Extreme Storm Surge Return Period Prediction Using Tidal Gauge Data and Estimation of Damage to Structures from Storm-Induced Wind Speed in South Korea

Sang Guk Yum

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University
2019

© 2019
Sang Guk Yum
All rights reserved

Abstract

Extreme Storm Surge Return Period Prediction Using Tidal Gauge Data and Estimation of Damage to Structures from Storm-Induced Wind Speed in South Korea

Sang Guk Yum

Global warming, which is one of the most serious consequence of climate change, can be expected to have different effects on the atmosphere, the ocean, icebergs, etc. Global warming has also brought secondary consequences into nature and human society directly. The most negative effect among the several effects of global warming is the rising sea level related to the large typhoons which can cause flooding on low-level land, coastal invasion, sea water flow into rivers and underground water, rising river level, and fluctuation of sea tides. It is crucial to recognize surge level and its return period more accurately to prevent loss of human life and property damage caused by typhoons.

This study researches two topics. The first purpose of this study is to develop a statistical model to predict the return period of the storm surge water related to typhoon Maemi, 2003 in South Korea. To estimate the return period of the typhoon, clustered separated peaks-over-threshold simulation (CSPS) has been used and Weibull distribution is used for the peak storm surge height's fitting. The estimated return period of typhoon Maemi's peak total water level is 389.11 years (95% confidence interval 342.27 - 476.2 years).

The second aim is related to the fragility curves with the loss data caused by typhoons. Although previous studies have developed various methods to mitigate damages from typhoons, the extent of financial loss has not been investigated enough. In this research, an insurance company provides their loss data caused by the wind speed of typhoon Maemi in 2003. The loss

data is very important in evaluating the extent of the damages. In this study, the damage ratio in the loss dataset has been used as the main indicator to investigate the extent of the damages. The damage ratio is calculated by dividing the direct loss by the insured amount.

In addition, this study investigates the fragility curves of properties to estimate the damage from typhoon Maemi in 2003. The damage ratios and storm induced wind speeds are used as the main factor for constructing fragility curves to predict the levels of damage of the properties. The geographical information system (GIS) has been applied to produce properties' spatial wind speeds from the typhoon. With the damage ratios, wind speeds and GIS spatial data, this study constructs the fragility curves with four different damage levels (Level I - Level IV). The findings and results of this study can be basic new references for governments, the engineering industry, and the insurance industry to develop new polices and strategies to cope with climate change.

Table of Contents

List of Figures

List of Tables

Acknowledgements

During my time at Columbia University, I have learned the most pivotal and essential lessons in both my academic and personal lives. The last six years of doctoral studies has not been an easy time in my life. There have been unexpected obstacles. However, those obstacles were eliminated by many people in different ways.

First, I would like to express boundless gratitude to my academic father Dr. George Deodatis. From my very first semester, he watched over me and my work. He was there for me both when I succeeded and when I failed. His warm-heartedness and encouragement helped me overcome many frustrations and difficulties, and his support and guidance on the dissertation led me to where I stand now. Without his thoughtful and thorough advising, I would not have completed my doctoral studies. His approach to academics, students, and life is one I would like to emulate.

I would also like to express sincere thanks to my committee members: Dr. Raimondo Betti, who improved this study by offering invaluable comments and personal advice at key stages in my research; Dr. Maria Feng and Dr. Simos Gerasimidis, who provided invaluable insights that enriched my dissertation; and Dr. Ibrahim Odeh, who supported and encouraged me in various ways to finish this far from solitary journey. Scott Kelly, the administrator of the department, always helped me secure financial resources and navigate the many different procedures and logistics of graduate school.

I should also thank my colleagues who I first met at Texas A&M University for their encouragement, support, and continuing good friendship: Dr. Ji-Myoung Kim at Mokpo University, South Korea; Dr. Junseo Bae at University of the West of Scotland, Scotland; and Dr. Youngjae Kim at Yeungnam University, South Korea. My Columbia colleagues, Jinwoo Jang of Florida Atlantic University and Sung-Hwan Jang of the University of Plymouth, made my time in New York productive as we shared doctoral life together. Sujin Lee and Seungmin Lim, who were postdocs at Columbia during my time there, shared their keen insights and knowledge. I cannot skip Vincent who supported me and continued his robust friendship even after graduating before me. It is also my great fortune to count Muqing, Lei, Ma Siwei, and Apostolos as my peers and friends. With them, life at Columbia was good.

Beyond Columbia, I have many people to thank. Life in New York City would be lonely and unbearable without Hyejo, Sojung, Jeewon, Dongkyun, Seungwhan, Donghyuk, Seokbae, and the rest of the Franciscan Young Adult Choir in New York. Hanging out and singing with them weekly were the most memorable moments in my life, as was my time playing soccer with Jiho and Seohyun whose friendships I hope will continue into the future. Closer to home, I would like to thank my family: Dr. KyungTaek Yum, Youngae Nam, Sangheon Yum, and Sukyoung Woo for their support and patience during my doctoral studies. Without their prayers, I would not be in the fortunate position to even write this acknowledgement.

Finally, I like to say thank you to my girlfriend Hyunju. Love you so much. I am blessed to have you in my life.

Dedication

To

Kyungtaek Yum
Youngae Nam
Sangheon Yum
Sukyoung Wu

Chapter 1

1 Introduction

1.1 Climate change and global warming

Global warming, which is one of the biggest issues of climate change, can affect the atmosphere, ocean, icebergs, etc. through various routes. In addition to this, global warming has secondary consequences for nature and human society. One of the negative effects among the several effects of global warming is sea level rise. Sea level rise can cause flooding on low-level land, coastal invasion, sea water flow into rivers and underground water, river level rise, and tidal changes and fluctuations of the sea.

According to recent research regarding hurricanes, the intensity and frequency of typhoons and hurricanes caused by global warming are continuously changing. As a consequence, natural hazards such as hurricanes could negatively affect natural systems, water resources, transportation, and infrastructure facilities. The importance of a highly accurate forecasting system for protecting human society has been emphasized. The inaccurate estimation of return periods of natural hazards can waste money and time on constructing waterfront facilities. Moreover, it could cause tremendous damage to society. The combined effect of sea level rise and tropical storms is an even more catastrophic hazard.

1.1.1 Sea level rise

According to the Intergovernmental Panel on Climate Change (IPCC, 2007) [18]'s recently released report, the average temperature in the world has increased 0.74°C (0.56°C - 0.92°C) during 1906-2005 [17]. The Fourth Assessment Report (AR4, 2007) by IPCC reported that since 1961, the increasing mean sea level over the world is 1.8 (1.3 – 2.3) mm/year. When melting icebergs are taken into account, the increasing sea level is 3.1 (2.4 – 3.8) mm/year [17]. Moreover, the arctic ice has decreased 2.7% annually since 1978, and the amount of snow in the mountains has declined (Kim and Cho, 2013 [27]). As a result, there is growing interest in how much sea level will increase, and relevant research is being conducted on how to cope with changes in the climate (National Academy of Science, 2012 [36]; Radic and Hock, 2011 [45]; Schaeffer et al., 2012 [47]). Most industrial facilities on the Korean Peninsula, such as plants, ports, roads, ship yards, and housing are located near the shore. Those topographical characteristics make big cities in South Korea especially susceptible to sea level rise. If rising sea levels begin to flood big cities, socioeconomic losses will be large.

1.1.2 Sea level rise in the city of Busan in South Korea

Yoon and Kim, 2012 [56] investigated long-term sea level changes using data from 17 tidal stations located around the Korean Peninsula with more than 25 years of tide-gauge data. Regression analysis was applied to calculate the general rising trend for long-term change (1960-–2010) in mean sea level (MSL) at each station. The results showed that rising trends of MSL around Korea is higher than that of global MSL. The linear rising trend of MSL was relatively small along the western coast (on average 1.3mm/year), large along the southern and eastern coasts (on average 3.2 and 2.0 mm/year, respectively), and very large around Jeju Island (on

average 5.6 mm/year). The rising rate around Jeju Island is about three times higher compared to the world's oceans. According to the AR4, the rate of rising sea level may accelerate after the 21st century. The trends in rising sea level should be considered when designing coastal structures in terms of coastal disaster prevention. Therefore, more accurate predictions of variation in sea level and surge height is required in areas affected by climate change. In the present study, Busan, a city in South Korea, has been used as a case study. Busan's rising sea level is 1.8mm/year according to a regression analysis based on data from 1960 to 2010 (Yoon and Kim, 2012) [56]. This rate of increase is similar with the rate of rising sea level over the world.

1.2 Tidal gauge data

1.2.1 The trend of typhoons in South Korea

The Korean Peninsula is surrounded by three different seas. This geographical characteristic can lead to severe damage to coastal regions. According to the Korea Ocean Observing and Forecasting System, Typhoon Maemi in September 2003 had a maximum wind speed of 54 m/s, which caused enormous damage to properties ($35 billion) as showed in Table 1.1. Also, all three highest peaks from tidal gauge stations were recorded in September 2003.

Name	Date	Amount of Damage (USD Dollar)	Max. Wind Speed (10 min avg. m/s)
Rusa	08/30/2002 – 09/01/2002	4.3 B (1^{th})	41
Maemi	09/12/2003 – 09/13/2003	3.5 B (2^{nd})	54
Bolaven	08/25/2012 – 08/30/2012	0.9 B (3^{th})	53

Table 1.1 Largest typhoons in the Korean Peninsula

The World Meteorological Organization (WMO) categorizes typhoons, or hurricanes, based on Table 1.2 below. Typhoons are classified into four levels according to the maximum wind speed (10 min sustained) recorded. Note that South Korea and Japan distinguish typhoons from tropical storms, classifying storms with a maximum wind speed above 17 m/s as typhoons.

Intensity Class	International Category	South Korean and Japanese Standard	Maximum Sustained Wind Speed (10 min mean)	Class
Tropical Depression	**TD**: Tropical Depression	**TD**: Tropical Depression	~ 17.2m/s	2
Typhoon	**TS**: Tropical Storm	**TY**: Typhoon	17.2m/s ~ 24m/s	3
	STS: Severe Tropical Storm		25m/s ~ 32m/s	4
Strong Typhoon	**TY**: Typhoon		33m/s ~ 43m/s	5
Very Strong Typhoon			44m/s ~ 53m/s	
Violent Typhoon			54m/s ~	

Table 1.2 Classification of intensity of typhoons in South Korea, National Typhoon Center [38]

According to a typhoon technical report published by the National Typhoon Center, 2011 [38], for typhoons that do not make landfall, the following criteria must be met for the storm to be classified as a typhoon with indirect effects:

- the storm in question must be between 32°N and 40°N latitude and 120°E and 138°E longitude; also the below two criteria define indirect effect typhoon.
- if the storm is located south of the Korean Peninsula, it must have at least an 8 second swell.
- if located northeast of the Korean Peninsula, it must have maximum wind speed of at least 20 m/s (10 min sustained), and at least 100 mm of rain.

The Korea Ocean Observing and Forecasting System reported that 1,019 typhoons have approached the Korean Peninsula, 327 typhoons have significantly impacted it, and 11 typhoons have made landfall in South Korea since 1904. Statistically, approximately three typhoons impact the Korean Peninsula annually. The most typhoon-heavy months, in descending order, are August, July, and September. Also, 66 % of the total number of typhoons occurs in July and August. Table 1.3 below shows statistics of typhoons in South Korea. Figure 1.1 shows the tracks of typhoons that made landfall in the Korean Peninsula (National Typhoon Center, 2011) [38]. The figure only shows the tracks of typhoons from 1977 to 2015. Before 1977, there were no systematic methods for tracking typhoons. As seen in Figure 1.1, all typhoons originate in South Asia and move northward. Figure 1.2 shows the track of Typhoon Maemi from September 4 to September 16, 2003. The differently colored dots in the right side figure indicate the changing storm status of the typhoon. As seen in Figures 1.1 and 1.2, Typhoon Maemi passed south east of Busan and the Korean Peninsula, causing direct damage upon landfall. When the typhoon made landfall at Busan, the maximum wind speed (10 min sustained) was 54 m/s.

Typhoon Maemi led the insurance industry, the Korean government, and many academic researchers to recognize the importance of prevention plans with regards to natural disasters such as typhoons.

Y/M	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Sum
Sum for 54 years (1952-2015)	29	15	15	45	67	115	245	351	322	238	152	73	1678
Sum of landfall for 54 years (1952-2015)	0	0	0	0	1	18	65	70	45	5	0	0	206
Average for 54 years (1952-2015)	0.54	0.28	0.46	0.83	1.24	2.13	4.54	6.52	5.96	4.41	2.81	1.35	
Average of landfall for 54 years (1952-2015)	0.0	0.0	0.0	0.0	0.02	0.33	1.2	1.3	0.87	0.09	0.0	0.0	
Sum of recent 10 years (2007-2016)	4	3	4	5	12	18	33	43	56	34	16	7	235
Sum of landfall for recent 10 years (2007-2016)	0	0	0	0	0	0	3	11	7	5	2	0	28
Average for recent 10 years (2007-2016)	0.4	0.3	0.4	0.5	1.2	1.8	3.3	4.3	5.6	3.4	1.6	0.7	
Average of landfall for recent 10 years (2007-2016)	0	0	0	0	0	0	0.3	1.1	0.7	0.5	0.2	0	

Table 1.3 Statistics of typhoons in South Korea, National Typhoon Center [38]

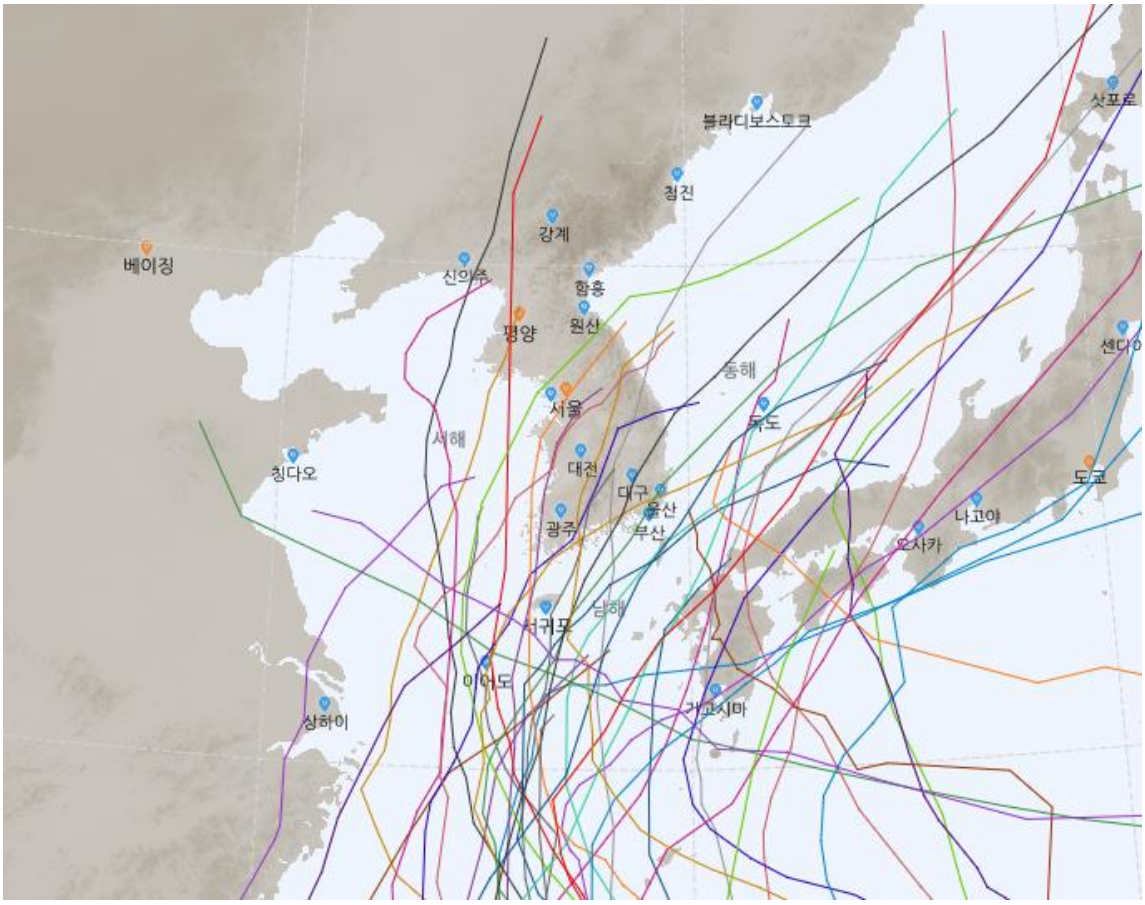

Figure 1.1 Landfall typhoons in South Korea (1977 - 2015), National Typhoon Center [38]

