may be error-free and more consistent (Vickery, Skerlj, and Twisdale 2000), wind speed is more relevant to represent a TC's strength in most of the actual projects. As shown in the figure, outcomes from the model clearly prove its ability to reproduce the actual TC characteristics.

Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
 Nguyen Binh Minh - 2015



(a) ITCs categorized by central pressure deficit

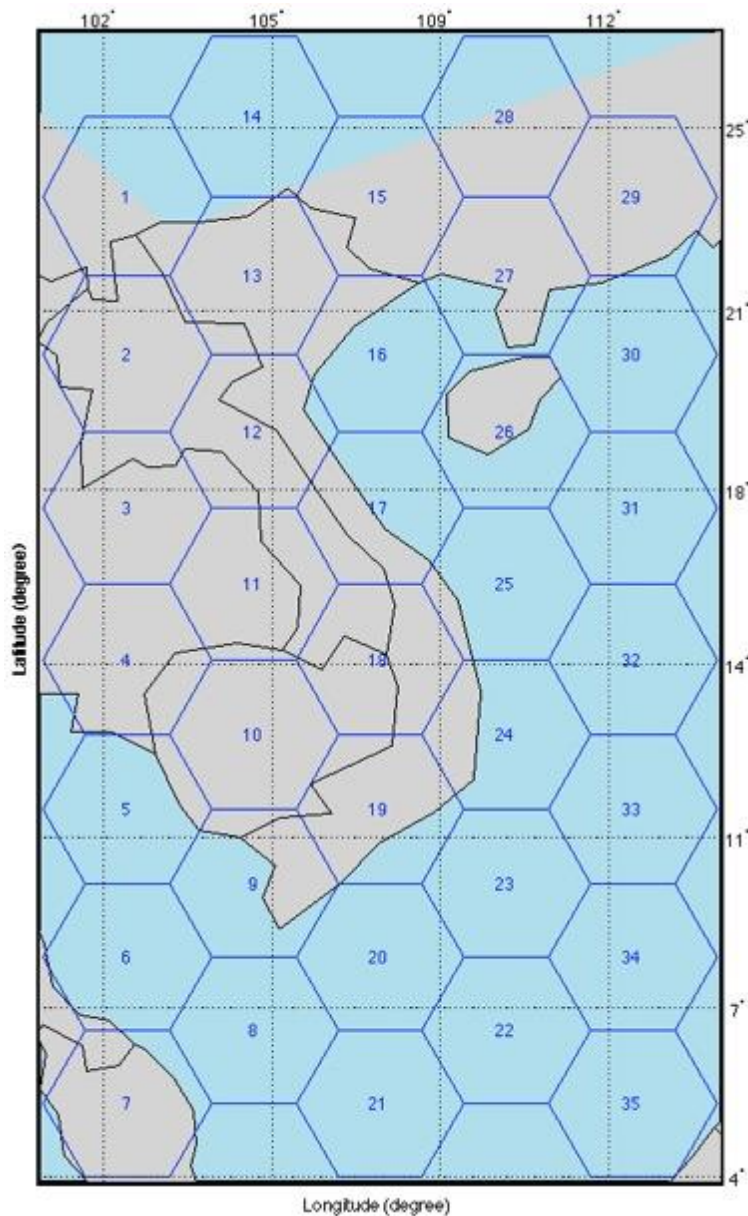


(b) ITCs categorized by maximum sustained wind speed

**Figure 5.7 Comparison of annual occurrence rate of the ITCs over the AOI**

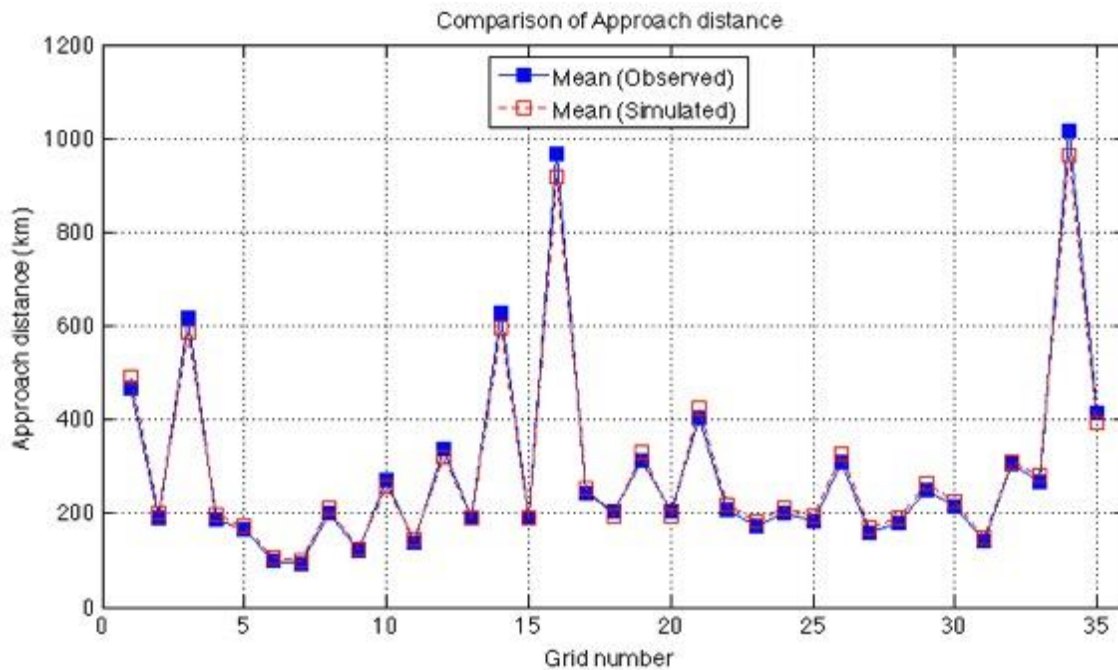
### 5.2.2 Validation for each grid within the AOI

Because TC characteristics heavily depend on local conditions, model coefficients used to estimate the TC's key parameters are the site-specific ones, which vary according to central position. Therefore, it is necessary to evaluate the methodology at each grid within the AOI. For testing purposes, a relatively coarse system of grids (i.e.,  $3.5^\circ$  hexagons, see Figure 5.8) is utilized. Note that, this system has roughly four times more resolution than the one employed in the track and central intensity sub-model (i.e.,  $7^\circ$  hexagons). The reason is that, while this sub-model is developed over the entire TA, model validation is only performed within a much smaller region (i.e., the AOI), which requires a finer system of grids. This substantial flexibility in model resolution is one of the significance of this method.

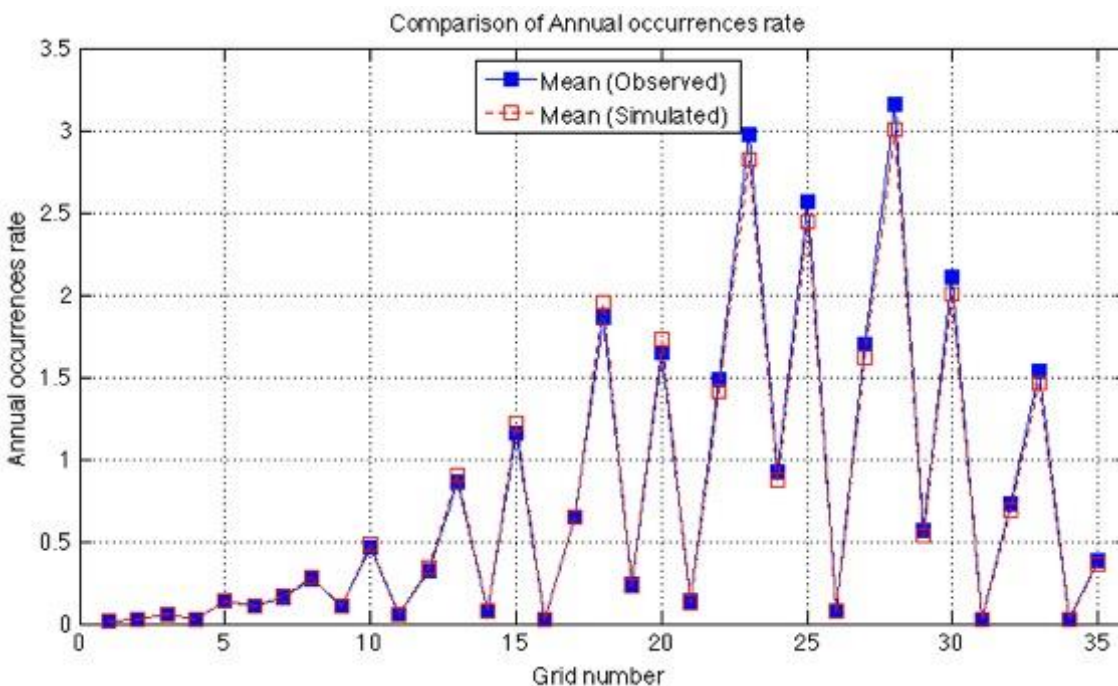


**Figure 5.8** Grids for model validation

Figure 5.9 and Figure 5.10 show the comparisons of the statistics for the various key parameters derived from the observed and simulated database at each grid. In the case that there are several DPs of a same TC located within a specific grid, values are only obtained at the DP that has the maximum intensity. Since each TC is assumed to be an independent realization of the same fundamental stochastic process, the selection of the DP ensures the independence of each sample in the databases. As can be seen in the figures, statistics of synthetic TCs agree well with those of the historical TCs at most locations.



Approach distance: distance from point of maximum intensity to the grid's centre



Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
Nguyen Binh Minh - 2015

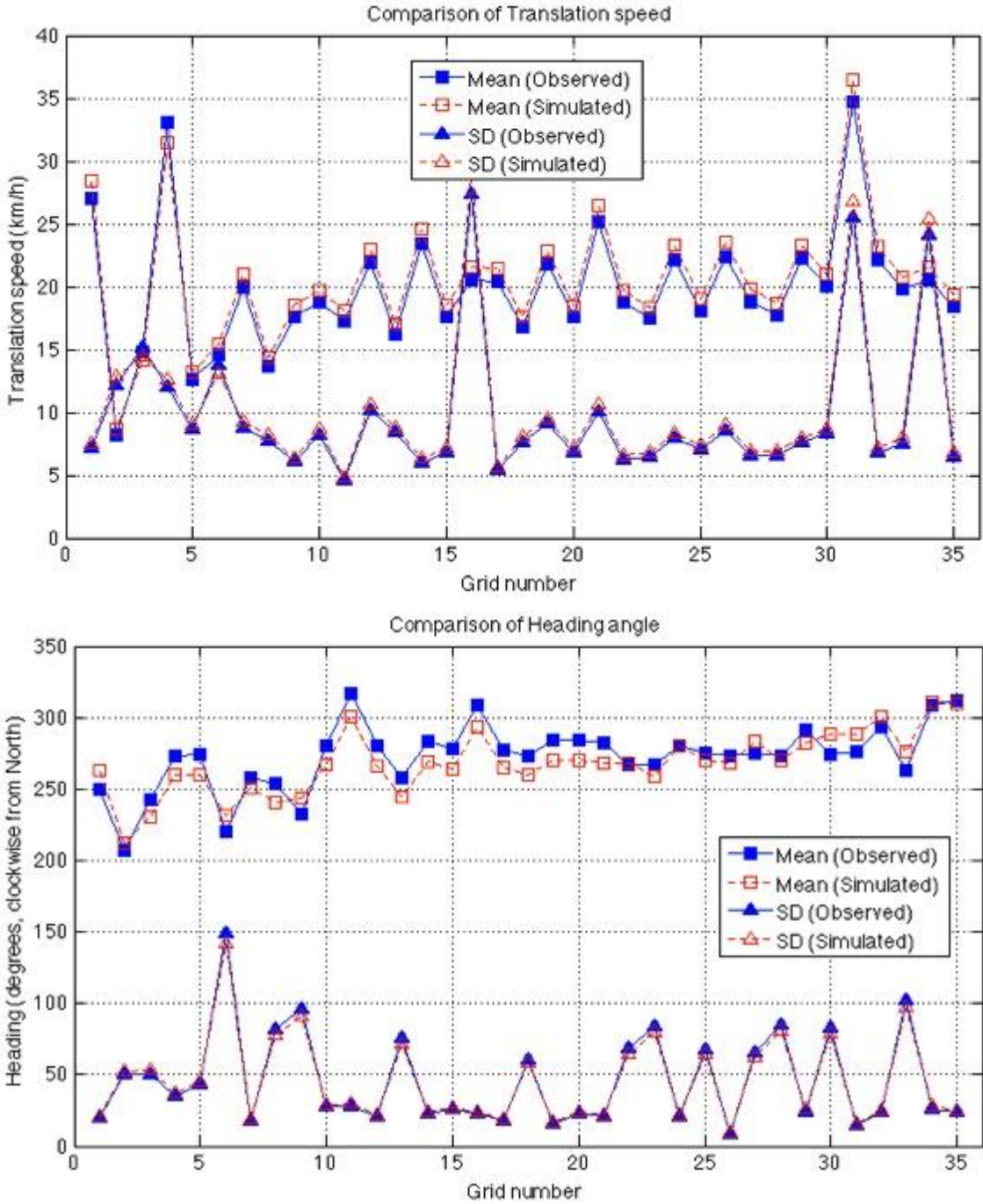
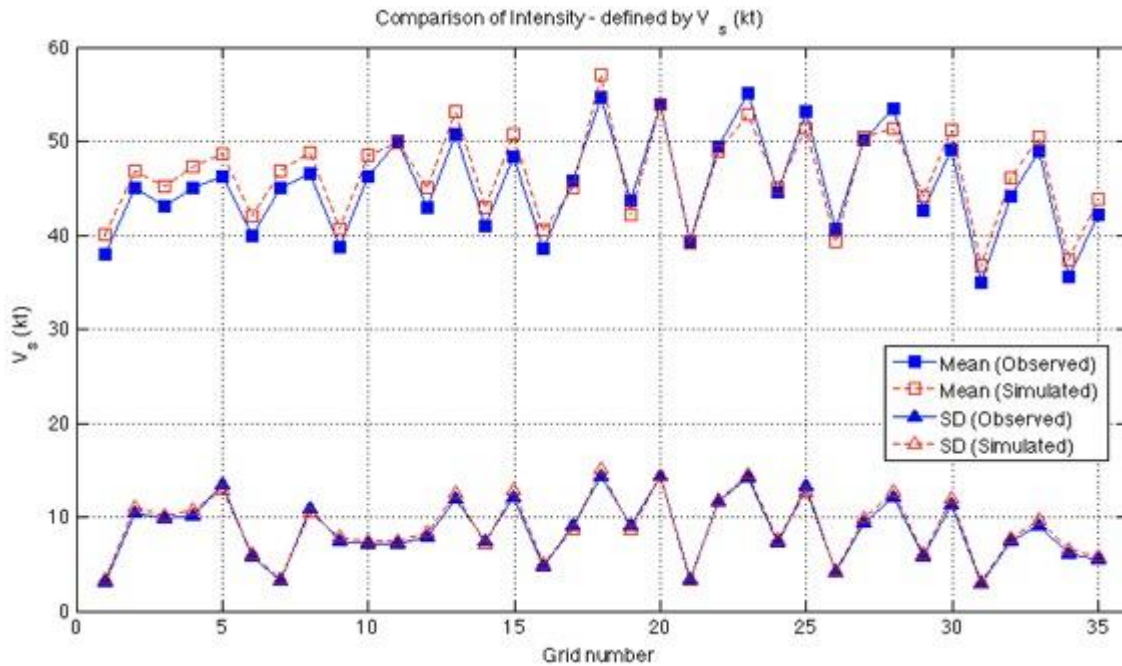


Figure 5.9 Track model validation

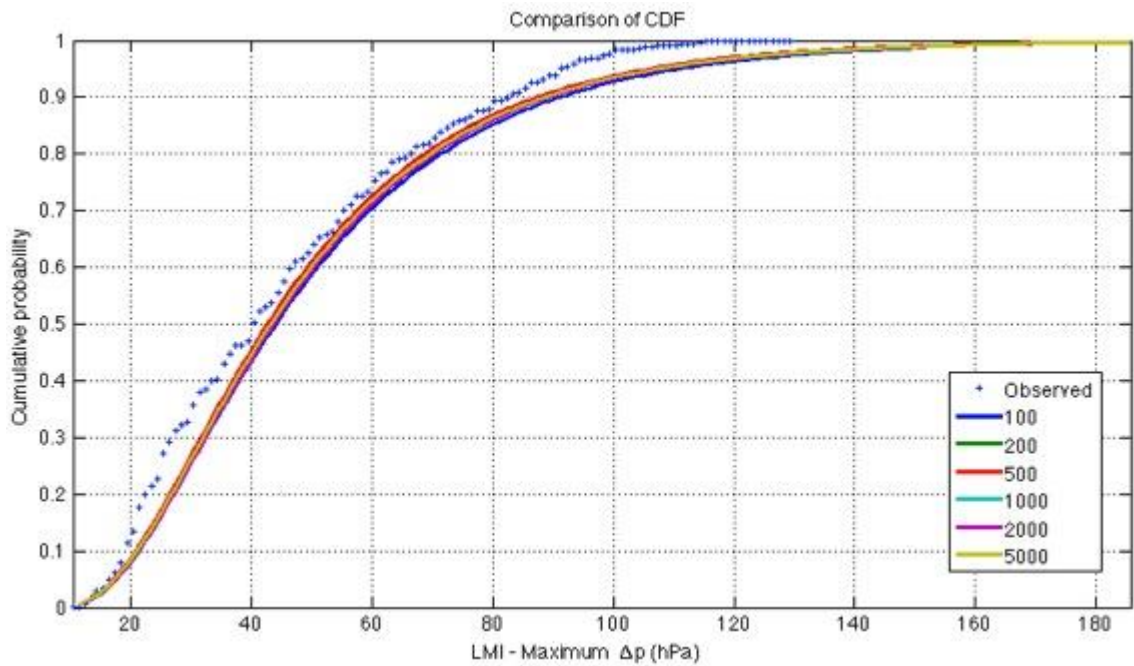


**Figure 5.10 Central intensity model validation**

A common feature of the above figures is the zigzag line connecting values at different grids, which means that in some cases, there is a fairly big difference in TC statistics at two adjacent grids. This can be explained by the fact that while the key parameters change gradually along the coastline, grids are numbered along the meridians and the parallels (see Figure 5.8). Since this research develops a generalized model that can be applied to any case study, it is impossible to number the computational grids precisely along the actual shape of the coastline. However, this is only an approach to display data in the graph and thus it is not considered a detriment to the model since no information is changed or lost.

### 5.2.3 Evaluation of the number of years to be simulated in the model

Finally, the simulation is run repeatedly with different lengths in time, ranging from a low of 100 years to a high of 5,000 years. A comparison of LMI over the entire AOI is carried out to define whether the model benefits much from an extremely long period. Figure 5.11 shows that the lines representing the CDF of LMI converge at a 5,000-year period, and thus the model hardly improves with a simulation length beyond this level. Obviously, if an application requires TC statistics associated with a return period higher than 5,000 years, a simulation must be carried out with the length that is at least equal to this return period. However, for the practical projects, in which key parameters corresponding to a much lower return period (i.e., 50, 100, or at most 1,000 years), a 5,000-year simulation provides sufficient data to derive reliable and meaningful TC statistics.



**Figure 5.11 Comparison of LMI over the AOI for different simulation length**

### 5.3 Possible applications

Since the previous section proves the model's ability to reproduce the actual TC characteristics, the simulation can generate a useable long-term database with an acceptable accuracy for many practical projects. In this section, the wind speed maps and the annual exceedance probability maps are provided as the possible applications of the model outcomes.

#### 5.3.1 Wind speed map

One of the most important factors to establish the building codes in storm-prone regions is the design wind speed map, in which wind speed at every site is determined based on the return period (recurrence interval). The estimated wind speed is shown visually on a map in the form of contour.

Given that a structure is designed to withstand a TC with a return period of  $T$  years, the probability of at least one TC that exceeds the design limit in any one year (i.e., annual exceedance probability) can be obtained from (see subsection 5.1.2):

$$P_a(x > X) = 1/T \quad (5.5)$$

Where:  $x$  - parameter used to design the structure (e.g., maximum sustained wind speed);  $X$  - design limit of  $x$ ;  $T$  - return period (years)

Furthermore, the probability that  $x$  is exceeding  $X$  during the time period of  $t$  years is as follows (Vickery, Skerlj, and Twisdale 2000):

$$P_t(x > X) = 1 - \sum_{x=0}^{\infty} P(x < X|n) p_t(n) \quad (5.6)$$

Where:  $P(x < X/n)$  - probability that  $x$  is less than  $X$  given that  $n$  TCs occur;  $p_t(n)$  - probability of  $x$  TCs occurring during the time period of  $t$  years

For  $t = 1$  years with  $p_t(n)$  defined as Poisson's (Vickery, Skerlj, and Twisdale 2000), the equation (5.6) becomes:

$$P_a(x > X) = 1 - \exp[-\lambda P(x > X)] \quad (5.7)$$

Where:  $\lambda$  - annual occurrence rate (TCs/year);  $P(x > X)$  - exceedance probability

Substituting (5.5) into (5.7), one can determine  $P(x > X)$  from:

$$1/T = 1 - \exp[-\lambda P(x > X)] \quad (5.8)$$

The cumulative probability associated with the design level  $X$  is:

$$CP(X) = 1 - P(x > X) \quad (5.9)$$

Using  $CP(X)$  and the CDF of the parameter derived from the long-term simulation, one can define the design level  $X$ .

Figure 5.12 gives the estimated values corresponding to different return periods for the maximum sustained wind speed at each grid within the AOI. These quantities are presented on the map as shown in Figure 5.13.

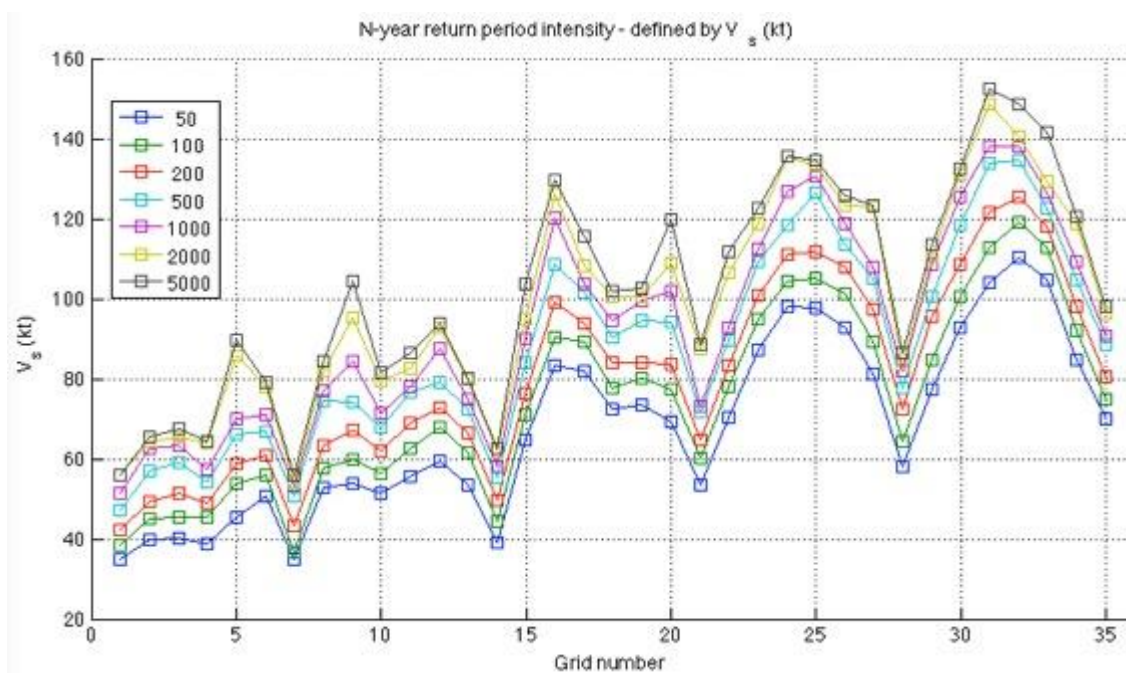
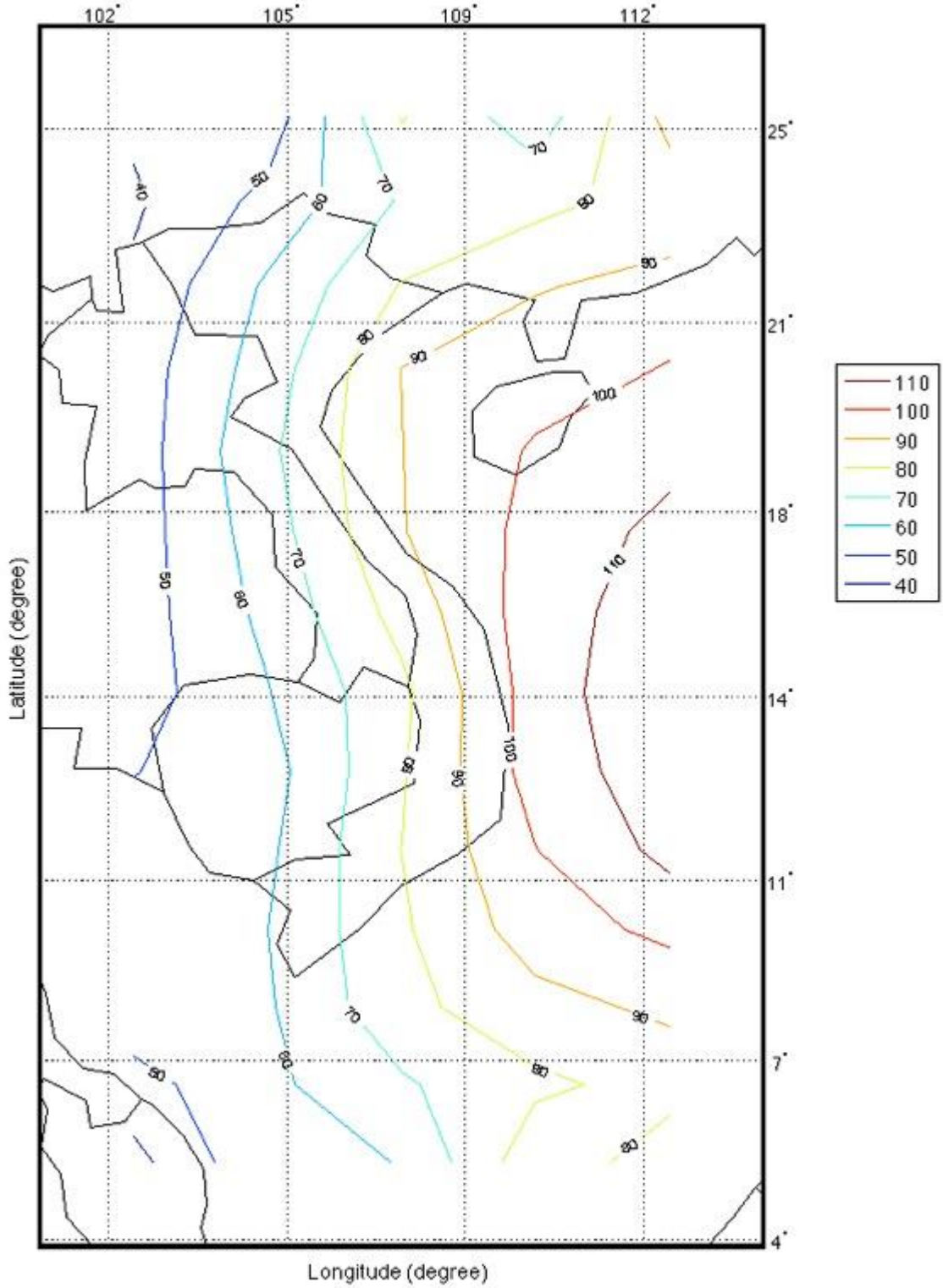