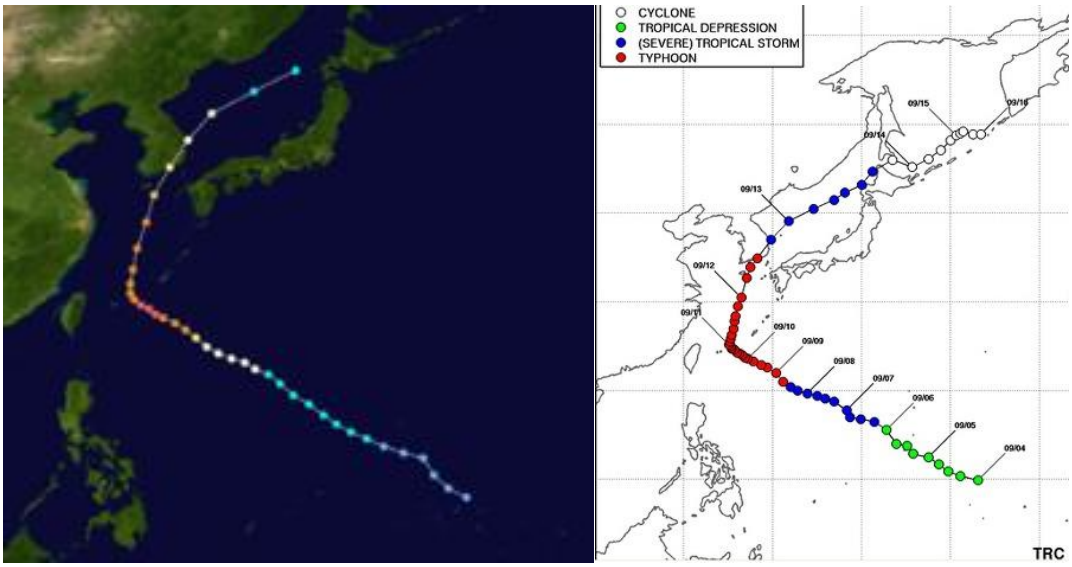

Figure 1.2 Track of Typhoon Maemi in 2003, National Typhoon Center [38]

1.2.2 Tidal gauge stations in South Korea

To reduce the damage caused by typhoons, accurate prediction of storm surge height has become critical. Accurate predictions can provide important information when building infrastructure on the waterfront to prevent damage from storm surges. In order to reduce damage to coastal infrastructure, this study seeks to predict when specific storm surge heights will recur. Using sea level heights recorded at tidal gauge observatory stations on the Korean Peninsula, the study focuses on Typhoon Maemi, which caused more property damage than any other typhoons in South Korea's history (Table 1.1). South Korea currently operates 17 tidal gauge stations, observing and recording sea water levels for more than 25 years. The tidal gauge stations are located on the South, West, and East Coasts as shown in Figure 1.3. Seventeen tidal gauge stations have observation records over 25 years while 13 of them have been observing for over 30 years. The tidal gauge station in Busan has recorded for 54 years.

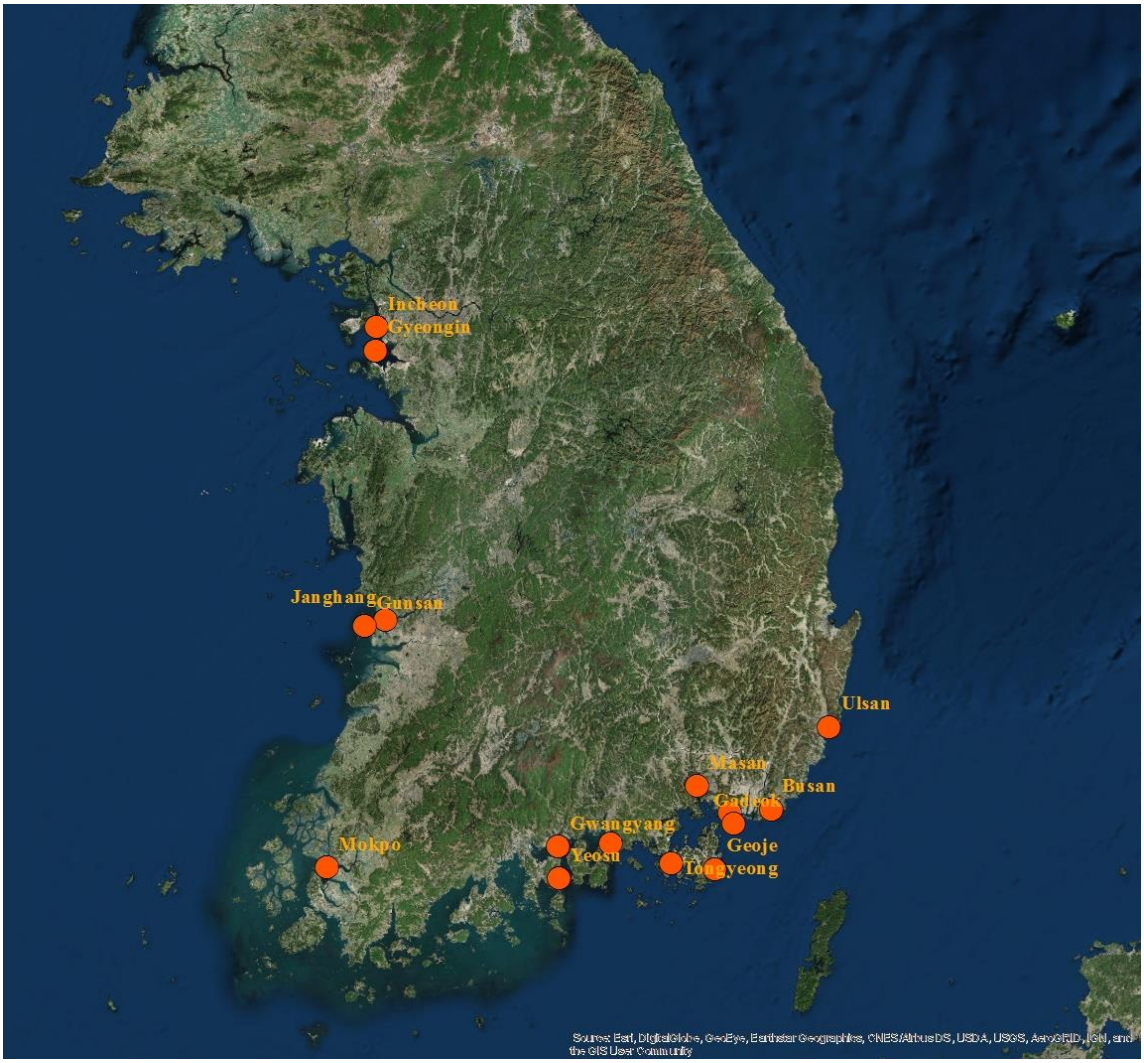

Figure 1.3 Location of tidal gauge stations in South Korea

This study focuses on 16 tidal gauge stations located on the South and West coasts. The reason for excluding the tidal gauge stations on the East Coast is that the majority of the typhoons do not pass from the East Coast. The tidal data for this study has been referenced from the Korea Hydrographic and Oceanographic Agency (KHOA) [28] which allowed the present study to use hourly tide data for research purposes.

1.2.3 Datum level in tidal gauge station of Busan

Each tidal gauge station has their own tidal data, according to their tidal patterns and location. Korea Hydrographic and Oceanographic Agency (KHOA) computes various elevation datums, such as the mean sea level (MSL), mean low water (MLW), mean high water (MHW), and approximately higher high water (AHHW), also known as mean higher high water (MHHW) for tidal gauge stations in the Korean Peninsula. Those are referenced to the local standard datum, which can be used to measure water height. Such datum level is a very important reference for planning, design, and construction of harbor and offshore facilities.

According to KHOA, the mean sea level datum level in Busan's tidal gauge station is 64.9 cm below the mean sea level. The below Figure 1.4 illustrates the datum elevation at the tidal gauge station in Busan, South Korea:

- Mean Sea Level (MSL): 64.9 cm
- Mean Low Water (MLW): 86 cm
- Mean High Water (MHW): 123.8 cm
- Approximately Highest High Water (AHHW) or
- Mean Higher High Water (MHHW): 129.8 cm

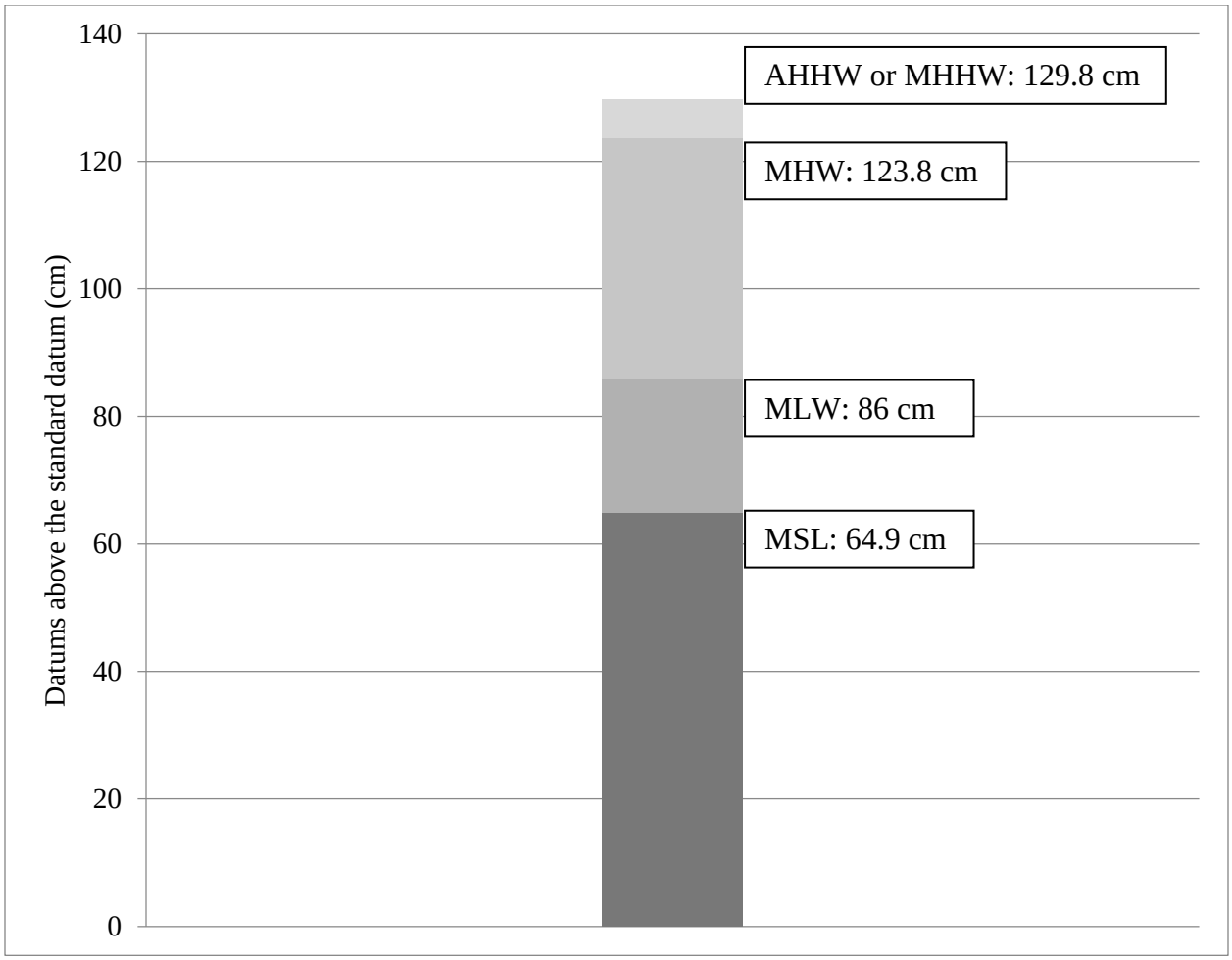

Figure 1.4 Datum elevations at the tidal gauge station in Busan, South Korea

1.2.4 Three highest total water levels at tidal gauge stations in South Korea

The West coast operates tidal gauge stations in Incheon, Gyeongin, Changwon, Gunsan, and Mokpo. The location of each station is shown in the above Figure 1.3. Each station has different observation lengths. After collecting hourly sea levels at each station within the observation period, this study identified the top three sea level heights measured at each station. It is obvious that the heights of the highest measured sea levels are related to the typhoons that occurred on the Korean Peninsula. Table 1.4 shows observation lengths for the tidal gauge stations located on the West coast and the top three sea water levels recorded of each station, respectively.

		Years	Top three high peaks (cm)	Date of the high peak
West coast	Incheon	18	987	7/24/2013, 10:00
			981	9/8/2002, 06:00
			980	10/27/2003, 18:00
	Gyeonin	2	993	9/30/2015, 19:00
			987	9/29/2015, 18:00
			986	10/29/2015, 18:00
	Janghang	14	798	9/30/2015, 17:00
			796	10/11/2014, 17:00
			794	9/29/2015, 16:00
	Gunsan	37	805	8/19/1997, 04:00
			799	8/21/1997, 05:00
			797	8/31/2000, 05:00
	Mokpo	61	544	7/4/2004, 04:00
			544	7/6/2004, 05:00
			538	11/16/2012, 16:00

Table 1.4 Top three high peak water levels recorded at tidal gauge stations in the West coast

The approach used to analyze data from tidal gauge stations on the West coast was applied in the same way to analyze data obtained from tidal gauge stations located on the South coast. Tables 1.5 and 1.6 below show observation lengths and the top three highest sea levels at each station on the South coast.

		Years	Top three high peaks (cm)	Date of the high peak
South coast	Busan	54	211	9/12/2003, 21:00
			190	9/12/2003, 20:00
			188	9/12/2003, 12:00
	New Busan	5	221	9/18/2012, 10:00
			219	9/17/2012, 09:00
			219	8/11/2014, 21:00
			219	8/11/2014, 22:00
			215	9/17/2012, 10:00
			215	9/18/2012, 22:00
	Gadeok	40	252	9/17/2012, 10:00
			246	9/17/2012, 09:00
			246	7/16/2987, 00:00
			241	7/16/1987, 01:00
	Masan	37	265	9/17/2012, 10:00
			264	9/17/2012, 11:00
			244	8/29/2004, 21:00
			244	8/21/2005, 23:00
	Ulsan	55	133	8/19/2004, 08:00
			120	9/12/2003, 21:00
			129	9/17/2012, 20:00

Table 1.5 Top three high peak water levels recorded at tidal gauge stations in the South coast

		Years	Top three high peaks (cm)	Date of the high peak
South coast	Tongyeong	41	426	9/12/2003, 21:00
			357	9/12/2012, 10:00
			356	9/12/2003, 20:00
	Samcheonpo	2	352	8/30/2015, 22:00
			350	10/28/2015, 09:00
			350	11/27/2015, 10:00
			350	4/8/2016, 22:00
			347	8/29/2015, 21:00
	Geoje	11	270	9/17/2012, 09:00
			259	9/17/2012, 10:00
			255	1/4/2006, 09:00
	Gwangyang	6	479	9/17/2012, 10:00
			443	9/17/2012, 11:00
			441	8/1/2014, 22:00
	Yeosu	52	440	8/18/1966, 23:00
			430	9/14/1966, 21:00
			129	8/17/1966, 22:00

Table 1.6 Top three high peak water levels recorded at tidal gauge stations in the South coast (Cont'd)

1.2.5 Tidal gauge station at the City of Busan in South Korea

Among the tidal gauge stations, there is a station that has observation records covering the past 54 years. The station is located on the South coast in Busan, the second largest city in South Korea. The position of the station is shown in the above Figure 1.3. Busan is located in the southeast of the Korean Peninsula. Because of its geographical location, international trade has boomed; as a consequence Busan has the largest port in South Korea. Busan also has the longest and widest river in South Korea passing through. It is called Nakdong River. Due to these geographical characteristics, Busan has been very vulnerable to natural disasters such as typhoons, and the importance of accurately predicting the statistics of future storms has become more and more recognized among governmental agencies and other stakeholders.

The top three sea level heights at the tidal gauge station in the Busan area are shown in the Table 1.7 below.

	Years	Top three high peak (cm)	Date of the high peak
Busan	54	211	9/12/2003, 21:00
		190	9/12/2003, 20:00
		188	9/12/2003, 12:00

Table 1.7 Top three high peak water levels recorded at the tidal gauge station in Busan, South Korea

Considering the top three highest sea level results above, this study analyzed sea level history from 1956 to 2016 in Busan. Figures 1.5 and 1.6 show the annual mean and fluctuations around the mean during this observation period at Busan's tidal gauge station.

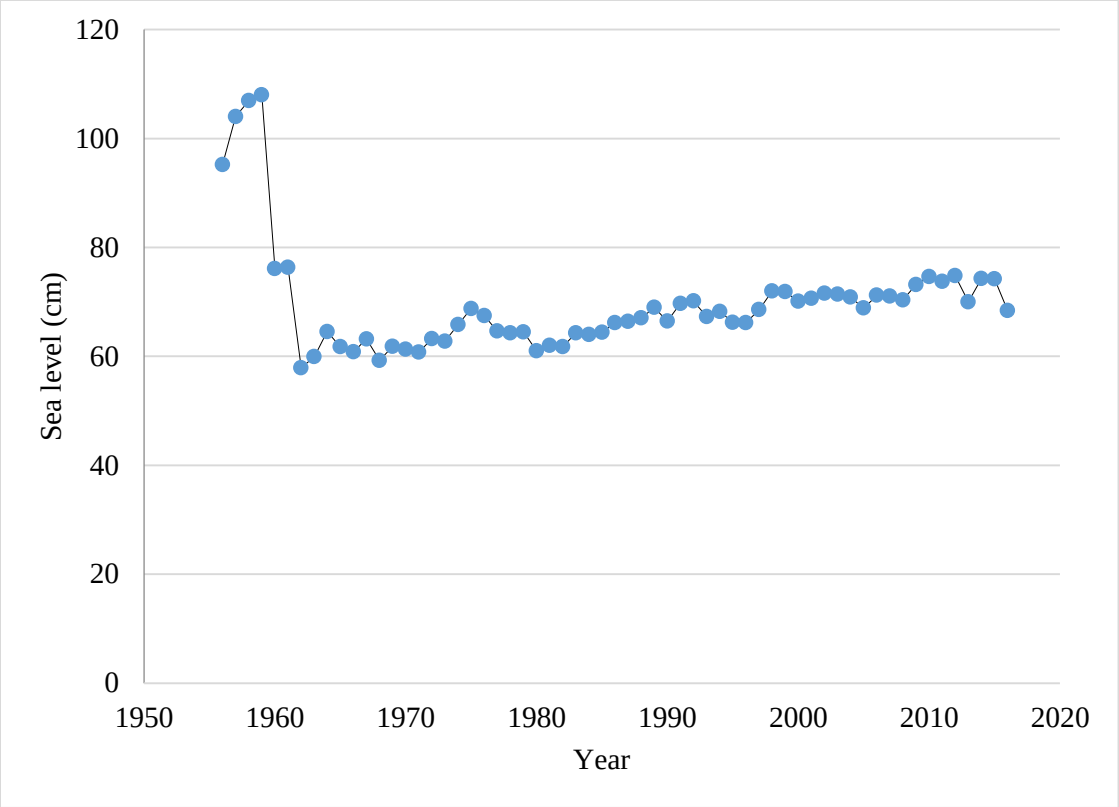

Figure 1.5 Time history of sea level in Busan, South Korea (1956-2016)

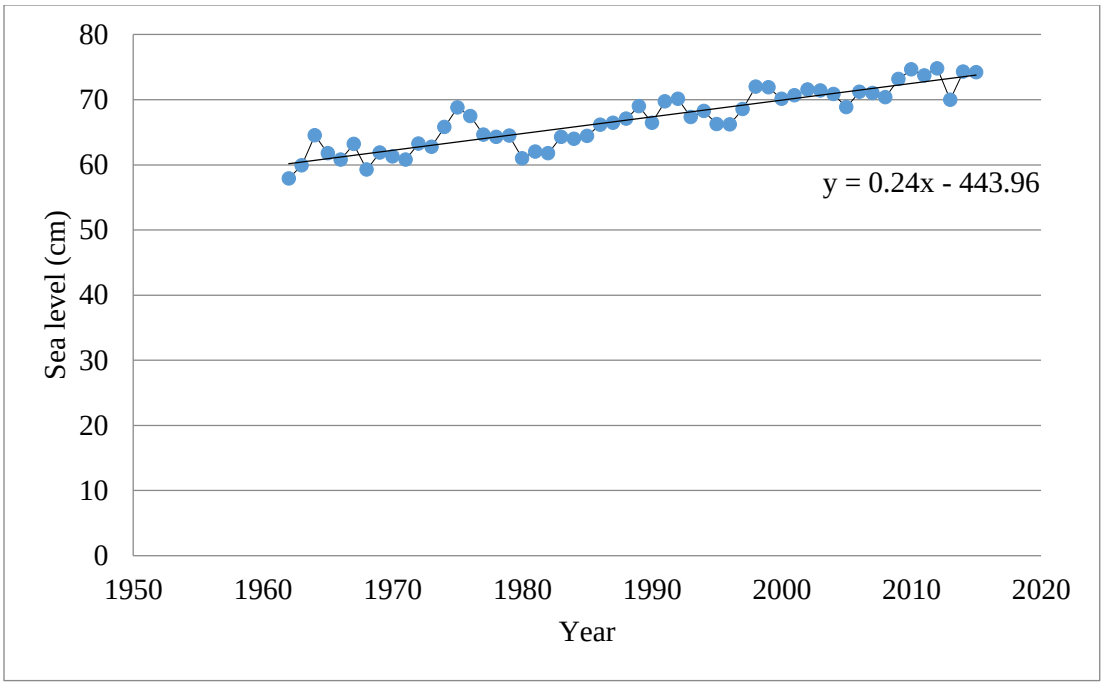

Figure 1.6 Time history of sea level in Busan, South Korea (1962-2015)

When looking at sea level history in Figure 1.5, the data trend between 1956 and 1961 is clearly anomalous. The data recorded during this period appear to have quality control issues, so they have been excluded from this study, and only data from 1962 to 2015 have been used, as shown in Figure 1.6.

KHOA provides hourly observed water height at the tidal gauge station. From this hourly data, each annual mean is calculated. Plotting mean water level for each year confirms water level variation. As seen in Figure 1.6, even if there are different fluctuations between years in the plot, the sea level continues to increase. Supposing mean sea level variation as a function of time, a linear regression is performed. The resulting coefficient of slope indicates the rate of increase (Yoon and Kim, 2012) [56]. Figure 1.6 shows:

$$y = 0.24x - 443.96$$

From the above equation, this study finds that the increasing rate of sea level at the Busan's tidal gauge station is 2.4 mm per year. The difference of mean sea level between 1962 and 2015 is 16.31 cm. Yoon and Kim, 2012 [56] also conducted similar research to find the approximate sea level rise in the Korean Peninsula, but applied a different time period (1960-2010). Nonetheless, their study calculated similar results (2.9 mm/year).

As previously mentioned, the reason why Busan has been chosen is that Busan's tidal gauge station has one of the longest observation histories, regularly updated to the present. Also, due to its topography and concentration of the industrial facilities, the city needs more accurate predictions of natural hazards.

1.2.6 Relationship between sea level and typhoons

This study investigates the relationship between typhoons and sea level. When a storm occurs, the height of the surge, which is the difference between the observed sea level height and the predicted sea level height, tends to increase. These significant surges can cause major natural disasters such as floods. In this study, before calculating the height of the surge, the date and time when the top three highest sea level heights were observed and the date and time when the typhoon occurred are compared.

Tables 1.8 and 1.9 show the date and time when the top three sea level heights for each tidal gauge station occurred, respectively. Also, the corresponding sea levels at these tidal gauge stations associated with typhoon occurrence are presented in Tables 1.8 and 1.9.

Location	Years	Peak	Date	Typhoon
Busan	54	211	9/12/2003, 21:00	Maemi
		190	9/12/2003, 20:00	Maemi
		188	9/12/2003, 12:00	Maemi
New Busan	5	221	9/18/2012, 10:00	Sanba
		219	9/17/2012, 09:00	Sanba
		215	9/18/2012, 22:00	Sanba
Gadeok	40	252	9/17/2012, 10:00	Sanba
		246	9/17/2012, 09:00	Sanba
		246	7/16/1987, 00:00	Thelma
Masan	37	265	9/17/2012, 10:00	Sanba
		264	9/17/2012, 11:00	·
		244	8/29/2004, 21:00	·
Ulsan	55	133	8/19/2004, 08:00	Megi
		120	9/12/2003, 21:00	Maemi
		129	9/17/2012, 20:00	Sanba

Table 1.8 Relation between sea level and typhoons

Location	Years	Peak	Date	Typhoon
Tongyeong	41	426	9/12/2003, 21:00	Maemi
		357	9/12/2012, 10:00	Sanba
		356	9/12/2003, 20:00	Maemi
Samcheonpo	2	352	8/30/2015, 22:00	·
		350	10/28/2015, 09:00	·
		350	11/27/2015, 10:00	·
Geoje	11	270	9/17/2012, 09:00	Sanba
		259	9/17/2012, 10:00	Sanba
		255	1/4/2006, 09:00	·
Gwangyang	6	479	9/17/2012, 10:00	Sanba
		443	9/17/2012, 11:00	Sanba
		441	8/1/2014, 22:00	·
Yeosu	52	440	8/18/1966, 23:00	·
		430	9/14/1966, 21:00	·
		129	8/17/1966, 22:00	·

Table 1.9 Relation between sea level and typhoons (Cont'd)

As can been seen in Tables 1.8 and 1.9 above, the three high peak sea levels at each tidal gauge station are related to the occurrences of typhoons. The date and time of the highest three sea level heights observed during Busan's entire observation period (1963-2015) coincide with the date and time when Typhoon Maemi passed from the Busan area in September 2003. Tables 1.8 and 1.9 also show that the highest three sea level heights at tidal gauge stations located in other cities relate strongly with observed typhoons.

Typhoon Maemi caused damage of $35 billion and many casualties (135 people) in Busan and nearby cities (National Typhoon Center, 2011) [38]. Other typhoons such as Thelma, Samba, and Megi also caused very significant damage. The amounts of damage are showed in Table 1.1.

1.3 Extreme value statistics

1.3.1 Generalized extreme value (GEV) distribution

Extreme events are difficult to predict because data points are very few. Predicting the probability of extreme events is particularly difficult due to their asymptotic nature. Extreme value probability theory deals with to find outlier information such as the maximum or minimum value of extreme situations. Examining the tail events in a probability distribution is very challenging. However, it is considered to be very important in the civil engineering and insurance industries since those industries need to consider low probability high consequence events to cope with various extreme situations. For example, bridges, breakwaters, dams, and industrial plants which are located near shore or flood areas should account for the low probability of extreme events to reflect the consequences of major natural disasters. In this research, extreme total high water level caused by typhoons is the main theme. There are various probability models to study extreme events. Extreme value theories can be divided into two

groups depending on how they are defined. First, the entire interval of interest is divided into a number of subintervals. When the maximum value from each subinterval is identified as the extreme value for this subinterval, the entirety of these extreme values converges into a generalized extreme value distribution. Second, when values that exceed a certain threshold are identified as extreme values, these extreme values converge to a generalized Pareto distribution (GPD). In the following chapters, the peaks-over-threshold and block maxima methods will be used to better understand these two groups of distributions (Coles, 2001) [7].

1.3.2 Peaks-over-threshold (POT) and block maxima (BM)

1.3.2.1 Block maxima (BM) method

One of the available methods for studying extreme values is block maxima. It uses the distribution of the maximum extreme values in the following equation, with X_n in order of maximum extreme values:

$$M_n = \max\{X_1 , \dots , X_n\}$$

n: number of observation in year

X_n: independent and identically random variables

M_n: annual maximum

Data is divided into blocks of specific time periods with the highest values within each block collectively serving as a sample for extreme values. One limitation of the block maxima method is the possibility of losing valuable extreme value data because the approach only takes the single largest data value in each block. Therefore, there is a possibility that the second largest or highest datum in one block could be larger or second higher than the highest or largest datum in another.

The other approach is the peak-over-threshold (POT) method. It can address the limitations of the block maxima method since the POT method can gather all the data points which exceed a certain prescribed threshold. That means that the POT method is not missing valuable extreme values compared to the block maxima method even when they occur close to each other. But deciding which threshold to select in order to get the best description of the extreme data can be challenging (Bommier, 2014) [3]. For example, if the threshold is too high, several extreme values might be lost. If it is set too low, values that are not really extremes can be considered.

1.3.2.2 Peaks-over-threshold (POT) method

The Peaks-over-threshold (POT) method has been used in various fields which need to study extreme events (Bommier, 2014) [3]. It uses limited data more efficiently by using relatively larger or higher values rather than the largest or highest ones. A threshold value is used to assort the larger or higher values from all data and the values above the threshold are called exceedances.

All data exceeding threshold value constitute the sample of extreme values. Determining the appropriate threshold value, however, requires significant trial and error. If the threshold is too high, valuable data may be lost. On the other hand, if the threshold is too low, both significant and insignificant data is collected and categorized as extreme values. Moreover, a low threshold could bring more bias due to decreased variance. Much research has been conducted on optimizing threshold values (Scarrott and Macdonald [51], 2012; Lopeman, 2015 [37]; Pickands, 1975 [54]).

Pickands, 1975 [54] suggested that the independent time series which exceed enough high thresholds would follow a generalized Pareto distribution (GPD) asymptotically. Pickands's POT avoids the inherent drawbacks of BM [54]. Whereas BM identifies only one extreme data point within specified time windows, POT isolates all data points that exceed a designated threshold. All exceedances above the designated threshold can be explained by the differentiated tail data distribution.