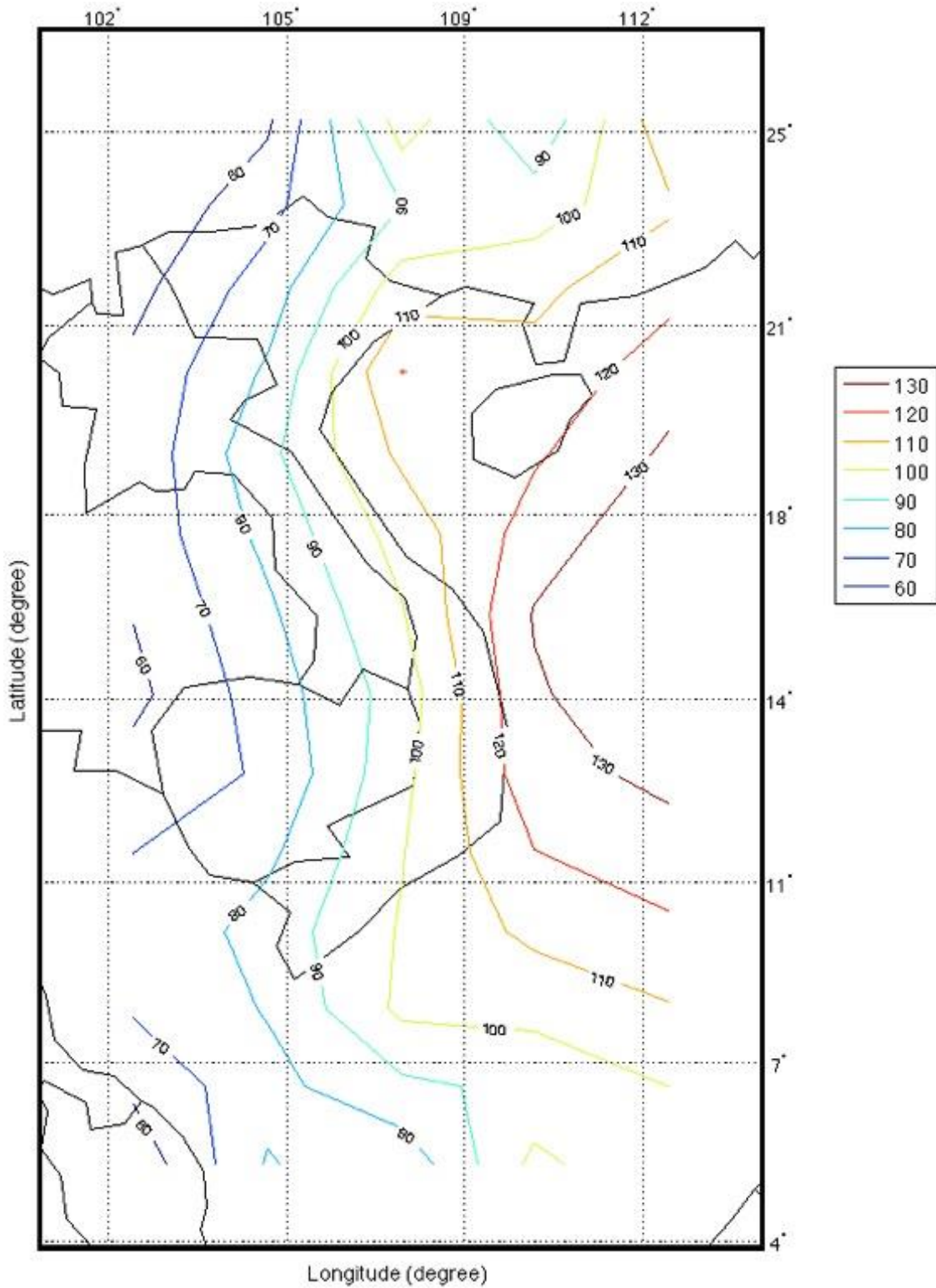
Figure 5.12 N-year return period wind speed at each grid within the AOI



Long-term regional simulation of tropical cyclones using a Generalized Stochastic Empirical Storm Model. A case study in the Western North Pacific  
Nguyen Binh Minh - 2015



(a) 100-year return period

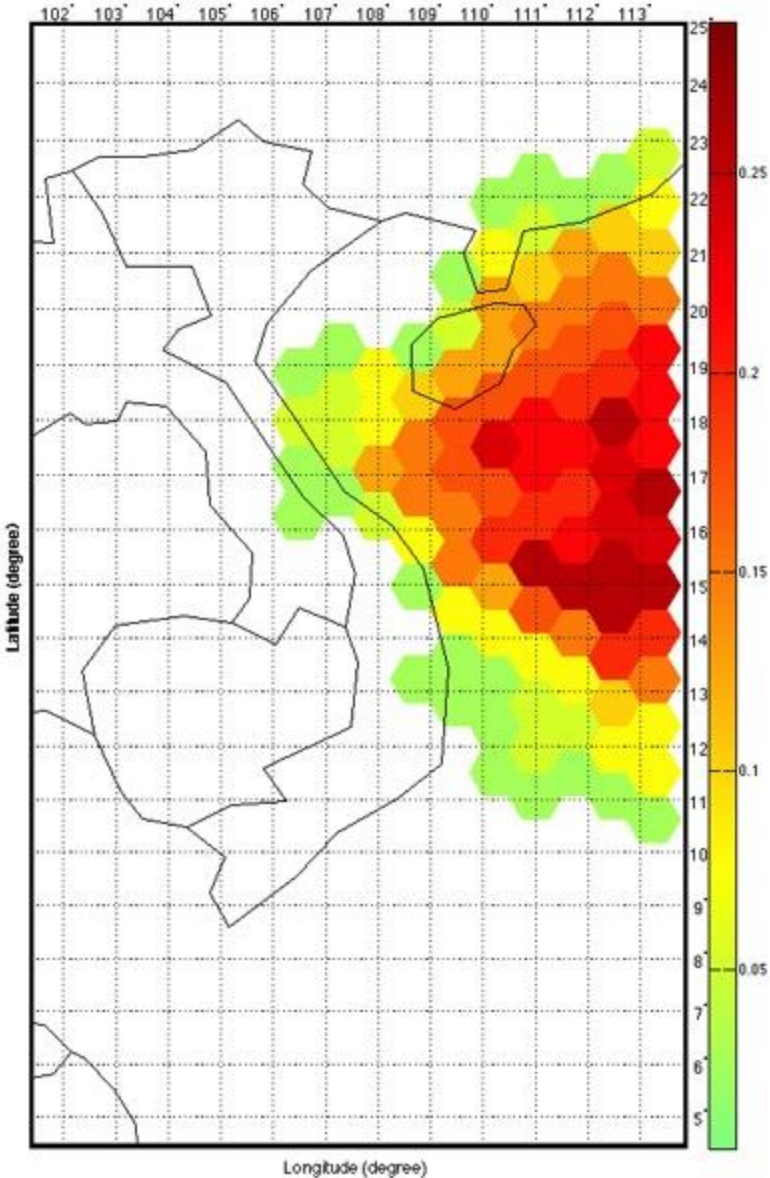


(b) 1000-year return period

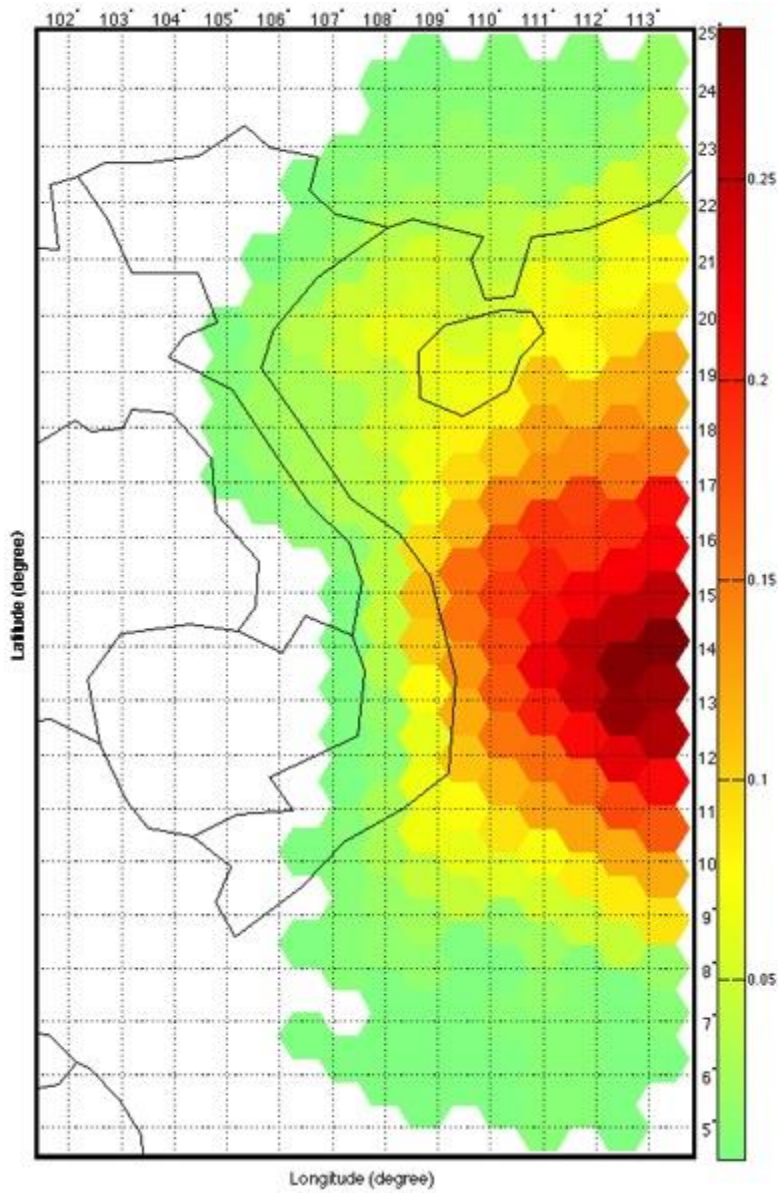
**Figure 5.13 Wind speed map**

### 5.3.2 Annual exceedance probability map

Another important application of the long-term simulation is the annual exceedance probability map. Because TC risk is computed as the product of probability and consequence, this map is used to assist in making decisions whether a project should be allowed to go forward in a zone of a certain risk, or which site within the AOI is suitable for construction with the lowest risk. For instance, if the wind speed of a category 3 hurricane (see Table 5.5) is selected as the threshold, Figure 5.14 gives the annual exceedance probability maps derived from the historical and simulated data. The figure shows that the model results not only agree well with the observations in the spatial distribution of the annual exceedance probability (e.g., the area of high probability in the middle of Vietnam), but also produce more data to allow risk analysis in many regions (e.g., the Mekong Delta), which is impossible to do with the observed data.