The below equation explains the distribution function F of exceedance above the threshold.

$$F_\theta(x) = P\{X - \theta \leq x | X > \theta\}, \quad x \geq 0$$

θ: threshold

X: random variable

Also, as shown in the below equation, F_u can be defined by conditional probabilities.

$$F_\theta(x) = \begin{cases} \dfrac{F(\theta + x) - F(\theta)}{1 - F(\theta)} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

θ: threshold

X: random variable

According to Bommier, 2014 [3], the distribution of exceedances $(Y_1, \dots, Y_{n_\theta})$ can be generalized by GPD with following assumption: When, $Y = X - \theta$ for $X > \theta$, and $X_1, \dots, X_n$, $Y_j = X_i - \theta$ can be described with i is jth exceedance, $i = 1, \dots, n_\theta$.

The equation below expresses the GPD:

$$G_x(x;\xi,\sigma,\theta) = \begin{cases} 1-\left(1+\xi\dfrac{(x-\theta)}{\sigma}\right)^{1/\xi} & \xi \neq 0 \\ 1-\exp\left(-\dfrac{(x-\theta)}{\sigma}\right) & \xi = 0 \end{cases}$$

In the above equation, x equals independent and identically random variables, σ is the scale, ξ is the shape, and θ is the threshold. All values above θ are considered tail data (extreme values).

When calculating return level, which is exceeded once every N years (N-year return periods x_N), the below equation describes the probability of exceedance over a threshold.

$$P\{X > x | X > \theta\} = \left[1+\xi\left(\frac{x-\theta}{\sigma}\right)\right]^{-1/\xi}$$

X: random variable

σ: scale

ξ: shape

θ: threshold

If the exceedances above threshold are rare events (number of observations per year), we can expect $P(X > \theta)$, which is the probability of an exceedance above a threshold, would follow Poisson distribution. The mean of exceedance per unit time ($\hat{\gamma}$) describes the Poisson distribution.

$$P(X > \theta) = \frac{\gamma}{\lambda}$$

λ: rare events (number of observations per year)

Here, γ can be estimated by dividing number of exceedances above the threshold by the number of years in the observation period.

Combining the POT and Poisson processes with GPD allows us to describe the conditional probability of the extreme values that exceed the designated threshold. The equation below defines the conditional probability (Lopeman and Deodatis, 2015) [33], [34], [35], [36]:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

When Bayes' theorem is applied to the role of GPD in conditional probability, we can rewrite the equation above as follows:

$$G_X(x) = \frac{P(\theta < X < x)}{P(X > \theta)}$$

x: independent and identically random variables (exceedance values)

1.4 Outline and objective

The purpose of this dissertation is two-fold: the first objective is to the estimate the return period in years of typhoon-induced high water levels in South Korea. The second objective is to calculate exceedance probabilities of damage by developing empirical fragility curves with respect to maximum wind speed induced by typhoons in the Korean Peninsula. This research adopts and further develops Lopeman and Deodatis's (2015) [33], [34], [35], [36] clustered separated peaks-over-threshold method (CSPS). CSPS provides statistical analysis of extreme values in long time series of natural phenomenon that can assume extreme values. Such an analysis can provide guidelines for coping with natural disasters in the Korean Peninsula, especially on the south coast of South Korea before they occur. Second, this dissertation continues by constructing fragility curves for buildings subjected to typhoon-induced maximum wind speeds. It uses for the first time data provided by an insurance company in South Korea of actual damages caused by Typhoon Maemi. The established fragility curves delineate

exceedance probabilities for various damage stages as a function of maximum wind speed. The findings from this research provide a realistic method to predict economic losses associated with typhoons and corresponding models to manage emergency situations arising from natural disasters for governmental agencies, insurance companies, and the construction industry in South Korea. Although this dissertation focuses on a specific region and city in the Korean Peninsula, the introduced probabilistic methodologies can be applied to other coastal regions in South Korea and around the world.

Chapter 2

2 Data processing and statistical model for determining return periods of typhoons in the Busan area

2.1 Literature review

2.1.1 Studies regarding Typhoon Maemi

Previous research has attempted to find approaches to reduce damages from natural disasters in South Korea. Most of these studies had focused on storm characteristics, such as storm track, rainfall, radius, wind field data, etc., and damages caused. Typically, they create synthetic storms predicting storm paths and estimate the extent of the associated damages.

Kang, 2005 [20] investigated the inundation and overflow caused by Maemi at one location near the coast. The study used a site survey and interviews from residents in the damaged area and found that inundation happened once the storm surge increased the water level by 80%. Using a numerical model, Hur et al., 2006 [14], [15] estimated storm surges in the Busan area caused by the most serious typhoons, such as Sarah, Thelma, and Maemi. The study compared the storm surge heights of the most serious typhoons at several points near Busan and found that Typhoon Maemi brought the highest storm surge compared to that of other typhoons. They then simulated storm surges to investigate tidal characteristics at Busan's coast. They created virtual typhoons to compare with the corresponding tracks of Sarah, Thelma, and Maemi. When the virtual typhoons followed the track of Typhoon Maemi, the simulated storm surge

height was higher than the surge heights produced when virtual typhoons followed the tracks of Sarah and Thelma.

Lee et al., 2008 [30] used atmospheric pressure and wind profiles of Typhoon Maemi, and introduced a multi-nesting grid storm surge model to simulate storm surges. To check their model's performance, they used numerical methods to test tidal calibration, the influence of open boundary conditions, and typhoon path. The study resulted in two findings: First, the location of the typhoon center was the most critical factor in calculating storm surges. Second, the track of the typhoon was a secondary, but still important, factor in storm surge prediction. The limitation of this research was that only recorded storm tracks were used, meaning that the simulation could not calculate storm surges from any other possible track. Like Lee et al., 2008 [30], Chun et al., 2008 [4] used a numerical model to simulate the storm surge of Typhoon Maemi, but used data from the coastal area of Masan, a city near Busan which was also damaged by the storm. The numerical model was combined with moving boundary conditions to explain the wave run-up. The study compared the predicted inundation area and depth with an actual site survey and found that the two were reasonably well correlated.

Kim and Suh, 2018 [22] created 25,000 random storms by modifying an automatic generation tool, Tropical Cyclone Risk Model (TCRM). They simulated surge elevations for each randomly generated storm. The tracks of these random, simulated storms had similar patterns to those of actual typhoons in South Korea. While past research on Typhoon Maemi has used such input data as atmospheric pressure, wind field, radius of typhoon, storm speed, latitude, and longitude, tidal gauge data, etc., no research has estimated return period using tidal gauge data. For instance, Kim and Suh's [22] study does not perform surge modeling and frequency analysis in the time domain. Although Chun et al., 2008 [4] provided valuable information

regarding the predicted inundation area and depth by numerical models, they did not consider tidal fluctuation which, if combined with the increased water level, would have given them different predicted inundation areas and depths.

2.1.2 Return period estimates of Hurricane Sandy, 2012, United States

While no research exists currently on the estimation of period return for typhoons, research does exist for the return period of hurricanes. Lopeman and Deodatis, 2015 [33], [34], [35], [36] used tidal gauge data to estimate the return period of Hurricane Sandy 2012. Other research on Sandy had also used tidal gauge data. For example, Talke et al., 2014 [52] studied the storm surge hazard in the New York harbor using tidal gauge data over a 37-year period. This research showed that the pattern of the storm surge hazard has been changing due to sea level rise caused by climate change and other possible meteorological factors. However, Talke et al. [52] did not estimate the specific return period of Hurricane Sandy.

Lin et al., 2010 [32], on the other hand, did estimate the return period of storm surges related to tropical cyclones in the NYC area. The estimated water level return period of Sandy in Lower Manhattan was found to be 500 years within a 95% confidence interval (approximately 400-700 years). Lin et al., 2012 [31] conducted a similar analysis using computational fluid dynamics Monte Carlo Simulations combining the randomness of the tidal phase angle. The result for the return period was 1,000 years with a 90% confidence interval of 750-1,050 years. Lin et al.'s study from 2010 [32] is considered less accurate than the 2012 [31] study because it did not consider different possibilities of surge height at different time windows within the tidal cycle.

Hall and Sobel, 2013 [13] developed an alternative method to estimate Sandy return periods. The study proposed that the track of Sandy was the primary reason for the damage in Lower Manhattan and other parts of NYC. Specifically, they argued that damage occurred due to Sandy's perpendicular impact angle with respect of the shore as it passed to the south of the Manhattan port. The study analyzed other hurricanes' tracks and concluded that the impact angle was the most critical factor in accounting for the damage caused by Sandy and other hurricanes of similar intensity. The return period of Hurricane Sandy's water level was estimated to be 714 years within a 95% confidence interval (435-1,429 years).

Zervas, 2013 [57] estimated the return period for extreme events using data from National Oceanographic and Atmospheric Administration (NOAA). NOAA has provided water level data to estimate return periods, recording the monthly mean water level at the tidal gauge station at Battery Park, New York. Zervas [57] used the GEV distribution and the Maximum Likelihood (MLE) method to estimate the return period for the highest observed water level. The GEV distribution showed that the return period for Sandy's peak water level was 3,500 years. However, sensitivity analysis suggested that the estimated results were too high given the GEV fit's sensitivity to the range of years used in the analysis. Once Sandy was excluded, the return period was 60,000 years. The difference in results suggests that the GEV distribution of the yearly maximum water level is not a realistic method for estimating extreme events for the New York Harbor area.

Building on past research, Lopeman and Deodatis 2015 [33], [34], [35], [36] estimated the return period for Sandy using tidal gauge data, the first researchers to do so, in order to estimate the total water level return period in New York Harbor. Lopeman and Deodatis [33], [34], [35], [36] proposed the clustered separated peaks-over-threshold simulation (CSPS) method

to be used and to separate tide fluctuation, surge, and sea level rise (only the surge being truly random). Lopeman and Deodatis calculated the return period of Sandy to be 103 years with a 95% confidence interval (38-452 years) [33], [34], [35], [36].

The differences in the results of past studies are due to the fact that different data and different assumptions were used in each one. The present study applies methods used in previous studies on Hurricane Sandy to estimate the return period of Typhoon Maemi and in the process establishes a new model.

2.2 Surge data

2.2.1 Storm surge data collection method

The Korea Hydrographic and Oceanographic Agency (KHOA) makes observed data regarding water height at designated observation stations available to the public. To determine the height of the surge (i.e., the difference between the observed and the predicted sea level height), the predicted sea level height is necessary.

The followed equation explain the relationship among observed water level, predicted water level, tidal fluctuation height, and residual (surge) at time t_i.

$$Y_i = X_i + S_i$$

$i = 1, 2, \dots, n$ (n is the time series of the input dataset).

- X_i is the predicted water height at time t_i
- Y_i is the observed water height t_i
- S_i is the residual (surge) water height

In this study, a standard harmonic analysis is performed to calculate the predicted sea level height based on hourly data.

2.2.2 Separation of tidal gauge data (Harmonic analysis)

Harmonic analysis is used to separate the observed data:

- First, harmonic analysis is used to estimate the tidal components from the total sea water level data, allowing residuals to be isolated so that surge data can be calculated once sea level rise is estimated
- Second, the estimated constituents are used to predict tidal fluctuations in the years simulated via Monte Carlo.

For the estimation of tidal components, the R package *TideHarmonics* by Stephenson [50] was used. The details of the estimation procedure are presented below:

Consider the time series $Y(t)$ of total water levels with t denoting time in hours. The representation of the tidal component with M harmonic constituents is given by:

$$\hat{Y}(t) = Z + \sum_{m=1}^{M} A_m cos\left(\frac{\pi}{180}(\omega_m t - \psi_m)\right),$$

where ω_m is the angular frequency of the m-th component in degrees per hour. The $2M + 1$ parameters to be estimated are: the amplitudes A_m, the phase lags ψ_m in degrees, and the mean sea level Z.

To account for long astronomical cycles (LAC), nodal correction functions for both the amplitude and phase are used. With these corrections, the tidal component takes the following form:

$$\hat{Y}_{LAC}(t) = Z + \sum_{m=1}^{M} H_m f_m(t) cos\left(\frac{\pi}{180}(\omega_m t - g_m + u_m(t) + V_m)\right),$$

where $f_m(t)$ and $u_m(t)$ represent the nodal corrections for the amplitude and phase, respectively. In this new formulation, the amplitude and phase parameters to be estimated are denoted by H_m and g_m (in degrees). Finally, V_m is the reference signal with respect to which the phase lag g_m is calculated and is set to refer to the origin $t = 0$.

The summation term in $\hat{Y}_{LAC}(t)$ can be alternatively written as:

$$\sum_{m=1}^{M} \beta_{m,1} f_m(t) \cdot cos\left(\frac{\pi}{180}(\omega_m t + u_m(t) + V_m)\right) +$$

$$\sum_{m=1}^{M} \beta_{m,2} f_m(t) \cdot sin\left(\frac{\pi}{180}(\omega_m t + u_m(t) + V_m)\right),$$

where $\beta_{m,1} = H_m cos(g_m)$ and $\beta_{m,2} = H_m sin(g_m)$.

What is gained with this new representation is a linear function with respect to the parameters $\beta_{m,1}$ and $\beta_{m,2}$ that need to be estimated; hence, a linear regression can be used. The full algorithm is given in the following Table 2.1:

Algorithm: *Harmonic analysis of sea-level height.*

Input: Sea level time series $Y(t)$, indexed by time in hours.

Number of tidal harmonic components M.

Angular frequency ω_m of each harmonic constituent m.

Nodal corrections $\left(f_m(t), u_m(t)\right)$ for amplitude and phase for each harmonic component m. (These are provided by in the R package TideHarmonics.)

Output: Estimated tidal component $\hat{Y}_{LAC}(t)$.

1. Set $\hat{Z} = mean(Y)$, where $\{Y(t)\}_t$ is the entire sea level time series.
2. Run zero-intercept linear regression on $Y(t) - \hat{Z}$ against the following $2M + 1$ variables:

$$X_{m,1} = f_m(t) \cdot cos\left(\frac{\pi}{180}(\omega_m t + u_m(t) + V_m)\right),$$

$$X_{m,2} = f_m(t) \cdot sin\left(\frac{\pi}{180}(\omega_m t + u_m(t) + V_m)\right),$$

for $m = 1, \dots, M$. Denote the respective coefficient estimates as $\hat{\beta}_{m,1}$ and $\hat{\beta}_{m,2}$, for $m = 1, \dots, M$.

3. For each $m = 1, \dots, M$, set:
   a. $\hat{H}_m = \sqrt{\beta_{m,1} + \beta_{m,2}}$;
   b. $\hat{g}_m = arctan\left(\frac{\beta_{m,2}}{\beta_{m,1}}\right)$.
4. The estimated tidal component is given by:

$$\hat{Y}_{LAC}(t) = \hat{Z} + \sum_{m=1}^{M} \hat{H}_m f_m(t) cos\left(\frac{\pi}{180}(\omega_m t - \hat{g}_m + u_m(t) + V_m)\right).$$

Table 2.1 Harmonic analysis of tidal and sea-level data

Given the large timespan covered by the data, $M = 60$ harmonic tidal constituents are estimated, and a constant mean sea level Z is assumed across the 54 years of data available.

2.2.3 Observed, predicted and residual water level

A harmonic analysis is conducted using the observed sea level data at Busan's tidal gauge station, which has the longest observation length among the tidal gauge stations operating in South Korea. Because observed sea level height is usually different from predicted sea level height, Figure 2.1 displays the height of observed sea level in blue, as calculated through harmonic analysis, predicted sea level height is in green, and surge height is in red.