(a) Historical data



(b) Simulated data

**Figure 5.14 Annual exceedance probability map**

# 6 CONCLUSIONS AND RECOMMENDATIONS

## 6.1 Conclusions

Tropical Cyclones (TCs) have an extensive negative impact on numerous aspects of human society and the ecological environment. In many regions, TCs can be the cause of most of the mortalities and economic losses due to natural disasters. Unfortunately, unlike other types of extreme weather events, current risk reduction measures for TCs are not completely successful in lessening their consequences. The reason for this inefficiency is the remaining uncertainty concerning the reliability of the estimates of TC key parameters. Because TCs are both relatively infrequent and small in terms of the length of coastlines affected by TCs each year, reliable observations on the tracks and winds of TCs having affected many regions are restricted to quite a small number. Thus, it is not feasible to derive accurate key parameters for the most intense TCs, solely based on historical records and on which risk analyses, building codes, and designs of coastal defence structures will rely, without producing large errors.

This research presents a comprehensive methodology to develop a numerical TC model, as an effective and widely accepted technique to overcome the observed data scarcity problem. The model is called the Generalized Stochastic Empirical Storm Model (GSESM). TCs are stochastically simulated over a period of hundreds or even thousands of years, which results in a long-term database of synthetic TCs, with specifications of the central track and intensity as well as the wind field at each time step. On the condition that statistical characteristics derived from these simulated TCs are proven to imitate those of the population of real TCs, results from a long-term TC model can be utilized as a complete input for any TC-related study. The approach has two important features.

- Firstly, in contrast with many other research, this model is carried out at a regional scale. Because TC evolution is heavily dependent on local conditions, a regional domain not only maintains a relative homogeneity

within both the input and outcome, but it also reduces computational demand.

- Secondly, since the model has a generalized theoretical framework and contains the worldwide historical weather data, it can be applied to any case study. Once users define the Area Of Interest (AOI), a stepwise calibration procedure is automatically performed by a computer program to achieve the most suitable approach and to determine every single detail of the model, specifically for this user-defined AOI.

Although the aim of this study is to provide a generalized model, which can be used at any location, a case study must be given in order to verify the theoretical framework and to evaluate the model performance. The Western North Pacific (WNP) as the case study is not only chosen because the basin has the highest TC occurrence rate and experiences the strongest TCs, but also due to the serious shortness of historical data, which is the exact problem that the model developed in this study intends to overcome. This section summarizes the arguments given in the course of this dissertation to answers the research questions proposed in subsection 1.5.2.

### 6.1.1 Questions associated with the model setup

The arguments, which are the answers to these questions, are given in chapter 2 as follows:

*What type of data must be collected to construct the GSESM?*

The basic and most important input for a TC climatology study is the so-called Best Track Data (BTD). A BTD generally contains positions and intensities, which are measured every 6 hours for each historical TC. In addition to the BTD record, other data sources are essential for the estimation of various required parameters as well as to present the model results.

*Which sources are available for each required type of data?*

In the WNP (because TCs are monitored by various agencies) there are at least 4 different BTD sources. Other input data include data from the Twentieth Century Reanalysis Project (Compo et al. 2011), mean monthly values of Sea Surface Temperature (SST) taken from the Extended Reconstructed SST V3b record (NOAA 2014a), and digital maps from the 1:110m Cultural Vectors (Natural Earth 2014).

*How to evaluate the quality of those sources?*

In order to evaluate the reliabilities of the available records, different BTD's are compared with independent reference data. Two trustworthy sources were chosen as references, namely the Blended Sea Winds (Zhang et al. 2006) for TCs with low intensities, and the aircraft measurements collected during the THORPEX Pacific Asian Regional Campaign [TPARC-2008 (NOAA 2008)] for extreme conditions.

*If there are several accessible sources, which one will be chosen and why?*

The historical record from the Japan Meteorological Agency [JMA (JMA 2014)] is selected as the BTD in this research due to its superior accuracy. The JMA is successful in keeping the homogeneity within its database by using the same

methods and information sources during the entire monitoring period. It makes JMA's record more reliable than other ones to derive TC statistics. Furthermore, the JMA's BTD also provides the observations of a valuable parameter, which is an advantage over other sources in defining model coverage.

*What is the AOI?*

The AOI is the region where TC parameters are derived from both historical and simulated data, in order to evaluate model results and assess risks due to TCs.

*How to define the geographic range of the AOI?*

In this study, the AOI is determined by considering its underlying meaning. That is, basically an AOI is the region that contains inhabited areas or properties, and which researchers would like to assess due to the risks of TCs. Therefore, the AOI consists of two parts. The first part is the mainland of a state or country on which the research will focus. The second part is the oceanic region where offshore constructions are located or proposed.

*What is the Threat Area (TA)?*

The main focus is on the TCs that are capable of influencing conditions in the AOI. The region that covers all the central locations of the TC is called the TA.

*Which criteria can be used to determine if a data point should be included in the research or not?*

One parameter, which is a very useful indicator of TC effects on a given region, is the most outer radius of 30 kt surface wind speed (10-min average).

*How to utilize this indicator to define the boundary of the TA?*

For the WNP basin, the JMA has included such a crucial TC parameter in its BTD since 1977, which formulates a basis for determining the TA. While this technique still best captures the statistical characteristics of historical TCs having affected the AOI, it also effectively reduces the computational demand by removing a large proportion of TC centres, which are completely irrelevant to the research.

*What is the shape and the size of the computational grids?*

The hexagon is used in this study, as it is the best compromise solution for the shape of computational grids. A hexagon absolutely does not have the limitation of uncovered or overlapping areas. Furthermore, other issues are also minimized because a hexagon has equal lengths of side and inside angles, and therefore it is the best approximate to the ideal shape of the circle. Unlike the shape of computational grids, the dimensions of these cells cannot be specified at the preparation stages of the model, because the influences of grid size on model performance are only revealed in the evaluation of the research equations.

### 6.1.2 Questions related to the modelling of central track, intensity, and surface wind field

The answers of these questions are given in chapter 3 and chapter 4 as follows:

*Which are the current theoretical frameworks for modelling key TC parameters?*

Although numerous models have been employed, they are all based on two distinct underlying approaches. These basic methods are the single site stochastic technique, which does not simulate the entire TC track, and the empirical track approach modelling the full trajectories.

*What are their pros and cons?*

Although the single site model is relatively simple, it is facing some critical, and in some cases unacceptable, problems. Even if one can accept all drawbacks, the serious shortages of observed data in many important areas still prevent the effective applications of the single site technique in deriving meaningful site-specific TC statistics.

In contrast, the Empirical Track Modelling (ETM) simulates a full track of each synthetic TC, from its initial point over water to the lysis. The ETM overcomes the disadvantages of the single site approach in many aspects. The only potential drawback of the ETM is that it involves creating a large modelled database, which may require an enormous computational demand.

*Which one should be chosen as the basis for the GSESM?*

Although one cannot expect any miracles, still being a long way from reality, the approach of the ETM certainly represents the state-of-art in long-term simulation.

*What are the limitations of the (chosen) existing method?*

None of the studies in the literature included either a detailed analyses of the ETM or any direct comparisons with the ETM. Thus, ETM's weaknesses, which are the basis for possible expansions or improvements, were not revealed and the provided modifications (if any) could not be regarded as beneficial or harmful. Furthermore, when coupled with the fact that the ETM was developed particularly for simulating TCs in the Atlantic, it is questionable whether the approach is applicable to other basins.

*How can the GSESM overcome these limitations?*

The remaining ambiguousness of the ETM is clarified in subsection 3.3.3.

A synthetic TC propagates in space along a trajectory defined by the track model. All the details of this model are calibrated with the specific case study to achieve the most suitable approach described in subsection 3.3.4. Regarding the central intensity, it also changes over time when a TC moves along its track. Depending on the eye's location, the central intensity is estimated by two different models that are also adjusted through the stepwise calibration procedures. When the TC is over water, the relative intensity method provided in section 3.4 is employed to simulate the intensity development. Once the TC makes landfall, a separated model given in section 3.5 must be used to describe the decay of this TC. The outcome of the track and central intensity models is a full trajectory of the simulated TC, together with the specification of its central location and strength.

As a TC propagates along its track, it also spreads its destructive effects over the nearby region. Thus, an analysis of TC's impacts on any particular area must include the wind fields produced by the TCs that directly hit this area, as well as the extreme "near miss" ones. Using this method, one can properly simulate the real TC's influences in many cases, which cannot be estimated alone by the



aforementioned track and central intensity model. This wind field model is presented in chapter 4.

*In comparison with the ETM, what will be the GSESM's improvements?*

Table 6.1 gives a summary of the GSESM's improvements over the ETM in all sub-models provided in chapter 3 and chapter 4. As can be seen from the table, this research presents an advanced method for modelling TCs.

**Table 6.1 Summary of the GSESM's improvements over the ETM**

Sub-model	R <sup>2</sup>		Improvement	
	ETM	GSESM	Absolute	Relative (%)
Track	0.1480	0.1778	0.0298	20
Intensity over water	0.9350	0.9591	0.0241	3
Inland intensity (decay model)	0.1418	0.3417	0.1999	141
Radius of maximum wind	0.0365	0.1230	0.0865	237
Holland B parameter	0.3729	0.4299	0.0570	15

### 6.1.3 Questions connected with the model run

The arguments, which give answers to these questions, are given in chapter 5 as follows:

*How long should a reasonable length of a synthetic TC database be?*

Although the approach is theoretically capable of implementing a model of unlimited length, a prolonged simulation significantly increases computational demand, which can be a crucial limitation. In subsection 5.2.3, a period of 5,000 years is defined to be the reasonable length of TC simulation in this study.

*Which basic discrete distribution should be employed to approximate the TC annual occurrence rate?*

The number of TCs to be simulated in any year is obtained by a sampling from a distribution, which is defined based on the historical annual occurrence rate of TCs in the research area. In this research, all of the three basic discrete distributions (i.e., the negative binomial, binomial, and Poisson distributions) are tested to the historical annual occurrence rate. Instead of relying only on the ordinary chi-square test, several fit indices are used to determine the best-fit distribution.

*How to define a set of initial points for TCs in the simulation?*

For each synthetic TC, the initial specifications (i.e., location, time, and all other key parameters) of the first time step are sampled from a set of starting conditions of the historical TCs.

*What is the Damage Distance Threshold (DDT)? How to define this value?*

The evaluation of TC's effects on a specific region will unnecessarily increase the computational demand if the TC is too far away. Therefore, the induced wind field is not examined unless a TC passes within a distance at which the TC can cause losses in the concerned area. Such a distance is called Damage Distance Threshold

(DDT) and must depend on a typical TC characteristic. As mentioned in subsection 2.2.2, the maximum radius of 30 kt wind speed ( $R_{max30}$ ) is the limit of the extreme winds, which most possibly can cause structural damage to the exposed properties in the storm-prone region. Thus, in this study the DDT is taken equal to  $R_{max30}$ .

*Which physical boundary conditions should be included in the model to introduce the realistic limits of the parameters as well as to define the lysis of a TC?*

As the model uses a combination of different mathematical expressions to approximate the real TC development, it is necessary to include some boundary conditions in the simulation procedure. The introduction of these limits represents the physical constraints of key TC characteristics, which leads to the more reasonable results. Generally, there are two types of boundary conditions for a TC model that are the ending conditions of the lysis and the realistic limits of simulated parameters. These limits are described in subsection 5.1.6.

*How to validate the model?*

The model is evaluated through comparisons of historical and simulated TC statistics in the AOI. This validation not only explicitly proves the model's ability to reproduce the actual TC characteristics, but also tests the assumption that the observed values come from the same statistical population as the simulated ones. The modelled parameters are derived from a 5,000-year simulation of TCs originating within the TA.