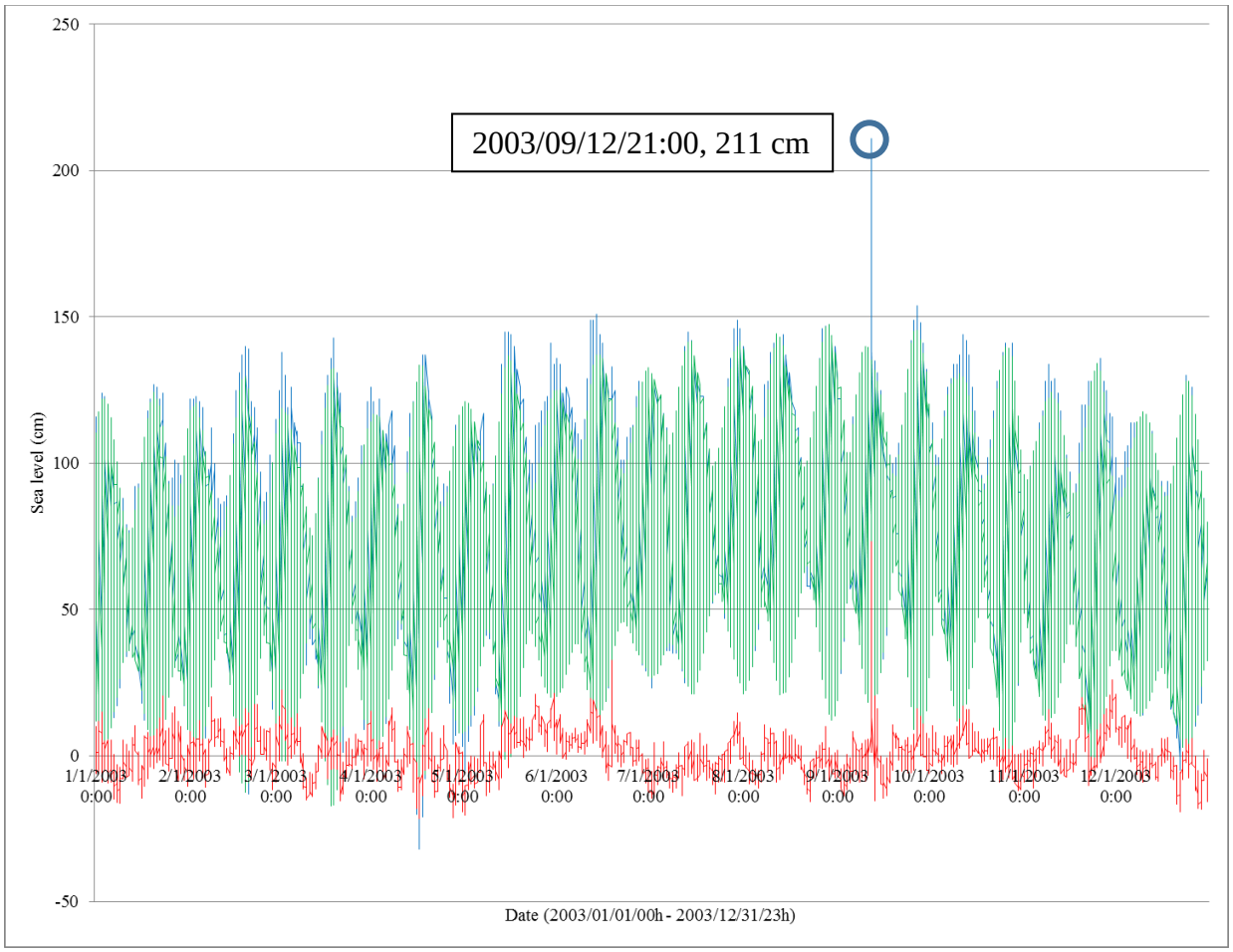

Figure 2.1 Observed, predicted, and residual water level in Busan, South Korea, 2003

As seen in Figure 2.1, the highest total water level coincided with the highest surge level during Typhoon Maemi. Figure 2.2 shows that the observed (in blue), predicted (in green), and surge height (in red) on the specified date (9/12/2003, 21:00) when the highest sea level occurred at the Busan tidal gauge station. Given a total water height of 211 cm, the surge height is calculated as 73.35 cm.

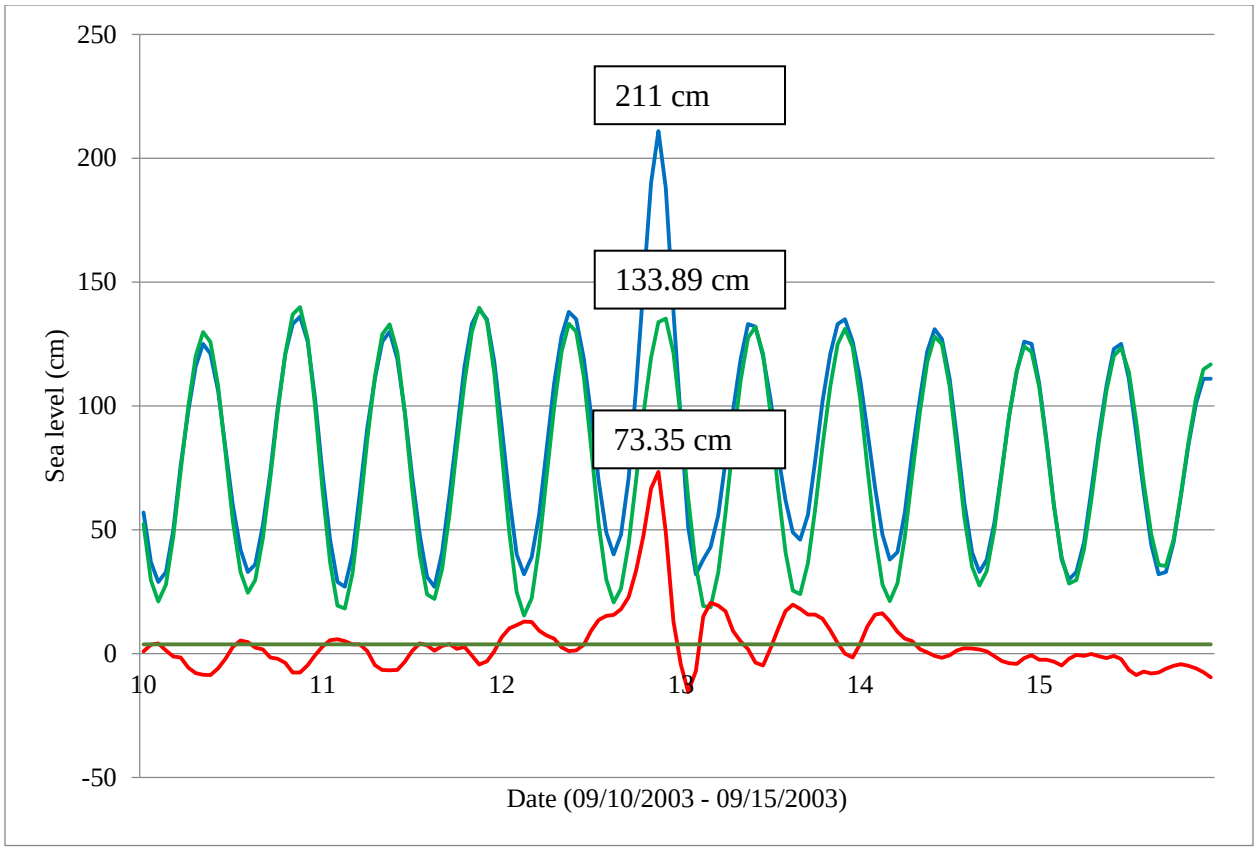

Figure 2.2 Observed (green), predicted (blue), and residual (red) water level, Typhoon Maemi, 2003

2.3 Threshold and target rate selection

For a given target annual rate, which is the number of the events (storms) per year, the algorithm proposed by Lopeman, 2015 [37] (shown in Figure 2.3) computes the threshold such that the given rate (within some tolerance) approximates the resulting yearly number of exceedance clusters. As mentioned in chapter 1.3, having the proper threshold is very important since too high a threshold can lose valuable tail data, and too low a threshold can include data that is likely to happen and therefore not does not qualify as extreme. An exceedance cluster is given by a set of consecutive surge observations that lie above the threshold. Hence, rather than choosing an "ideal" threshold according to some criteria, the algorithm finds the threshold that forces a previously chosen target rate. We can set such a target rate to be equal to the average rate observed over a given period or, more generally, to a value that we find reasonable given past data.

The algorithm uses iterative updates of the threshold to allow a computationally intensive, but not exhaustive, exploration of all the possible threshold values from its minimum (i.e., minimum observed surge height) to its maximum value (i.e., maximum observed surge height). In particular, the algorithm first sets the threshold to zero centimeters, and then iteratively overwrites it according to the following steps:

1. At a given iteration with a certain threshold, the resulting exceedance clusters are identified and the resulting annual storm rate computed.
2. If the resulting annual storm rate is close enough to the chosen target, then the threshold from the previous iteration is the result and the algorithm is stopped.
3. If the resulting annual storm is not close enough to the chosen target

- … but is smaller than the target rate, then the threshold from the previous iteration is the result and the algorithm is stopped.
- … but is greater than the target rate, then a vector collecting the maximum peak height of the clusters is built and sorted in descending order. The threshold updated by setting it equal to the C-th element of this vector, where C is the closest integer to the product of 54 (the number of years covered by the dataset) with the target rate. This updated threshold is used in the next iteration of algorithm and steps 1 through 3 are repeated.

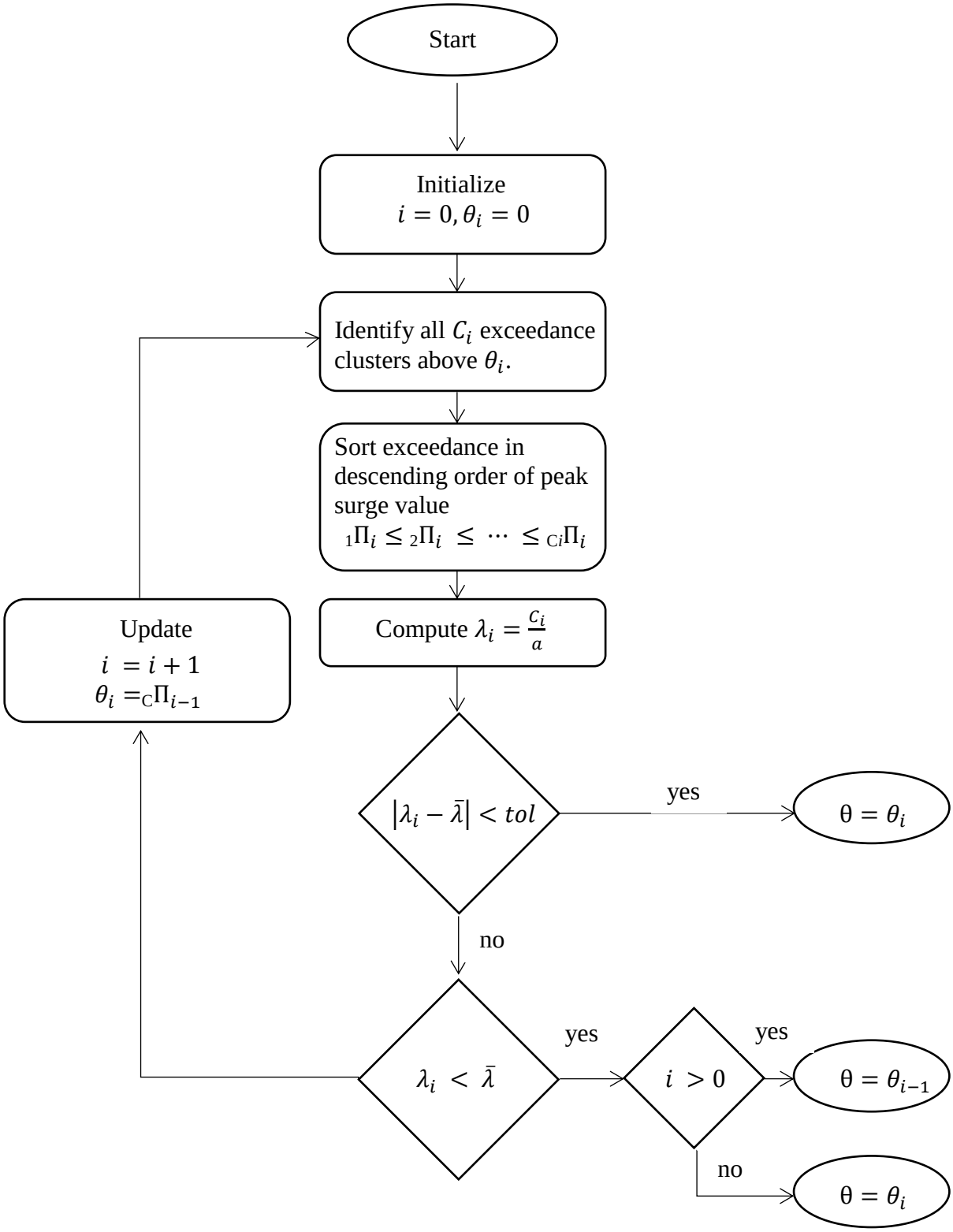

Figure 2.3 Flowchart of threshold algorithm proposed by Lopeman, 2015 [37]