The overall model's performance is first evaluated through comparisons of historical and simulated TC statistics in the AOI. Two parameters, which represent the general TC's behaviour, are examined in subsection 5.2.1. Because TC characteristics heavily depend on local conditions, the model coefficients used to estimate the TC's key parameters are the site-specific ones, which vary according to their central position. Therefore, it is necessary to evaluate the methodology at each grid within the AOI. This validation is given in subsection 5.2.2

*What are the possible applications of the model outcomes?*

Since the previous section proves the model's ability to reproduce the actual TC characteristics, the simulation can generate a useable long-term database with an acceptable accuracy for many practical projects. The wind speed maps (subsection 5.3.1) and the annual exceedance probability maps (subsection 5.3.2) are provided as the possible applications of the model outcomes.

## 6.2 Recommendations for future research

Although the objective of this study is to develop a complete methodology, in which every single detail is given a comprehensive description, there still remain some limitations that may be the interesting topics for future research.

Firstly, all of the fundamental equations, which are employed to simulate TCs in this study, are in linear form. Although a simple formula can remarkably reduce the required computational demand, especially when subsequent analyses are needed, this simplification results in a relatively low value of coefficient of determination (i.e.,  $R^2$ , see Table 6.1). Thus, other forms of equations can be considered to improve the model's performance. Furthermore, the basic formulae

do not include several physical factors that influence the TC's evolution. Therefore, a detailed analysis that explores and examines all parameters, which may possibly contribute to the developments of TCs, is necessary to introduce more realistic factors to the approach and to produce more reasonable results.

Secondly, as indicated in subsection 4.1.3, for the case study of Vietnam in the WNP, due to the significant scatter of field data, there is no detailed validation for the wind field model. Although this model has already proven to be a reasonable approach for practical applications in many other areas, a validation should be carried out as soon as the field data are available to test the suitability for this specific case study.

Finally, because at the WNP no similar research has been performed and has been described in the literature, it is not possible to directly compare the results of this model with the outcomes of other studies. Thus, if a TC model using a different methodology is carried out in a future research, it should be compared to the model in this study.

## REFERENCES

- Aixue Hu, G. A. M. (2009). "Effect of the Atlantic hurricanes on the oceanic meridional overturning circulation and heat transport." *Geophysical Research Letters - GEOPHYS RES LETT*, 36(3).
- ANSI. (1982). *Minimum Design Loads for Buildings and Other Structures*. American National Standards Institute, Inc., New York.
- Apivatanagul, P., Davidson, R., Blanton, B., and Nozick, L. (2011). "Long-term regional hurricane hazard analysis for wind and storm surge." *Coastal Engineering*, 58(6), 499–509.
- Ariffin, M., and Moten, S. (2009). *Relationship Between Maximum Sustained Wind Speed and Central Pressure of Tropical Cyclones*. Malaysian Meteorological Department, Malaysia.
- ASCE. (1993). *ASCE-7 Minimum Design Loads for Buildings and Other Structures*. American Society of Civil Engineers, New York.
- ASCE. (1998). *Minimum design loads for buildings and structures*. American Society of Civil Engineers, Reston, VA.
- ASCE. (2003). *Minimum design loads for buildings and structures*. American Society of Civil Engineers, Reston, VA.
- ASCE. (2006). *ASCE-7 Minimum Design Loads for Buildings and Other Structures*. American Society of Civil Engineers, New York.
- AS/NZS. (2002). *Standards Australia/Standards New Zealand. Structural design actions, part 2: wind actions*.
- Atkinson, G. D., and Holliday, C. R. (1977). "Tropical Cyclone Minimum Sea Level Pressure/Maximum Sustained Wind Relationship for the Western North Pacific." *Monthly Weather Review*, 105(4), 421–427.
- Bankoff, G., Hilhorst, D., and Frerks, G. (2004). *Mapping Vulnerability: Disasters, Development and People*. Routledge, London ; Sterling, VA.
- Barcikowska, M. (2012). "Variability and trends of tropical cyclones over the western North Pacific for the last decades."
- Barcikowska, M., Feser, F., and von Storch, H. (2012). "Usability of Best Track Data in Climate Statistics in the Western North Pacific." *Monthly Weather Review*, 140(9), 2818–2830.
- Batts, M. E., Simiu, E., and Russell, L. R. (1980). "Hurricane Wind Speeds in the United States." *Journal of the Structural Division*, 106(10), 2001–2016.
- Belanger, J. I., Curry, J. A., and Hoyos, C. D. (2009). "Variability in tornado frequency associated with U.S. landfalling tropical cyclones." *Geophysical Research Letters*, 36(17), L17805.
- Bendimerad, F. (2004). *Disaster risk reduction and sustainable development*. World Bank.

- Booij, N., Ris, R. C., and Holthuijsen, L. H. (1999). "A third-generation wave model for coastal regions: 1. Model description and validation." *Journal of Geophysical Research: Oceans (1978–2012)*, 104(C4), 7649–7666.
- Brettschneider, B. (2006). "Estimating Atlantic Basin Tropical Cyclone Landfall Probability for the United States."
- Brettschneider, B. (2008). "Climatological Hurricane Landfall Probability for the United States." *Journal of Applied Meteorology and Climatology*, 47(2), 704–716.
- Burnham, K. P., and Anderson, D. R. (2013). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.
- Camargo, S. J., Barnston, A. G., and Zebiak, S. E. (2005). "A statistical assessment of tropical cyclone activity in atmospheric general circulation models." *Tellus A*, 57(4).
- Cardone, V. J., and Cox, A. T. (2009). "Tropical cyclone wind field forcing for surge models: critical issues and sensitivities." *Natural Hazards*, 51(1), 29–47.
- CBO. (1995). *Federal disaster assistance : report of the Senate Task Force on Funding Disaster Relief*. United States Government Printing Office.
- Chu, P.-S., and Wang, J. (1998). "Modeling Return Periods of Tropical Cyclone Intensities in the Vicinity of Hawaii\*." *Journal of Applied Meteorology*, 37(9), 951–960.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Mauerer, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J. (2011). "The Twentieth Century Reanalysis Project." *Quarterly Journal of the Royal Meteorological Society*, 137(654), 1–28.
- Courtney, J., Buchan, S., Cerveny, R. S., Bessemoulin, P., Peterson, T. C., Rubiera Torres, J. M., Beven, J., King, J. C., Trerwin, B., and Rancourt, K. (2012). "Documentation and verification of the world extreme wind gust record: 113.3 m s<sup>-1</sup> on Barrow Island, Australia, during passage of tropical cyclone Olivia." *Australian Meteorological and Oceanographic Journal*, 62(1), 1–9.
- Cox, A. T., Greenwood, J. A., Cardone, V. J., and Swail, V. R. (1995). "An interactive objective kinematic analysis system." *Proceedings of 4th International Workshop on Wave Hindcasting and Forecasting*, Atmospheric Environment Service, Alberta, Canada, 109–118.
- CRED. (2008). "EM-DAT International Disaster Database." <<http://www.emdat.be/>>.
- CUBiC. (1985). *Caribbean Uniform Building Code. Structural design requirements WIND LOAD, part 2, section 2*. Caribbean Community Secretariat, Georgetown, Guyana.
- Darling, R. W. R. (1991). "Estimating Probabilities of Hurricane Wind Speeds Using a Large-Scale Empirical Model." *Journal of Climate*, 4(10), 1035–1046.

- Dasgupta, S., Laplante, B., Meisner, C., Wheeler, D., and Yan, J. (2009). "The impact of sea level rise on developing countries: a comparative analysis." *Climatic Change*, 93(3-4), 379–388.
- DeMaria, M., and Kaplan, J. (1999). "An Updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and Eastern North Pacific Basins." *Weather Forecast*, 14(3), 326–337.
- DeMaria, M., Knaff, J. A., and Kaplan, J. (2006). "On the Decay of Tropical Cyclone Winds Crossing Narrow Landmasses." *Journal of Applied Meteorology and Climatology*, 45(3), 491–499.
- Depperman, R. C. E. (1947). "Notes on the origin and structures of Philippine typhoons." *Bulletin of the American Meteorological Society*, 28, 399–404.
- Dunnavan, G. M., and Diercks, J. W. (1980). "An Analysis of Super Typhoon Tip (October 1979)." *Monthly Weather Review*, 108(11), 1915–1923.
- Dvorak, V. F. (1975). "Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery." *Monthly Weather Review*, 103(5), 420–430.
- Elsner, J. B., and Kara, A. B. (1999). *Hurricanes of the North Atlantic: Climate and Society*. Oxford University Press.
- Emanuel, K. (2001). "Contribution of tropical cyclones to meridional heat transport by the oceans." *Journal of Geophysical Research: Atmospheres*, 106(D14), 14771–14781.
- Emanuel, K. A. (1988). "The maximum intensity of hurricanes." *Journal of the Atmospheric Sciences*, 45, 1143–1155.
- Emanuel, K., Ravela, S., Vivant, E., and Risi, C. (2006). "A Statistical Deterministic Approach to Hurricane Risk Assessment." *Bulletin of the American Meteorological Society*, 87(3), 299–314.
- FEMA. (2003). *Multi-hazard loss estimation methodology - Hurricane Model - Technical manual*. Federal Emergency Management Agency, Washington, DC.
- FEMA. (2013). *Surviving the Storm: A Guide to Hurricane Preparedness*. Federal Emergency Management Agency.
- Foster, R. C. (2005). "Why Rolls are Prevalent in the Hurricane Boundary Layer." *Journal of the Atmospheric Sciences*, 62(8), 2647–2661.
- Georgiou, P. N. (1985). "Probability distribution for Texas Gulf coast hurricane effects of engineering interest." Ph.D. Thesis, University of Western Ontario, London, Ontario, Canada.
- Georgiou, P. N., Davenport, A. G., and Vickery, B. J. (1983). "Design wind speeds in regions dominated by tropical cyclones." *Journal of Wind Engineering and Industrial Aerodynamics*, 13(1–3), 139–152.
- Goklany, I. M. (2009). "Deaths and Death Rates from Extreme Weather Events: 1900-2008." *Journal of American Physicians and Surgeons*, 14(4).
- Green, R. J. (2009). *Coastal Towns in Transition: Local Perceptions of Landscape Change*. Springer Science & Business Media.

- Hallegatte, S. (2007). "The Use of Synthetic Hurricane Tracks in Risk Analysis and Climate Change Damage Assessment." *Journal of Applied Meteorology and Climatology*, 46(11), 1956–1966.
- Hall, T. M., and Jewson, S. (2007). "Statistical modelling of North Atlantic tropical cyclone tracks." *Tellus A*, 59(4), 486–498.
- Harmeling, S. (2009). *Global Climate Risk Index 2010*. Germanwatch, Germany.
- Harmeling, S. (2010). *Global Climate Risk Index 2011*. Germanwatch, Germany.
- Harper, B. A. (1999). "Numerical modelling of extreme tropical cyclone winds." *Journal of Wind Engineering and Industrial Aerodynamics*, 83(1–3), 35–47.
- Harper, B. A., and Holland, G. J. (1999). "An updated parametric model of tropical cyclone." *Proceedings of 23rd Conference on Hurricanes and Tropical Meteorology*, American Meteorological Society, Dallas, TX.
- Ho, F. P., Schwerdt, R. W., and Goodyear, H. V. (1975). *Some climatological characteristics of hurricanes and tropical storms, Gulf and East Coast of the United States*. NOAA Technical Report, U.S. Department of Commerce.
- Ho, F. P., Su, J. C., and Hanevich, K. L. (1987). *Hurricane climatology for the Atlantic and Gulf coasts of the United States*. NOAA technical report NWS, U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, MD.
- Holland, G. J. (1980). "An Analytic Model of the Wind and Pressure Profiles in Hurricanes." *Monthly Weather Review*, 108(8), 1212–1218.
- Holland, G. J. (1997). "The Maximum Potential Intensity of Tropical Cyclones." *Journal of the Atmospheric Sciences*, 54(21), 2519–2541.
- Holland, G. J. (2008). "A Revised Hurricane Pressure–Wind Model." *Monthly Weather Review*, 136(9), 3432–3445.
- Holland, G. J., Belanger, J. I., and Fritz, A. (2010). "A Revised Model for Radial Profiles of Hurricane Winds." *Monthly Weather Review*, 138(12), 4393–4401.
- Holmes, J., Affairs, H., and Relief, E. (2005). "Hyogo Framework for Action 2005-2015: Building the Resilience of Nations and Communities to Disasters."
- Hope, J. R., and Neumann, C. J. (1971). *Digitized Atlantic Tropical Cyclone Tracks*.
- Hu, K., Chen, Q., and Fitzpatrick, P. (2012). "Assessment of a Parametric Hurricane Surface Wind Model for Tropical Cyclones in the Gulf of Mexico." *Advances in Hurricane Research - Modelling, Meteorology, Preparedness and Impacts*, K. Hickey, ed., InTech.
- Hu, K., Chen, Q., and Kimball, S. K. (2012). "Consistency in hurricane surface wind forecasting: an improved parametric model." *Natural Hazards*, 61(3), 1029–1050.
- Jagger, T., Elsner, J. B., and Niu, X. (2001). "A Dynamic Probability Model of Hurricane Winds in Coastal Counties of the United States." *Journal of Applied Meteorology*, 40, 853–863.
- James, M., and Mason, L. (2005). "Synthetic Tropical Cyclone Database." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 131(4), 181–192.

- Jarvinen, B. R. (1984). *A tropical cyclone data tape for the North Atlantic Basin, 1886-1983: contents, limitations, and uses*. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, [Washington, D.C.].
- JMA. (2014). "Best Track Data." <<http://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html>>.
- Kamahori, H., Yamazaki, N., Mannoji, N., and Takahashi, K. (2006). "Variability in Intense Tropical Cyclone Days in the Western North Pacific." *Sola*, 2, 104–107.
- Kaplan, J., and DeMaria, M. (1995). "A Simple Empirical Model for Predicting the Decay of Tropical Cyclone Winds after Landfall." *Journal of Applied Meteorology*, 34(11), 2499–2512.
- Kaplan, J., and Demaria, M. (2001). "On the Decay of Tropical Cyclone Winds after Landfall in the New England Area." *Journal of Applied Meteorology*, 40(2), 280–286.
- Kass, R. E., and Raftery, A. E. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90(430), 773–795.
- Knaff, J. A., and Sampson, C. (2006). "Reanalysis of West Pacific tropical cyclone intensity 1966-1987." Monterey.
- Knaff, J. A., and Zehr, R. M. (2007). "Reexamination of Tropical Cyclone Wind–Pressure Relationships." *Weather and Forecasting*, 22(1), 71–88.
- Knapp, K. R., and Kruk, M. C. (2009). "Quantifying Interagency Differences in Tropical Cyclone Best-Track Wind Speed Estimates." *Monthly Weather Review*, 138(4), 1459–1473.
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010a). "IBTrACS: The International Best Track Archive for Climate Stewardship."
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010b). "The International Best Track Archive for Climate Stewardship (IBTrACS)." *Bulletin of the American Meteorological Society*, 91(3), 363–376.
- Kossin, J. P., and Velden, C. S. (2004). "A Pronounced Bias in Tropical Cyclone Minimum Sea Level Pressure Estimation Based on the Dvorak Technique." *Monthly Weather Review*, 132(1), 165–173.
- Koutsoyiannis, D. (2012). "Clausius–Clapeyron equation and saturation vapour pressure: simple theory reconciled with practice." *European Journal of Physics*, 33(2), 295.
- Kruk, M. C., Knapp, K. R., and Levinson, D. H. (2009). "A Technique for Combining Global Tropical Cyclone Best Track Data." *Journal of Atmospheric and Oceanic Technology*, 27(4), 680–692.
- Kwun, J. H., Kim, Y.-K., Seo, J.-W., Jeong, J. H., and You, S. H. (2009). "Sensitivity of MM5 and WRF mesoscale model predictions of surface winds in a typhoon to planetary boundary layer parameterizations." *Natural Hazards*, 51(1), 63–77.
- Lander, M. A. (2008). "A comparison of typhoon best track data in the western North Pacific: irreconcilable differences."



- Landsea, C. (2010). "Tropical Cyclones Records." <<http://www.aoml.noaa.gov/hrd/tcfaq/E1.html>>.
- Lee, K., and Rosowsky, D. (2007). "Synthetic Hurricane Wind Speed Records: Development of a Database for Hazard Analyses and Risk Studies." *Natural Hazards Review*, 8(2), 23–34.
- Legg, M. R., Nozick, L. K., and Davidson, R. A. (2010). "Optimizing the selection of hazard-consistent probabilistic scenarios for long-term regional hurricane loss estimation." *Structural Safety*, 32(1), 90–100.
- Li, M., Zhong, L., Boicourt, W. C., Zhang, S., and Zhang, D.-L. (2006). "Hurricane-induced storm surges, currents and destratification in a semi-enclosed bay." *Geophysical Research Letters*, 33(2).
- Lin, N., Emanuel, K. A., Smith, J. A., and Vanmarcke, E. (2010). "Risk assessment of hurricane storm surge for New York City." *Journal of Geophysical Research: Atmospheres*, 115(D18), D18121.
- Luettich Jr, R. A., Westerink, J. J., and Scheffner, N. W. (1992). *ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL*. DTIC Document.
- Luong, N. O., Nguyen, T. T. T., Wilderspin, I., and Coulier, M. (2011). *A preliminary analysis of flood and storm disaster data in Viet Nam*. United Nations Development Programme.
- Malkin, W. (1959). *Filling and Intensity Changes in Hurricanes Over Land*. U.S. Weather Bureau.
- Mandal, M., Mohanty, U. C., and Raman, S. (2004). "A Study on the Impact of Parameterization of Physical Processes on Prediction of Tropical Cyclones over the Bay of Bengal With NCAR/PSU Mesoscale Model." *Natural Hazards*, 31(2), 391–414.
- Mattocks, C., and Forbes, C. (2008). "A real-time, event-triggered storm surge forecasting system for the state of North Carolina." *Ocean Modelling*, 25(3–4), 95–119.
- McConochie, J. D., Hardy, T. A., and Mason, L. B. (2004). "Modelling tropical cyclone over-water wind and pressure fields." *Ocean Engineering*, 31(14–15), 1757–1782.
- McDonald, J. H. (2009). *Handbook of biological statistics*. Sparky House Publishing Baltimore, MD.
- Miller, B. I. (1964). "A study of the filling of hurricane donna (1960) over land." *Monthly Weather Review*, 92(9), 389–406.
- Mitzenmacher, M., and Upfal, E. (2005). *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York.
- Natural Earth. (2014). "1:110m Cultural Vectors." <<http://www.naturalearthdata.com/downloads/110m-cultural-vectors/>> (Aug. 22, 2014).

- Neumann, C. J. (1991). *The National Hurricane Center Risk Analysis Program (HURISK)*. National Oceanic and Atmospheric Administration (NOAA), Washington, DC.
- Neumann, C. J. (1993). *Global Overview - Chapter 1: Global Guide to Tropical Cyclone Forecasting*. World Meteorological Organization, Geneva, Switzerland.
- Newell, R. E. (1973). *The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes - Vol. 1*. The MIT Press, Cambridge, Mass.
- NHC. (1997). "Summary of the NHC/TPC tropical cyclone track and intensity guidance models." <<http://www.nhc.noaa.gov/modelsummary.shtml>>.
- NHC. (2006). "An overview of NHC prediction models." <<http://www.srh.noaa.gov/ssd/nwpmmodel/html/nhcmodel.htm>>.
- NHC. (2014). "Saffir-Simpson Hurricane Wind Scale." <<http://www.nhc.noaa.gov/aboutsshws.php>>.
- NOAA. (2008). "THORPEX Pacific Asian Regional Campaign." <[http://www.aoml.noaa.gov/hrd/data\\_sub/hurr.html](http://www.aoml.noaa.gov/hrd/data_sub/hurr.html)>.
- NOAA. (2014a). "ESRL: PSD: NOAA Extended Reconstructed SST V3b." <<http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.html>> (Aug. 22, 2014).
- NOAA. (2014b). "Marine Safety." <<http://www.nhc.noaa.gov/prepare/marine.php>>.
- NOAA. (2014c). "Glossary of NHC Terms." <<http://www.nhc.noaa.gov/aboutgloss.shtml>>.
- Novlan, D. J., and Gray, W. M. (1974). "Hurricane-Spawmed Tornadoes." *Monthly Weather Review*, 102(7), 476–488.
- Ott, S. (2006). *Extreme Winds in the Western North Pacific*.
- Pasquero, C., and Emanuel, K. (2008). "Tropical Cyclones and Transient Upper-Ocean Warming." *Journal of Climate*, 21(1), 149–162.
- Pava, J., Fleites, F., Ruan, F., Chatterjee, K., Chen, S.-C., and Zhang, K. (2010). "A three-dimensional geographic and storm surge data integration system for evacuation planning." *2010 IEEE International Conference on Information Reuse and Integration (IRI)*, 181–188.
- Pielke, R., Gratz, J., Landsea, C., Collins, D., Saunders, M., and Musulin, R. (2008). "Normalized Hurricane Damages in the United States: 1900-2005." *Natural Hazards Review*, 9(1), 29–42.
- Powell, M. D. (1990). "Meteorological aspects of Hurricane Hugo." *Proceedings Hurricane Hugo One Year Later*, Charleston, S.C.
- Powell, M. D., Houston, S. H., Amat, L. R., and Morisseau-Leroy, N. (1998). "The HRD real-time hurricane wind analysis system." *J. Wind Engineer. and Indust. Aerodyn.*, (77), 53–64.
- Powell, M. D., Murillo, S., Dodge, P., Uhlhorn, E., Gamache, J., Cardone, V., Cox, A., Otero, S., Carrasco, N., Annane, B., and Fleur, R. S. (2010). "Reconstruction of Hurricane Katrina's wind fields for storm surge and wave hindcasting."

- Ocean Engineering*, A Forensic Analysis of Hurricane Katrina's Impact: Methods and Findings, 37(1), 26–36.
- Powell, M. D., Vickery, P. J., and Reinhold, T. A. (2003). "Reduced drag coefficient for high wind speeds in tropical cyclones." *Nature*, 422(6929), 279–283.
- Powell, M., Soukup, G., Cocke, S., Gulati, S., Morisseau-Leroy, N., Hamid, S., Dorst, N., and Axe, L. (2005). "State of Florida hurricane loss projection model: Atmospheric science component." *Journal of Wind Engineering and Industrial Aerodynamics*, 93(8), 651–674.
- Ren, F., Liang, J., Wu, G., Dong, W., and Yang, X. (2011). "Reliability Analysis of Climate Change of Tropical Cyclone Activity over the Western North Pacific." *Journal of Climate*, 24(22), 5887–5898.
- Riehl, H. (1954). *Tropical Meteorology*. McGraw-Hill.
- Rosendal, H. E., and Shaw, S. L. (1982). *Relationship Of Maximum Sustained Winds To Minimum Sea Level Pressure In Central North Pacific Tropical Cyclones*. NOAA Technical Memorandum, National Oceanic and Atmospheric Administration.
- Rumpf, J., Weindl, H., Höpfe, P., Rauch, E., and Schmidt, V. (2007). "Stochastic modelling of tropical cyclone tracks." *Mathematical Methods of Operations Research*, 66(3), 475–490.
- Russell, L. R. (1968). "Design windspeeds in tropical cyclone-prone regions." Ph.D. Thesis, Stanford University.
- Russell, L. R. (1971). "Probability Distributions for Hurricane Effects." *Journal of the Waterways, Harbors and Coastal Engineering Division*, 97(1), 139–154.
- Scheffner, N., Borgman, L., and Mark, D. (1996). "Empirical Simulation Technique Based Storm Surge Frequency Analyses." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 122(2), 93–101.
- Schloemer, R. W., United States, Weather Bureau, United States, Hydrologic Services Division, Hydrometeorological Section, United States, Army, and Corps of Engineers. (1954). *Analysis and synthesis of hurricane wind patterns over Lake Okechobee, Florida*. U. S. Department of Commerce, Weather Bureau, Washington, D.C.
- Schwerdt, R. W., Ho, F. P., and Watkins, R. R. (1979). *Meteorological criteria for standard project hurricane and probable maximum hurricane windfields, gulf and east coasts of the United States*. Dept. of Commerce, National Oceanic and Atmospheric Administration, [Office of Oceanic and Atmospheric Services], National Weather Service, Washington.
- Shapiro, L. J. (1983). "The Asymmetric Boundary layer Flow Under a Translating Hurricane." *Journal of the Atmospheric Sciences*, 40(8), 1984–1998.
- Shen, W., Ginis, I., and Tuleya, R. E. (2002). "A Numerical Investigation of Land Surface Water on Landfalling Hurricanes." *Journal of the Atmospheric Sciences*, 59(4).
- Song, J.-J., Wang, Y., and Wu, L. (2010). "Trend discrepancies among three best track data sets of western North Pacific tropical cyclones." *Journal of Geophysical Research: Atmospheres*, 115(D12), D12128.