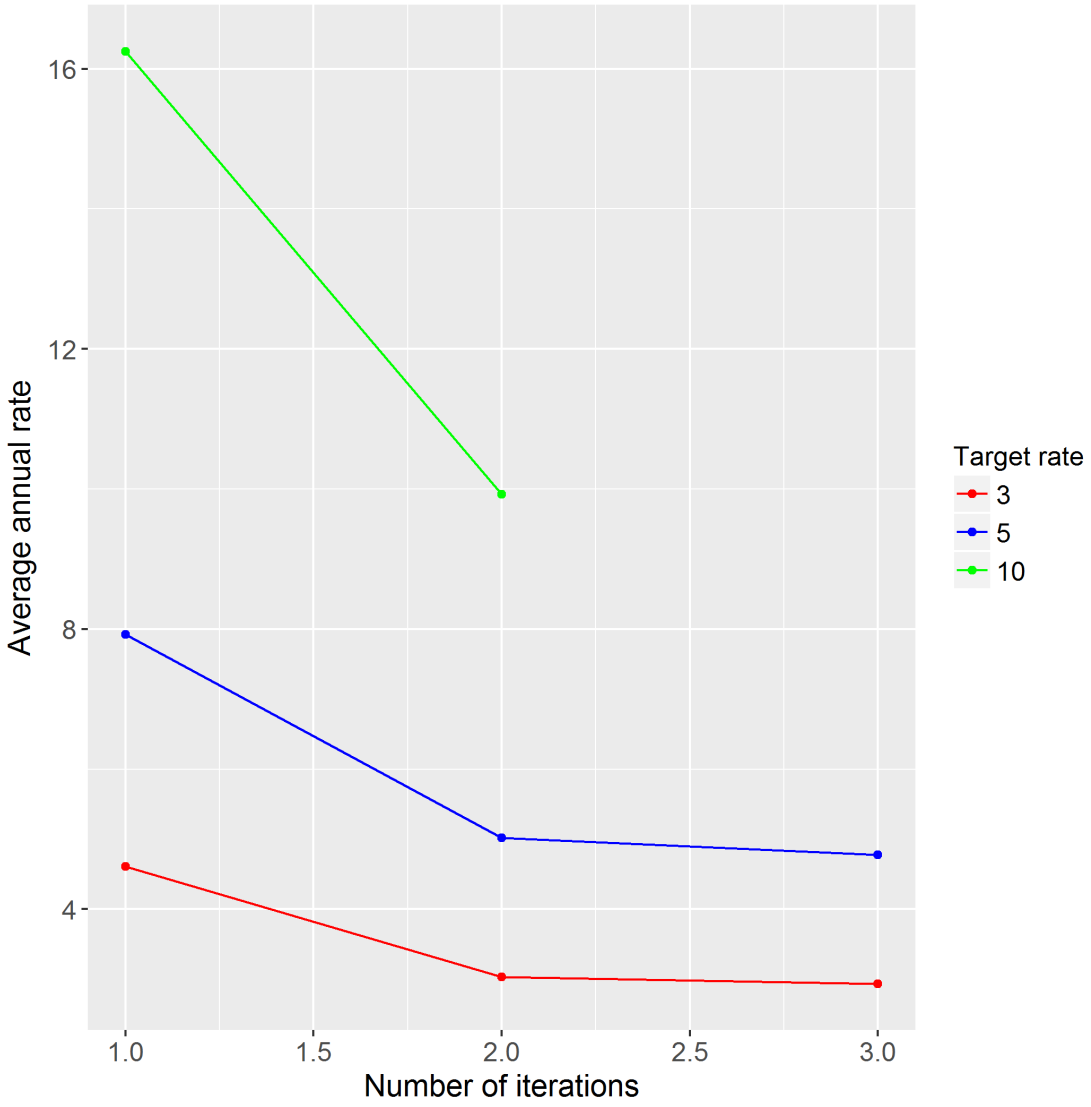

Figure 2.4 Iterative process using the threshold selection

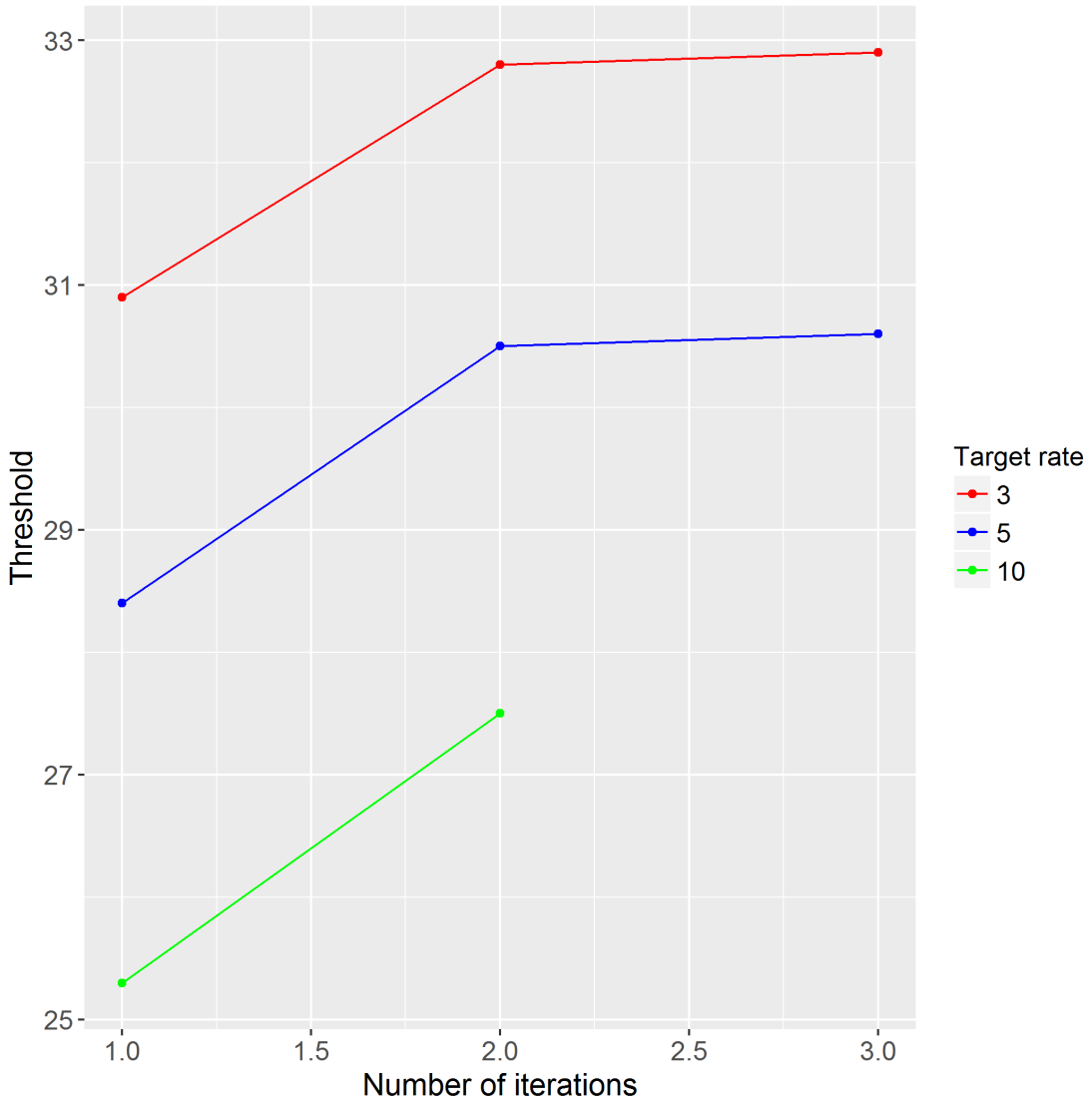

Figure 2.5 Iterative process using the threshold selection (Cont'd)

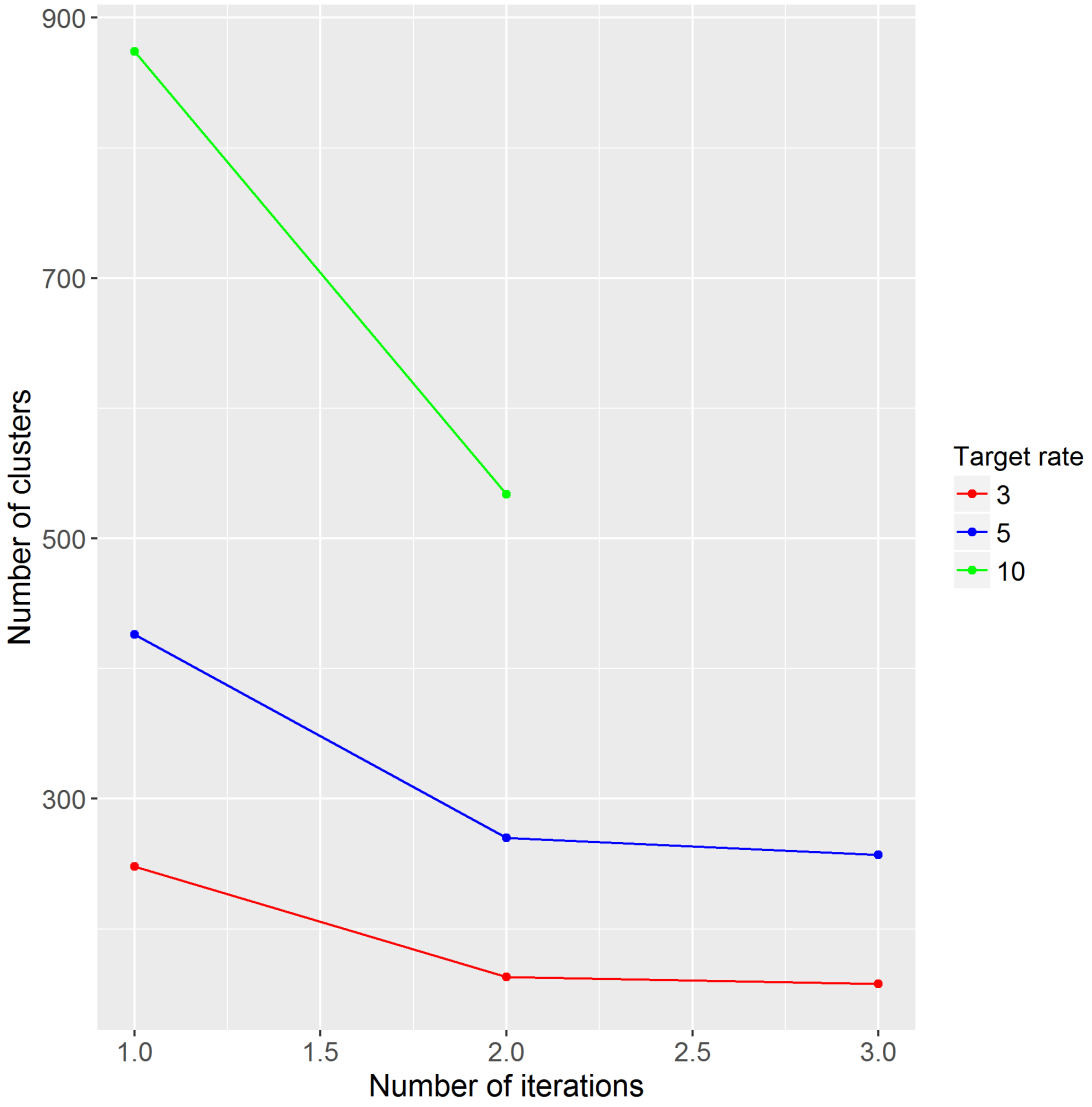

Figure 2.6 Iterative process using the threshold selection (Cont'd)

As shown in Figures 2.4, 2.5, and 2.6, the convergence of the threshold algorithm is pretty fast for all the different target rates selected (3, 5, 10) and the number of iterations needed for convergence ranges from 3 (target rate = 3) to 5 (target rate = 10).

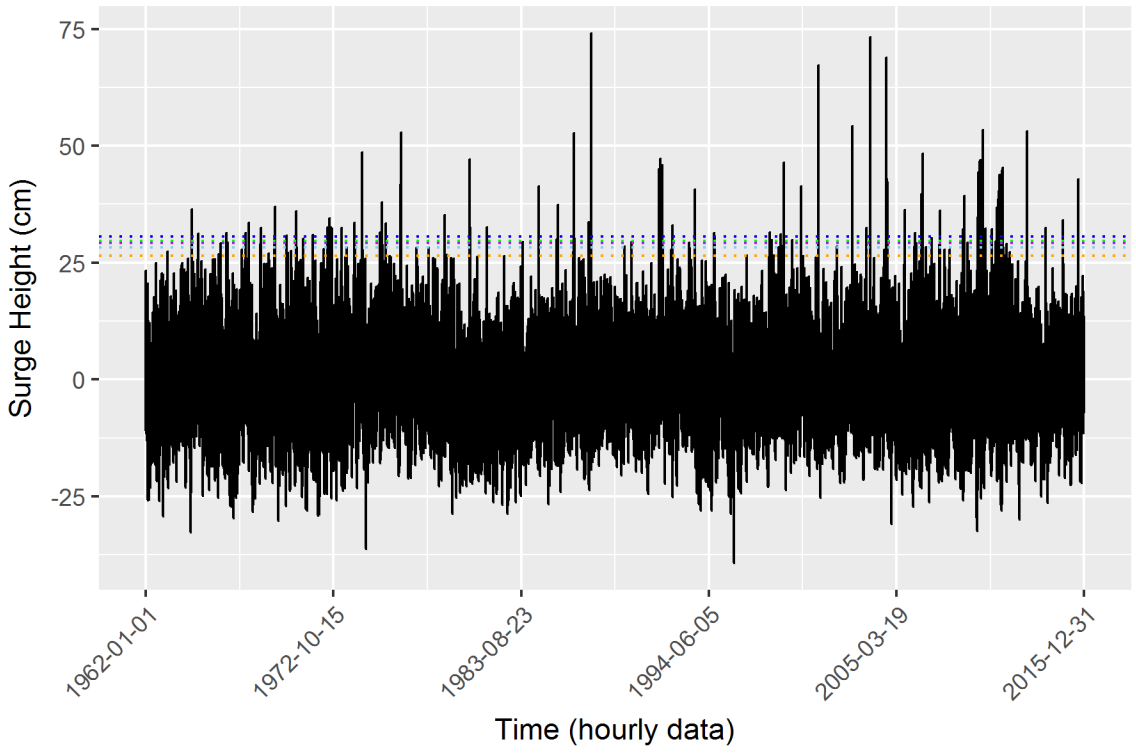

Figure 2.7 Various thresholds considered

Figure 2.7 displays the threshold of 31.2 cm (Target rate: 3.5, number of clusters: 189) in red, the threshold of 30.54 cm (Target rate: 4, number of clusters: 217) in blue, the threshold of 29.56 cm (Target rate: 4.5, number of clusters: 246) in purple, the threshold of 29.15 cm (Target rate: 5, number of clusters: 274) in green, the threshold of 28.33 cm (Target rate: 6, number of clusters: 324) in sky-blue, and the threshold of 26.53 cm (Target rate: 8, number of clusters: 431) in orange. As expected, as the target rate increases, the threshold decreases and as the threshold decreases, the number of clusters increases.

2.4 Clustering of the storm surge data

2.4.1 Relationship among target rate, threshold and clusters

As shown in Figures 2.4, 2.5, and 2.6, the convergence of the threshold algorithm is rapid for all the different target rates that have been tested, namely 3, 5, and 10 storm per years: the number of iterations needed for convergence ranges from 3 (target rate = 3) to 5 (target rate = 10). As expected, the lower the target rate, the lower the number of clusters (storm events) per year and the higher the threshold. If a certain target rate is desired, the strategy is to increase the threshold (excluding all observations below the threshold) until the target rate (in terms of storms per year) is achieved. For example, if the desired target annual rate is 3 storms, the algorithm will converge in 3 iterations and set the threshold level to 32.01 cm; this results in an average of 3 storms per year, for a total of 164 storm events (clusters) over the 54-year timespan covered by the data. Conversely, if the desired target rate is set to 10 storms per year, the chosen threshold is significantly lower (25.43 cm), and the total number of storm events more than triples (539 storm events clusters). Figures 2.8, 2.9, and 2.10 show the stages of the clustering surges (target rate: 5, threshold: 29.15 cm, number of clusters: 274). Figure 2.8 indicates only the number of surges because of the difficulty of visually indicating all surge dates and times.