- Southern, R. L. (1979). "The global socio-economic impact of tropical cyclones." *Australian Meteorol.*, 27, 175–195.
- Sparks, P. R., and Huang, Z. (1999). "Wind speed characteristics in tropical cyclones." *Proceedings of the 10th International Conference on Wind Engineering*, Copenhagen, Denmark.
- Sparks, P. R., and Huang, Z. (2001). "Gust factors and surface-to-gradient wind-speed ratios in tropical cyclones." *Journal of Wind Engineering and Industrial Aerodynamics*, 10th International Conference on Wind Engineering, 89(11–12), 1047–1058.
- Srifer, R. L., and Huber, M. (2007). "Observational evidence for an ocean heat pump induced by tropical cyclones." *Nature*, 447(7144), 577–580.
- Theil, H. (1961). *Economic forecasts and policy*. North-Holland Pub. Co.
- Thompson, E., and Cardone, V. (1996). "Practical Modeling of Hurricane Surface Wind Fields." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 122(4), 195–205.
- Tompkins, H. (2002). "Climate change and extreme weather events: Is there a connection?" *Cicerone*, 3.
- Tryggvason, B. V., Davenport, A. G., and Surry, D. (1976). "Predicting Wind-Induced Response in Hurricane Zones." *Journal of the Structural Division*, 102(12), 2333–2350.
- UN. (1982). "1982 United Nations Convention on the Law of the Sea | Centre for International Law."
- UNDP. (2004). *Reducing disaster risk: a challenge for development*. United Nations Development Programme.
- UNDP. (2007). *Human Development Report 2007/2008: Fighting Climate Change - Human Solidarity in a Divided World*. United Nations Development Programme.
- UNISDR. (2008). *Links between Disaster Risk Reduction, Development and Climate Change*. United Nation International Strategy for Disaster Reduction, Sweden.
- Vallis, G. K. (2006). *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*. Cambridge University Press, Cambridge.
- Velden, C., Harper, B., Wells, F., Beven, J. L., Zehr, R., Olander, T., Mayfield, M., Guard, C., Lander, M., Edson, R., Avila, L., Burton, A., Turk, M., Kikuchi, A., Christian, A., Caroff, P., and McCrone, P. (2006). "The Dvorak Tropical Cyclone Intensity Estimation Technique: A Satellite-Based Method that Has Endured for over 30 Years." *Bulletin of the American Meteorological Society*, 87(9), 1195–1210.
- Vickery, P. J. (2005). "Simple Empirical Models for Estimating the Increase in the Central Pressure of Tropical Cyclones after Landfall along the Coastline of the United States." *Journal of Applied Meteorology*, 44(12), 1807–1826.
- Vickery, P. J., Masters, F. J., Powell, M. D., and Wadhera, D. (2009). "Hurricane hazard modeling: The past, present, and future." *Journal of Wind*

- Engineering and Industrial Aerodynamics*, 12th International Conference on Wind Engineering, 97(7–8), 392–405.
- Vickery, P. J., Skerlj, P., Steckley, A., and Twisdale, L. (2000). "Hurricane Wind Field Model for Use in Hurricane Simulations." *Journal of Structural Engineering*, 126(10), 1203–1221.
- Vickery, P. J., Skerlj, P., and Twisdale, L. (2000). "Simulation of Hurricane Risk in the U.S. Using Empirical Track Model." *Journal of Structural Engineering*, 126(10), 1222–1237.
- Vickery, P. J., and Twisdale, L. A. (1995a). "Wind-Field and Filling Models for Hurricane Wind-Speed Predictions." *Journal of Structural Engineering*, 121(11), 1700–1709.
- Vickery, P. J., and Twisdale, L. A. (1995b). "Prediction of Hurricane Wind Speeds in the United States." *Journal of Structural Engineering*, 121(11), 1691–1699.
- Vickery, P. J., and Wadhera, D. (2008). "Statistical Models of Holland Pressure Profile Parameter and Radius to Maximum Winds of Hurricanes from Flight-Level Pressure and H\*Wind Data." *Journal of Applied Meteorology and Climatology*, 47(10), 2497–2517.
- Vickery, P. J., Wadhera, D., Powell, M. D., and Chen, Y. (2009). "A Hurricane Boundary Layer and Wind Field Model for Use in Engineering Applications." *Journal of Applied Meteorology and Climatology*, 48(2), 381–405.
- Vickery, P. J., Wadhera, D., Twisdale, L., and Lavelle, F. (2009). "U.S. Hurricane Wind Speed Risk and Uncertainty." *Journal of Structural Engineering*, 135(3), 301–320.
- Wang, Y., and Rosowsky, D. V. (2012). "Joint distribution model for prediction of hurricane wind speed and size." *Structural Safety*, 35, 40–51.
- Weisberg, R. H., and Zheng, L. (2006). "Hurricane storm surge simulations for Tampa Bay." *Estuaries and Coasts*, 29(6), 899–913.
- Whittingham, H. E. (1958). *The Bathurst Bay hurricane and associated storm surge*. Bureau of Meteorology, Melbourne, Vic.
- Wikipedia. (2014a). "List of the most intense tropical cyclones." *Wikipedia, the free encyclopedia*.
- Wikipedia. (2014b). "Saffir–Simpson hurricane wind scale." *Wikipedia, the free encyclopedia*.
- Willoughby, H. E., Darling, R. W. R., and Rahn, M. E. (2006). "Parametric representation of the primary hurricane vortex. Part II: A new family of sectionally continuous profiles." *Monthly weather review*, 134(4), 1102–1120.
- Willoughby, H. E., and Rahn, M. E. (2004). "Parametric Representation of the Primary Hurricane Vortex. Part I: Observations and Evaluation of the Holland (1980) Model." *Monthly Weather Review*, 132(12), 3033–3048.
- Wisner, B., Blaikie, P., Cannon, T., and Davis, I. (2003). *At Risk: Natural Hazards, People's Vulnerability and Disasters*. Routledge, London ; New York.
- World Vision. (2009). *Reduce Risk and Raise Resilience*. Climate Change Series, World Vision International, Australia.

- Wu, M.-C., Yeung, K.-H., and Chang, W.-L. (2006). "Trends in western North Pacific tropical cyclone intensity." *Eos, Transactions American Geophysical Union*, 87(48), 537–538.
- Xie, L., Bao, S., Pietrafesa, L. J., Foley, K., and Fuentes, M. (2006). "A Real-Time Hurricane Surface Wind Forecasting Model: Formulation and Verification." *Monthly Weather Review*, 134(5), 1355–1370.
- Xie, L., Liu, H., and Peng, M. (2008). "The effect of wave–current interactions on the storm surge and inundation in Charleston Harbor during Hurricane Hugo 1989." *Ocean Modelling*, 20(3), 252–269.
- Zhang, H.-M., Bates, J. J., and Reynolds, R. W. (2006). "Assessment of composite global sampling: Sea surface wind speed." *Geophysical Research Letters*, 33(17).

## LIST OF TABLES

TABLE 1.1 TOP FIVE DEADLIEST EXTREME WEATHER EVENTS, 1970-2001 [FROM (TOMPKINS 2002)].....	2
TABLE 1.2 TOP FIVE COSTLIEST EXTREME WEATHER EVENTS, 1970-2001 [FROM (TOMPKINS 2002)].....	2
TABLE 1.3 ANNUAL GLOBAL DEATHS AND DEATH RATES FOR VARIOUS CATEGORIES OF DISASTERS, 1900-1989 AND 1990-2008 [FROM (GOKLANY 2009)].....	4
TABLE 1.4 TOP FIVE MOST EXTREME TCs [ADAPTED FROM (WIKIPEDIA 2014A)] .....	9
TABLE 1.5 STRONGEST TCs BY BASINS [ADAPTED FROM (WIKIPEDIA 2014A)] .....	9
TABLE 1.6 ANNUAL TC OCCURRENCE BY BASINS [FROM (NEUMANN 1993)] .....	10
TABLE 2.1 TC'S CENTRES FOR THE CASE STUDY OF VIETNAM.....	27
TABLE 2.2 SUMMARY OF VARIOUS SHAPES OF COMPUTATIONAL GRIDS .....	30
TABLE 3.1 ANALYSIS OF HISTORICAL DATA (TRACK MODEL) .....	40
TABLE 3.2 EXCLUSION OF DPs IN THE TA (TRACK MODEL) .....	41
TABLE 3.3 SCENARIOS BASED ON HEADINGS (TRACK MODEL) .....	43
TABLE 3.4 OPTIONS FOR LINEAR REGRESSION SOLUTION (TRACK MODEL) .....	44
TABLE 3.5 POSSIBLE COMBINATIONS OF AVAILABLE PARAMETERS TO DETERMINE THE CHANGES IN TRANSLATION SPEED .....	46
TABLE 3.6 POSSIBLE COMBINATIONS OF AVAILABLE PARAMETERS TO DEFINE THE CHANGES IN HEADING .....	47
TABLE 3.7 COMPARISON BETWEEN 5° SQUARES AND 7° HEXAGONS (TRACK MODEL).....	48
TABLE 3.8 ANALYSIS FOR 6° HEXAGONS (TRACK MODEL).....	49
TABLE 3.9 ANALYSIS FOR 8° HEXAGONS (TRACK MODEL).....	50
TABLE 3.10 SUMMARY OF SCENARIOS FOR COMPUTATIONAL GRIDS (TRACK MODEL) .....	50
TABLE 3.11 COMPARISON BETWEEN APPROACHES UTILIZED IN THE ETM AND THE GSESM FOR THE CASE STUDY OF VIETNAM (TRACK MODEL) .....	54
TABLE 3.12 ANALYSIS OF HISTORICAL DATA (INTENSITY MODEL).....	64
TABLE 3.13 EXCLUSION OF DATA POINTS IN THE TA (INTENSITY MODEL) .....	64
TABLE 3.14 SCENARIOS BASED ON HEADINGS (INTENSITY MODEL).....	65
TABLE 3.15 OPTIONS FOR LINEAR REGRESSION SOLUTION (INTENSITY MODEL).....	66
TABLE 3.16 5° SQUARES COMPARE WITH 7° HEXAGONS (INTENSITY MODEL).....	67
TABLE 3.17 ANALYSIS FOR 6° HEXAGONS (INTENSITY MODEL) .....	68
TABLE 3.18 ANALYSIS FOR 8° HEXAGONS (INTENSITY MODEL) .....	69
TABLE 3.19 SUMMARY OF SCENARIOS FOR COMPUTATIONAL GRIDS (INTENSITY MODEL) .....	69
TABLE 3.20 COMPARISON BETWEEN APPROACHES UTILIZED IN THE ETM AND THE GSESM FOR THE CASE STUDY OF VIETNAM (INTENSITY MODEL) .....	71
TABLE 3.21 ANALYSIS OF DPs FOR USE IN A DECAY MODEL .....	74

TABLE 3.22 ANALYSIS OF NONLINEAR REGRESSION IN THE DECAY MODEL.....	75
TABLE 3.23 COMBINATIONS OF USABLE TERMS (DECAY RATE COEFFICIENT MODEL) .....	77
TABLE 3.24 OPTIONS FOR REGRESSION SOLUTION (DECAY RATE COEFFICIENT MODEL) .....	77
TABLE 3.25 COMPARISON BETWEEN THE GSESM' EQUATION AND VARIOUS EMPIRICAL FORMULAE FOR THE CASE STUDY OF VIETNAM (DECAY RATE COEFFICIENT MODEL) .....	77
TABLE 4.1 EXAMPLE VALUES OF WIND SPEED REDUCTION FACTORS [ADAPTED FROM (VICKERY, MASTERS, ET AL. 2009)].....	86
TABLE 4.2 ANALYSIS OF HISTORICAL DATA (WIND FIELD MODEL).....	87
TABLE 4.3 EXCLUSION OF DPs IN THE TA (WIND FIELD MODEL) .....	87
TABLE 4.4 WIND FIELD CALCULATIONS .....	89
TABLE 4.5 COMBINATIONS OF AVAILABLE TERMS (RADIUS OF MAXIMUM WIND MODEL) .....	92
TABLE 4.6 OPTIONS FOR REGRESSION SOLUTION (RADIUS OF MAXIMUM WIND MODEL).....	92
TABLE 4.7 COMPARISON BETWEEN THE GSESM' APPROACH AND VARIOUS EMPIRICAL FORMULAE FOR THE CASE STUDY OF VIETNAM (RADIUS OF MAXIMUM WIND MODEL) .....	94
TABLE 4.8 COMBINATIONS OF AVAILABLE TERMS (HOLLAND PARAMETER MODEL) .....	96
TABLE 4.9 OPTIONS FOR A REGRESSION SOLUTION (HOLLAND PARAMETER MODEL) .....	96
TABLE 4.10 COMPARISON BETWEEN THE GSESM' APPROACH AND VARIOUS EMPIRICAL FORMULAE FOR THE CASE STUDY OF VIETNAM (HOLLAND PARAMETER MODEL) .....	97
TABLE 5.1 BIC DIFFERENCE CRITERION [ADAPTED FROM (KASS AND RAFTERY 1995)] .....	101
TABLE 5.2 AIC DIFFERENCE CRITERION [ADAPTED FROM (BURNHAM AND ANDERSON 2013)] .....	101
TABLE 5.3 COMPARISON OF FIT INDICES (ANNUAL OCCURRENCE RATE).....	104
TABLE 5.4 NUMBER OF DPs USED IN MODEL VALIDATION AND MODEL DEVELOPMENT.....	109
TABLE 5.5 SAFFIR-SIMPSON HURRICANE SCALE [FROM (VICKERY, SKERLJ, AND TWISDALE 2000)] .....	111
TABLE 6.1 SUMMARY OF THE GSESM'S IMPROVEMENTS OVER THE ETM.....	127