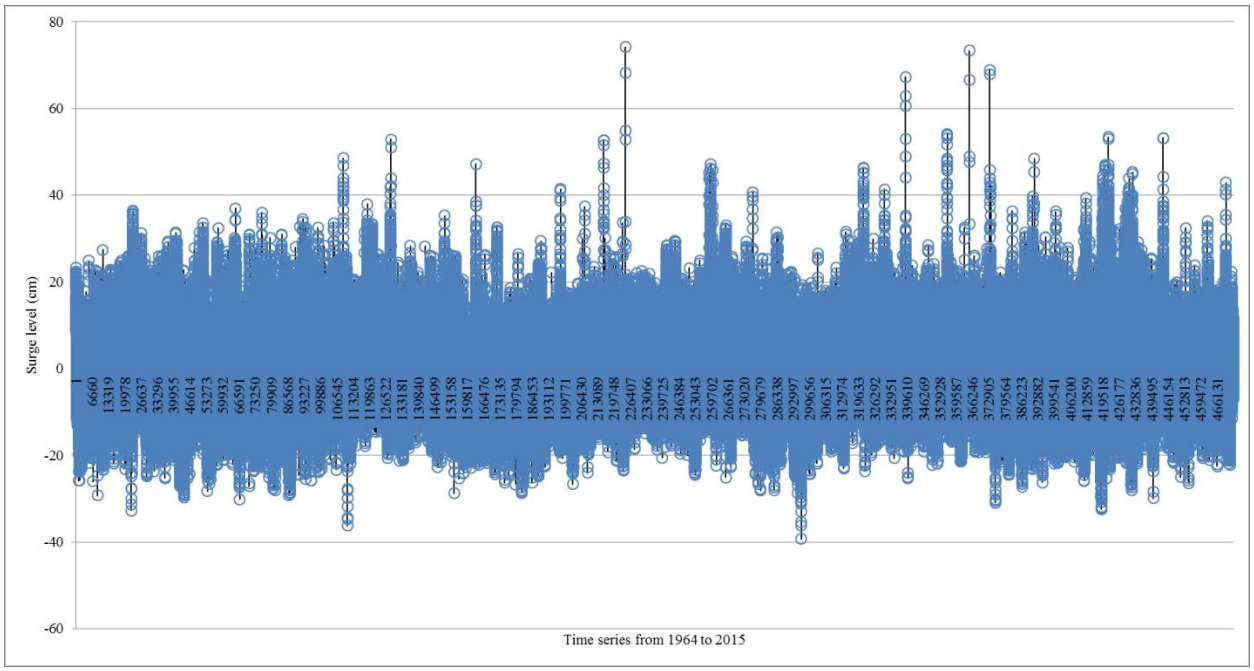

Figure 2.8 Complete set of data before any threshold application from Busan tidal gauge station

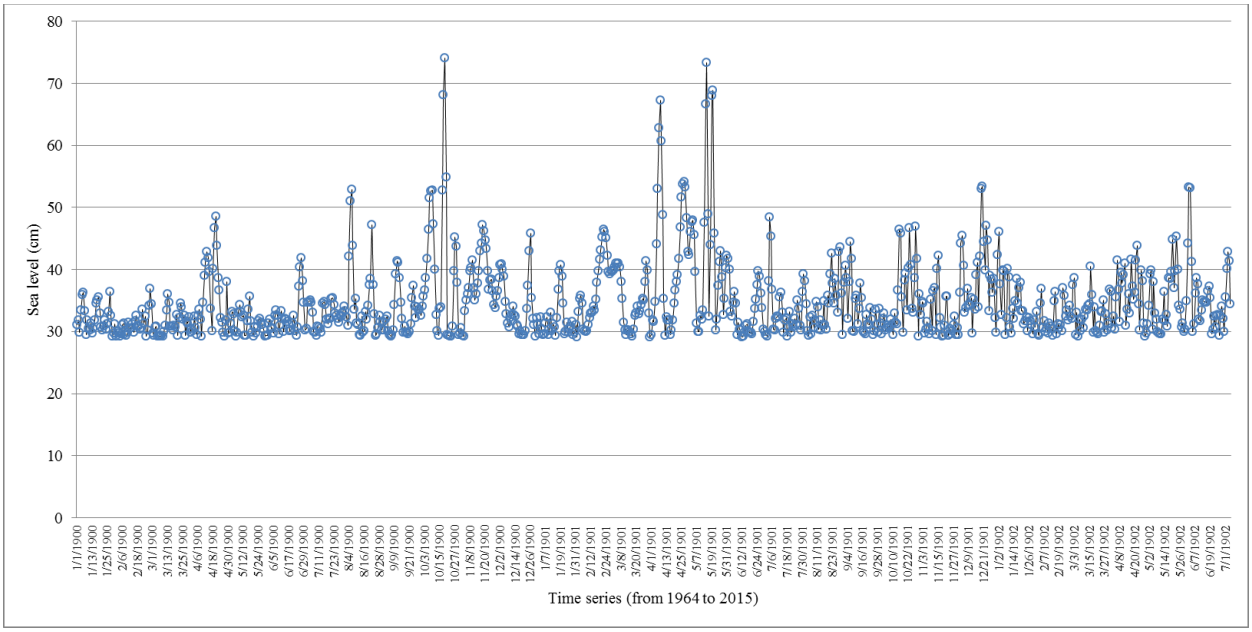

Figure 2.9 Surges above the threshold (29.15 cm) before clustering from Busan tidal gauge station

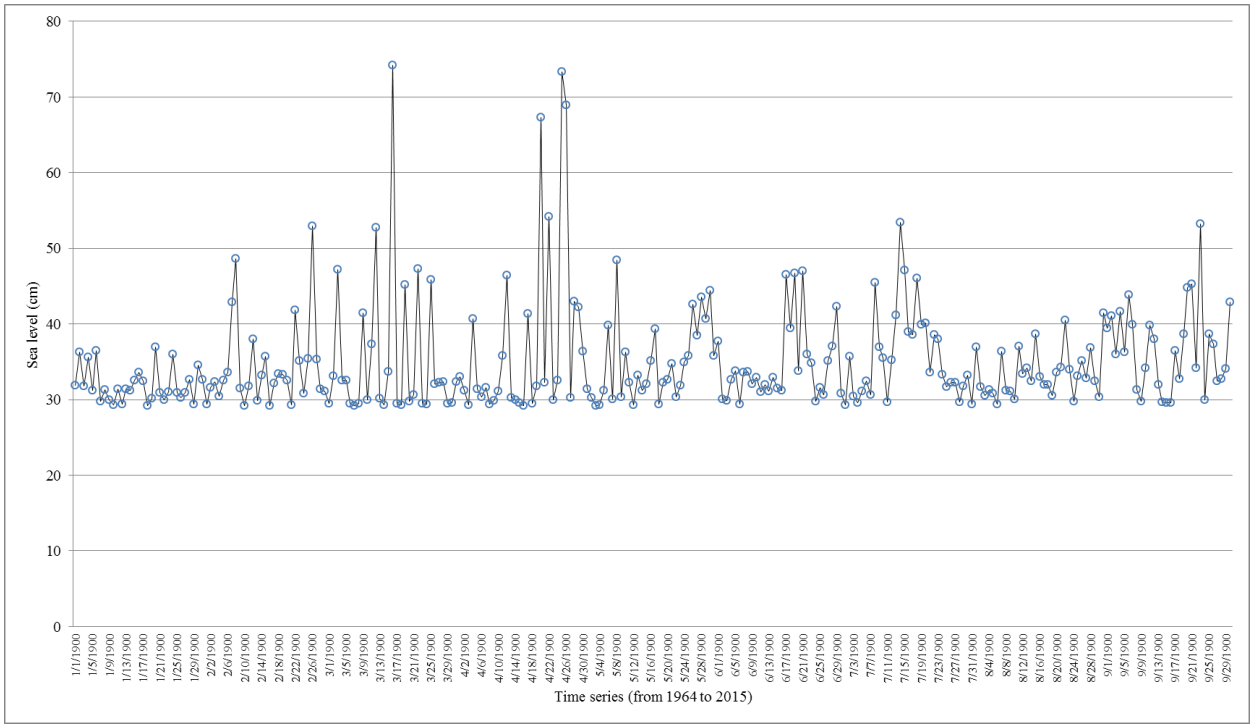

Figure 2.10 Surges above the threshold (29.15 cm) after clustering from Busan tidal gauge station

Figures 2.11 and 2.12 show the stages of the clustering surges for target rate: 3, threshold: 32.01 cm, number of clusters: 164.

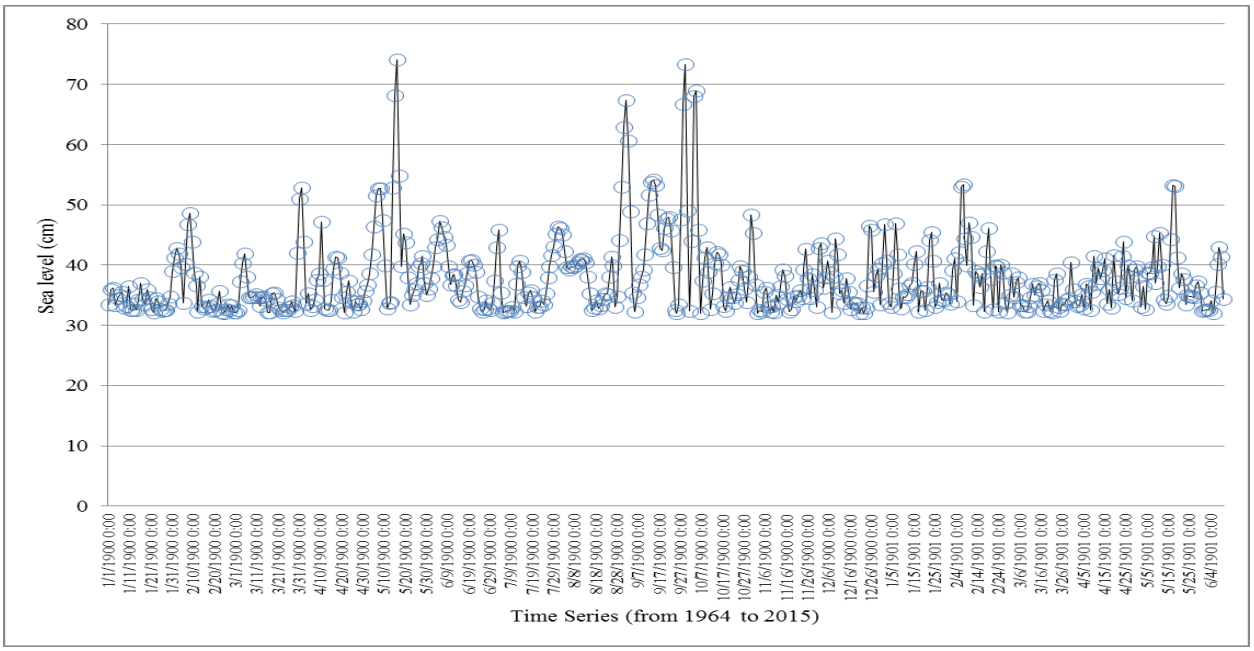

Figure 2.11 Surges above the threshold (32.01 cm) before clustering from Busan tidal gauge station

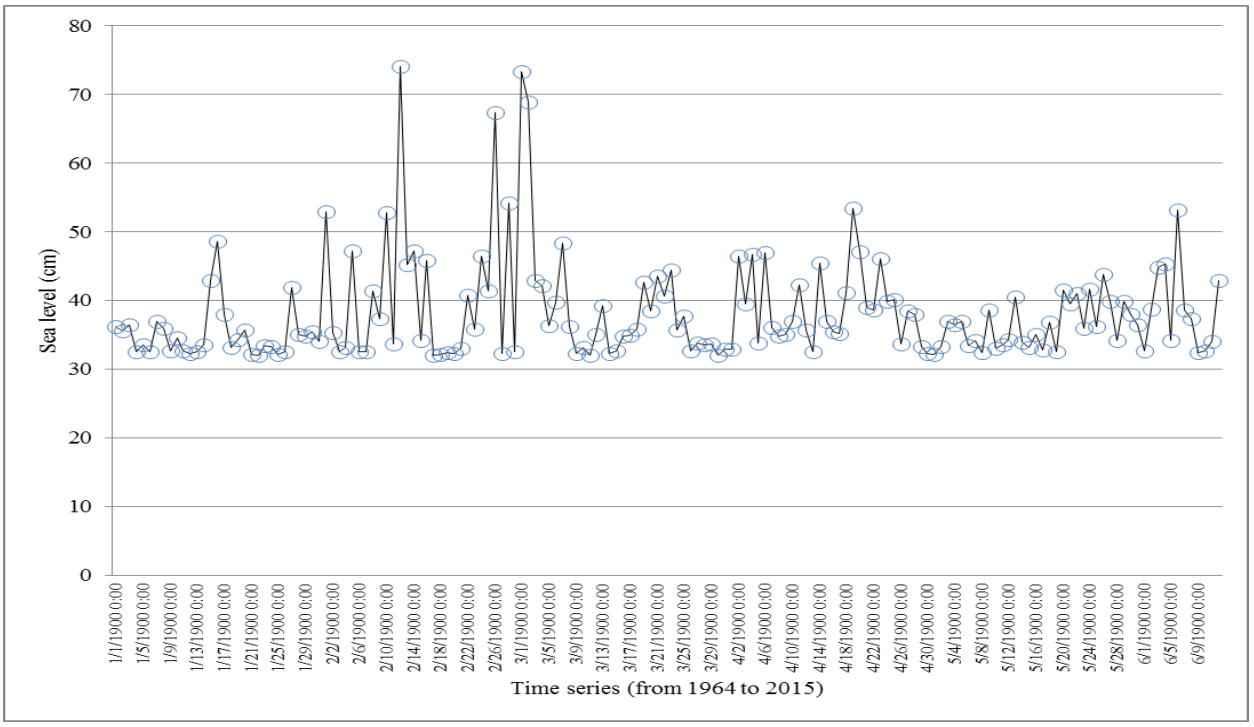

Figure 2.12 Surges above the threshold (32.01 cm) after clustering from Busan tidal gauge station

Figures 2.13 and 2.14 indicate the results of target rate: 10, threshold: 25.43 cm, number of clusters: 164.

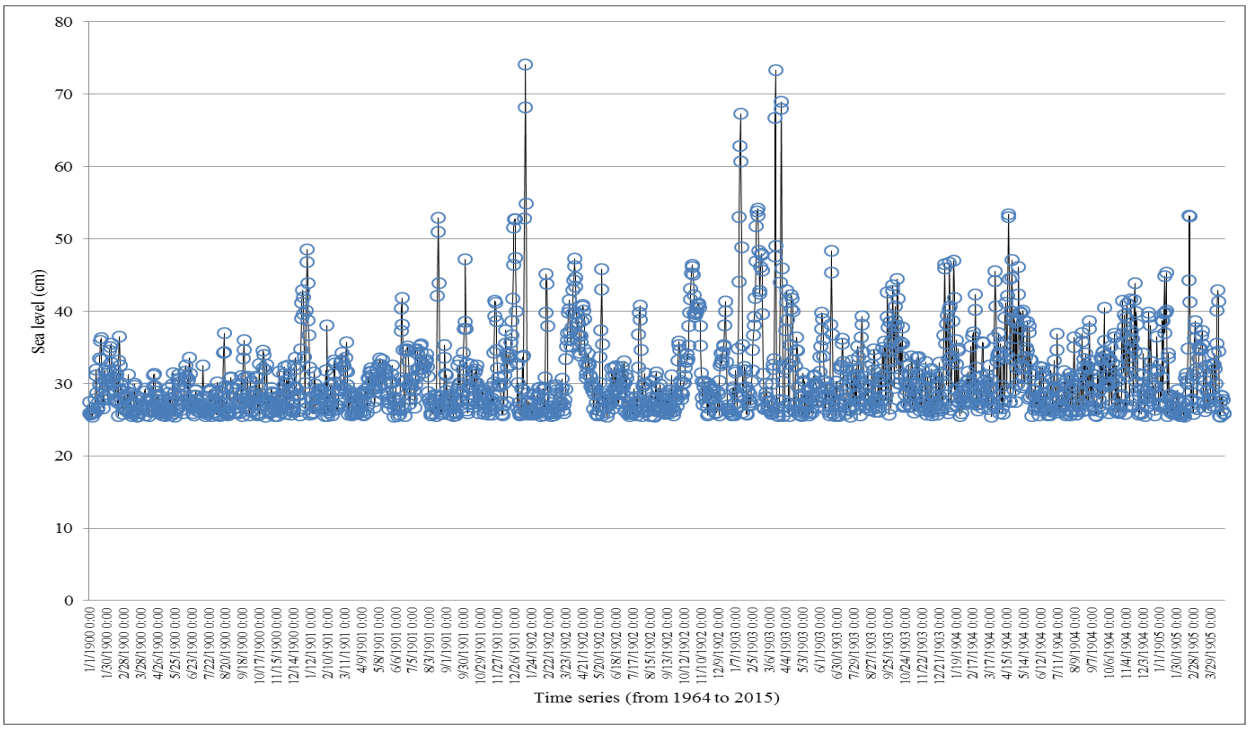

Figure 2.13 Surges above the threshold (25.43 cm) before clustering from Busan tidal gauge station