## LIST OF FIGURES

FIGURE 1.1 DEATHS AND DEATH RATES DUE TO HURRICANES IN THE U.S., 1900–2006 [FROM (GOKLANY 2009)] .....	5
FIGURE 1.2 GLOBAL TC BASINS [FROM (CAMARGO ET AL. 2005)] .....	9
FIGURE 1.3 GLOBALLY RECORDED TC'S ACTIVITY [FROM (KNAPP ET AL. 2010A)] .....	10
FIGURE 1.4 GEOGRAPHIC EXTENT OF VIETNAM AND THE NEARBY WATER AREAS .....	12
FIGURE 2.1 NUMBER OF AVAILABLE BTD SOURCES [FROM (KRUK ET AL. 2009)] .....	16
FIGURE 2.2 INTERAGENCY DIFFERENCES IN ESTIMATIONS OF TRACK AND INTENSITY, TC PEKE (1987) [FROM (KNAPP ET AL. 2010A)] .....	17
FIGURE 2.3 DISCREPANCIES (IN PERCENTAGE) IN OBSERVED SURFACE MAXIMUM SUSTAINED WINDS BETWEEN DIFFERENT AGENCIES [FROM (KNAPP ET AL. 2010A)] .....	17
FIGURE 2.4 COMPARISONS BETWEEN DATA DERIVED FROM BTD SETS AND REFERENCES FOR DIFFERENT TCs [FROM (BARCIKOWSKA 2012)] .....	18
FIGURE 2.5 THE AOI AND THE TA FOR THE QUEENSLAND COAST [DEFINED BY JAMES AND MASON (2005)] .....	20
FIGURE 2.6 THE AOI FOR THE CASE STUDY OF VIETNAM .....	21
FIGURE 2.7 THE TA FOR THE STATE OF FLORIDA [DEFINED BY (POWELL ET AL. 2005)] .....	22
FIGURE 2.8 RULE OF TC AVOIDANCE [FROM (NOAA 2014B)] .....	24
FIGURE 2.9 SAFFIR-SIMPSON HURRICANE SCALE [FROM (WIKIPEDIA 2014B)] .....	24
FIGURE 2.10 FLOW CHART OF THE APPROACH TO DEFINE THE TA FOR THE CASE STUDY .....	25
FIGURE 2.11 TCs IN THE WNP OF THE 1982 SEASON (ZOOM-OUT MAP) .....	26
FIGURE 2.12 TCs IN THE WNP OF THE 1982 SEASON (ZOOM-IN MAP) .....	27
FIGURE 2.13 SEVERAL ALTERNATIVES FOR THE SHAPES OF COMPUTATIONAL GRIDS .....	29
FIGURE 3.1 EXAMPLE MAP SHOWING PRINCIPLES TO DEFINE THE SUBSTITUTIVE GRIDS .....	40
FIGURE 3.2 COMPUTATIONAL GRIDS AND DOMAIN FOR THE CASE STUDY OF VIETNAM (TRACK MODEL) .....	51
FIGURE 3.3 OPTIMIZATION FLOW CHART (TRACK MODEL) .....	53
FIGURE 3.4 SEASONAL VARIATION IN THE ENVIRONMENTAL PRESSURE .....	59
FIGURE 3.5 SEASONAL VARIATION IN THE CENTRAL PRESSURE DEFICIT .....	60
FIGURE 3.6 SEASONAL VARIATION IN THE TC OCCURRENCE .....	60
FIGURE 3.7 OPTIMIZATION FLOW CHART (INTENSITY MODEL) .....	63
FIGURE 3.8 COMPUTATIONAL GRIDS AND DOMAIN FOR THE CASE STUDY OF VIETNAM (INTENSITY MODEL) .....	70
FIGURE 3.9 OBSERVED AND ESTIMATED DECAY AFTER LANDFALL, TC NANCY (1982) .....	75
FIGURE 4.1 DECOMPOSITION OF WIND SPEED AT GRADIENT-LEVEL [FROM (WANG AND ROSOWSKY 2012)] .....	81
FIGURE 4.2 AN EXAMPLE OF RADIAL PROFILES OF A TC WIND, HURRICANE KATRINA (2005) [FROM (WANG AND ROSOWSKY 2012)] .....	82

FIGURE 4.3 FLOW CHART (WIND FIELD MODEL) .....	90
FIGURE 5.1 FLOW CHART (SIMULATION) .....	99
FIGURE 5.2 COMPARISON OF PMF (ANNUAL OCCURRENCE RATE) .....	102
FIGURE 5.3 COMPARISON OF CDF (ANNUAL OCCURRENCE RATE) <sup>23</sup> .....	102
FIGURE 5.4 STARTING POSITIONS OF OBSERVED TCS .....	105
FIGURE 5.5 COMPARISON OF LMI OVER THE ENTIRE AOI.....	110
FIGURE 5.6 COMPARISON OF MONTHLY OCCURRENCE RATE OF THE ITCs OVER THE AOI.....	111
FIGURE 5.7 COMPARISON OF ANNUAL OCCURRENCE RATE OF THE ITCs OVER THE AOI.....	112
FIGURE 5.8 GRIDS FOR MODEL VALIDATION.....	113
FIGURE 5.9 TRACK MODEL VALIDATION .....	115
FIGURE 5.10 CENTRAL INTENSITY MODEL VALIDATION .....	116
FIGURE 5.11 COMPARISON OF LMI OVER THE AOI FOR DIFFERENT SIMULATION LENGTH .....	117
FIGURE 5.12 N-YEAR RETURN PERIOD WIND SPEED AT EACH GRID WITHIN THE AOI.....	118
FIGURE 5.13 WIND SPEED MAP .....	120
FIGURE 5.14 ANNUAL EXCEEDANCE PROBABILITY MAP .....	122

## LIST OF ABBREVIATIONS

The following list summarises the main abbreviations that have been used in this dissertation. All individual abbreviations have been defined in the text of this dissertation and those definitions are also repeated the first time they appear in each chapter.

AIC	Akaike Information Criterion
AOI	Area Of Interest
BIC	Bayesian Information Criterion
BTD	Best Track Data
CDF	Cumulative Distribution Function
DDT	Damage Distance Threshold
DP	Data Point
ETM	Empirical Track Modelling
GSESM	Generalized Stochastic Empirical Storm Model
JMA	Japan Meteorological Agency
LMI	Lifetime Maximum Intensity
OLS	Ordinary Least Squares
PDF	Probability Density Function
PMF	Probability Mass Function
TA	Threat Area
TC	Tropical Cyclone
WNP	Western North Pacific

## ACKNOWLEDGEMENTS

If a Ph.D. degree is a treasure, then it must be hidden in a paradisiacal island somewhere in the middle of an infinite ocean. At the beginning of the journey, everything seems so fresh and exciting that it makes you believe the research is truly the pursuit of happiness. However, as time passes by, you realize that the way to achieve your dream is not so enjoyable as you initially thought. Sometimes, you find yourself alone in the middle of nowhere, without a clear direction to move forward. Unfortunately, this is especially the case when I devote my youth to develop a numerical model. Most of the time, it's just me and the computer. These lonely moments may deter my will and drag me away from my delicious achievement, especially if I do not have my loved ones right beside me to constantly advise, support, and encourage me to stay in the right track.

My story began 8 years ago, when I conquered an international English certificate, which was quite a challenge for me at that moment. Then fate smiled upon me as I got a scholarship to study in the Netherlands and to start the voyage of discovery of knowledge that, now I cannot believe, took me such a long time. There are many people that influenced my path, from the beginning to this end. And because I cannot list all their names here, it does not mean that I will forget them or do not acknowledge all the valuable things that they have done for me. I would like to thank them all and they really know who they are.

First and foremost, I would like to thank my promoter, who is also my daily supervisor, Prof. Pieter van Gelder for his permanent support. Although he has been so busy all the time that we could not meet each other daily as it supposed to be, he still adjusted his bustling schedule to accommodate a discussion whenever I really needed one. Until now, after 8 years, I still remember his first lesson in statistics that allured me strongly, and remember how he praised me when I solved a mathematical problem in a special way. That all encouraged me much. Pieter, I am grateful to you for all your help.

I would like to thank my promoter, Prof. Han Vrijling for his precious time and energy spent to advise and support my work. I continue the list to acknowledge my professor at UNESCO-IHE, when I was a master student, Prof. Dano Roelvink, for his important guidance that directly led me to the theoretical framework used in this research. I would like to thank Mariette van Tilburg for her hard work to revise my writing. Mariette, thanks for your kind and timely support at the exact moment when I really needed it, and please keep working with your independent consultancy - M.A.Q.T. (M. van Tilburg Academic Quality Text).

I would like to send special thanks to the committee members for their significant presence at my defence ceremony and their valuable knowledge that contributes to the contents of this dissertation.

Most importantly, I thank my friends and family for their sweet love and deep care. Without you, I cannot go that far.

# CURRICULUM VITAE

Surname: NGUYEN  
First name: BINH MINH  
Date of birth: 05<sup>th</sup> July, 1982  
Place of birth: Hanoi, Vietnam  
Email: nguyenbinhminh@me.com

## **EDUCATION**

1997-2000 Gifted High school belongs to Hanoi University of Science,  
a member of Vietnam National University, Vietnam

2000-2005 Bachelor of Engineering,  
Water Resources University, Vietnam

2007-2009 Master of Science in Water Science and Engineering,  
UNESCO-IHE Institute for Water Education, the Netherlands

2010-2015 Ph.D. Candidate, Faculty of Civil Engineering and Geosciences,  
Delft University of Technology, the Netherlands

## **PROFESSIONAL EXPERIENCE**

Since 2006 Lecturer, Faculty of Hydraulic Engineering,  
Water Resources University, Vietnam

## LIST OF PUBLICATIONS

Nguyen, B.M., Roelvink, J.A., and Gelder, P.H.A.J.M. van (2011). "Empirical modelling of storm processes." *Proceedings of the 3rd international conference of European Asian civil engineering forum (EACF) 2011* (pp. 189-194). Yogyakarta, Indonesia.

Nguyen, B.M. (2010). "Estimating main storm parameters using an extended empirical track model in Vietnam." *Proceedings of the 6th risk and design symposium* (pp. 37-42). Delft, the Netherlands.

Nguyen, B.M. and Gelder, P.H.A.J.M. van (2012). "General hurricane track and intensity model." In F Klijn & T Schweckendiek (Eds.), *Comprehensive flood risk management: Research for policy and practice, Proceedings of the 2nd European conference on floodrisk management FLOODrisk2012* (pp. 95-101). London: CRC Press/Balkema.

Nguyen, B.M. and Gelder, P.H.A.J.M. van (2013). "Hurricane risk assessment in a long temporal scale." In D Hardjito & Antoni (Eds.), *International conference of The Euro-Asia civil engineering forum* (pp. 1-6). Surabaya: Institute for research and community service. Singapore.

Nguyen, B.M. and Gelder, P.H.A.J.M. van (2011). "Hurricane simulation at local level." *Proceedings of the 7th risk and design symposium* (pp. 89-96). Delft, the Netherlands.

Nguyen, B.M. and Gelder, P.H.A.J.M. van (2012). "Investigation of long-term hurricane activity." In NQ Kim (Ed.), *Proceedings of the fourth international conference on estuaries and coasts, volume 2 (ICEC 2012)* (pp. 306-313). Science and technics publishing house. Hanoi, Vietnam.

Nguyen, B.M. and Gelder, P.H.A.J.M. van (2012). "Risk estimate in storm-prone areas." In D Hanea & S Sillem (Eds.), *Proceedings of the risk and design symposium* (pp. 107-113). Delft: Faculteit TBM, Risk Centre.