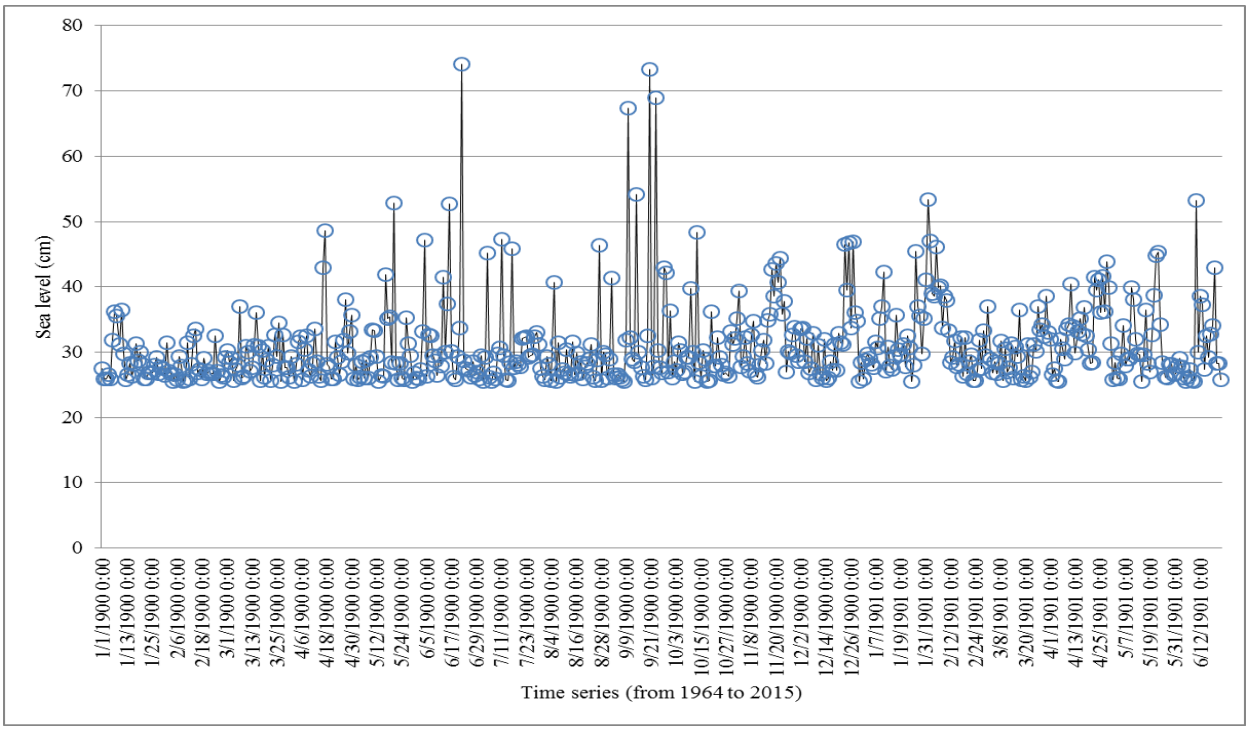

Figure 2.14 Surges above the threshold (25.43 cm) after clustering from Busan tidal gauge station

2.5 The relationship between storm surge parameters

2.5.1 Storm surge parameters

Storm surges are characterized by four major parameters:

- peak time
- peak surge height
- duration
- rise ratio

Peak time follows a gamma distribution because peaks over threshold (POT) produces a Poisson process of exceedance occurrence, and the waiting times of consecutive exceedances in a Poisson process are, by definition, exponentially distributed (Lopeman and Deodatis, 2015) [33], [34], [35], [36].

For the peak time (interarrival times), we use a gamma (exponential) distribution, because:

- the gamma (exponential) distribution converts the arrival process of the storm peaks into a Poisson process (assuming that the arrival maxima of exceedance clusters (storm peaks) are regulated by a Poisson process which is a common assumption/convention of the POT method).

For peak height, a generalized Pareto distribution (GPD) is typically used because some results (representation theorems) from extreme value statistics indicate that, if the cluster maxima follow a Poisson process, then the intensity (height) of the cluster peaks follows a GPD distribution (Lopeman and Deodatis, 2015 [33], [34], [35], [36]; Zhong et al., 2014 [59]). However, a Weibull distribution is applied to peak storm surge heights in this study because it better fits the data (especially with regards to the right tail).

The rise ratio which does not appear to be evenly distributed along the interval [0, 1] follows a beta distribution. A beta distribution was used because:

- the rise ratio is by definition between 0 and 1, and the beta distribution is commonly used to model continuous random variables that take values between 0 and 1 (Lopeman and Deodatis, 2015) [33], [34], [35], [36].

Duration follows a lognormal distribution which was used for the following two reasons (Lopeman and Deodatis, 2015) [33], [34], [35], [36]:

- it models a continuous and positive random variable (the duration, by definition, is positive).
- it is quite flexible as it has two parameters, so it can fit the data better than other distributions with just one parameter. For example, the exponential distribution models a positive random variable, but it has only one parameter.

2.5.2 The relationship between storm surge parameters

Figures 2.15, 2.16, and 2.17 indicate that there is no evident relationship between the rise ratio and the other two variables, duration and exceedance. However, the peak exceedance and the cluster duration appear to have a liner relationship.

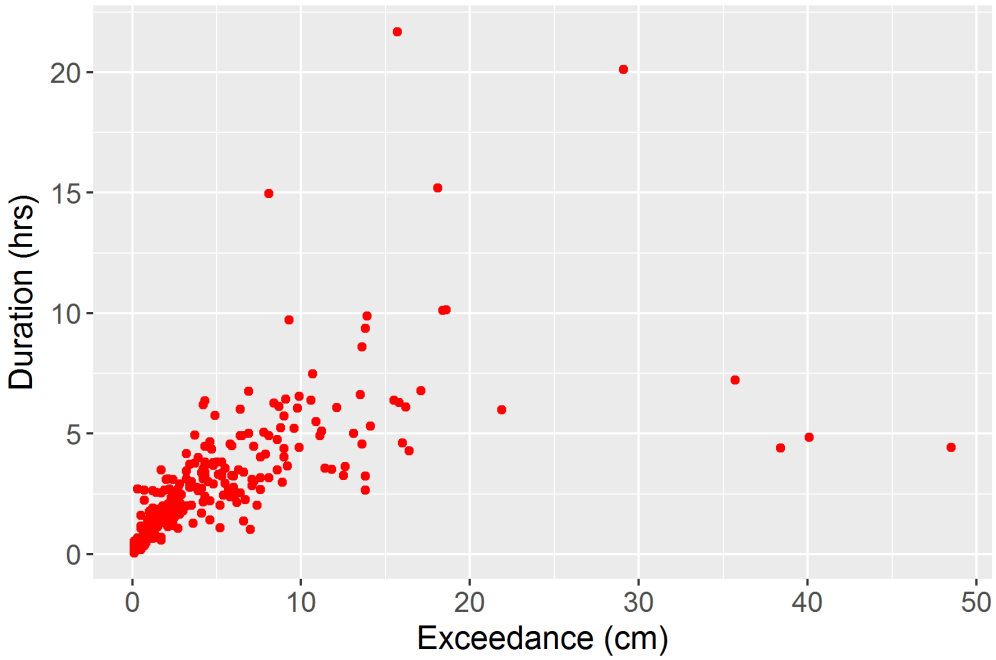

Figure 2.15 Relationship between exceedance and duration

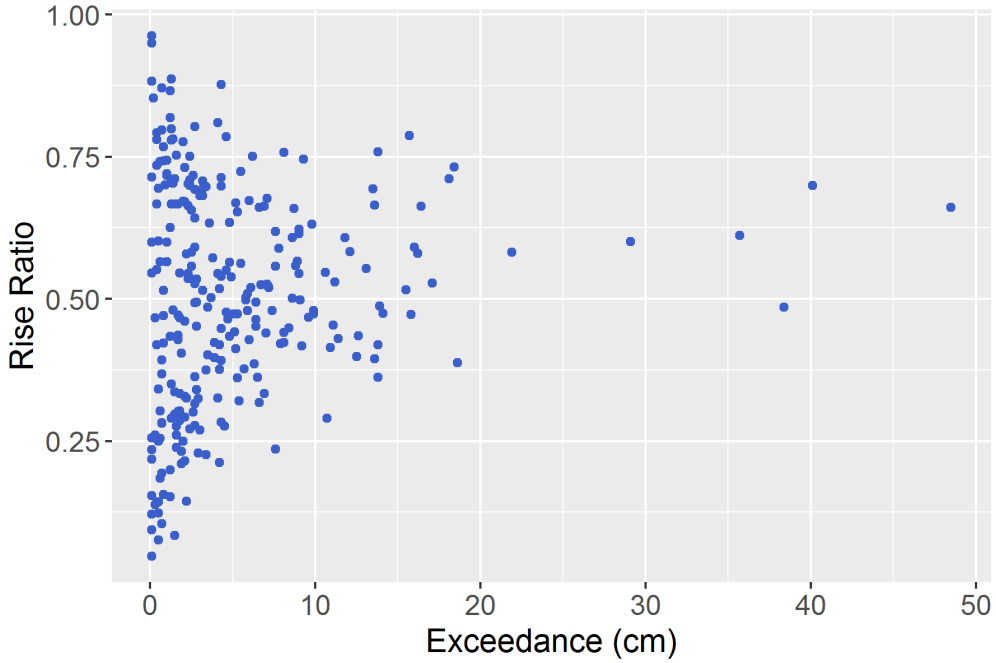

Figure 2.16 Relationship between exceedance and rise ratio

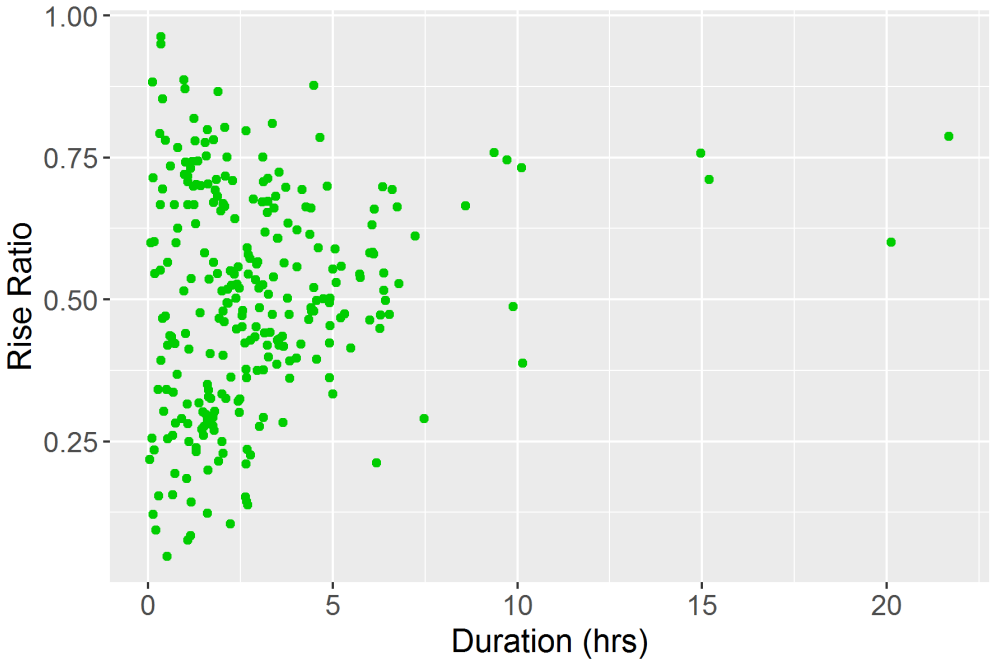

Figure 2.17 Relationship between duration and rise ratio

Chapter 3

3 Storm surge simulation and estimation of the return period of Typhoon Maemi

3.1 Simulation

3.1.1 Fitting the statistical model in this study to the actual surge data

After finding the threshold resulting from a given target rate, interarrival times, rise ratios, peak height, and cluster duration are computed for each exceedance cluster. These calculations are grouped by season and then used to estimate the parameters of the statistical model via maximum likelihood. In particular, for each season, interarrival times are fitted with an exponential distribution; rise ratios are fitted with a beta distribution; peak height is fitted with a Weibull distribution where the Weibull location parameter is equal to the threshold. A detailed description of the methodological aspects is provided in what follows.

Maximum likelihood estimation

Assume an independent and identically distributed data sample $(x_1, \dots, x_n)$ is observed from a population with a distribution of interest parametrized by an unknown variable θ, which the study wants to estimate. The maximum likelihood estimator $\hat{\theta}_{MLE}$ is defined as:

$$\hat{\theta}_{MLE}(x_1, \dots, x_n) = \ argmax_{\theta_0} \prod_{i=1}^{n} f(x_i; \theta_0) \ ,$$

where $f(\cdot\,; \theta)$ denotes the probability density function of the distribution of interest, parametrized by θ_0. The distributions of interest for the data in this study are chosen as follows:

- $T_i \sim Exponential(\lambda)$, where T_i denotes the interarrival time between the peak of the $i-1$-th cluster and the peak of the i-th cluster. This distributional assumption is equivalent to assuming that a Poisson process governs the surge peak arrivals.
- $\Phi_i \sim Beta(\alpha, \beta)$, where Φ_i denotes the rise ratio of the i-th cluster.
- $\prod_i \sim GPD(\xi, \sigma, \theta^*)$, where GPD denotes the generalized Pareto distribution, $\prod_i$ denotes the peak surge height of the i-th cluster, and θ^* is the selected threshold.

For the exponential distribution (i.e., interarrival times), the exact solutions of the maximization problem stated above can be derived in closed form. For the GPD distribution (i.e., peak exceedances) and the beta distribution (i.e., rise ratios), the problem is solved numerically. A full description of the estimation algorithm is detailed below Table 3.1.

Algorithm: *Maximum likelihood estimation for interarrival times, rise ratios, and peak exceedances.*

Input: Observed interarrival times $t_1, \dots, t_C$ of the clusters' surge peaks.

Observed rise ratios $\phi_1, \dots, \phi_C$.

Observed peak surge heights $\gamma_1, \dots, \gamma_C$.

Number of clusters C.

Threshold rate θ^*.

Output: Maximum likelihood estimates of the model parameters:

$\hat{\lambda}_{MLE}, \hat{\alpha}_{MLE}, \hat{\beta}_{MLE}, \hat{\xi}_{MLE}, \hat{\sigma}_{MLE}$.

Compute the MLE estimate $\hat{\lambda}_{MLE}$ for the exponential interarrival rate λ as:

$$\hat{\lambda}_{MLE} = \left(\sum_{c=1}^{C} t_c \right)^{-1}.$$

Compute the MLE estimates $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ for the Beta parameters α and β, by solving the following first order equations numerically:

$$C\left(\psi\left(\hat{\alpha}_{MLE} + \hat{\beta}_{MLE}\right) - \psi(\hat{\alpha}_{MLE})\right) + \sum_{c=1}^{C} \log \phi_i = 0;$$

$$C\left(\psi\left(\hat{\alpha}_{MLE} + \hat{\beta}_{MLE}\right) - \psi(\hat{\alpha}_{MLE})\right) + \sum_{c=1}^{C} \log(1 - \phi_i) = 0,$$

where $\psi(\cdot)$ denotes the digamma function.

Compute the MLE estimates $\hat{\xi}_{MLE}$ and $\hat{\sigma}_{MLE}$ for the GPD parameters ξ and σ (for more details on this estimation, see documentation provided by *ismev* package.)

Return the MLE estimates $\hat{\lambda}_{MLE}, \hat{\alpha}_{MLE}, \hat{\beta}_{MLE}, \hat{\xi}_{MLE}, \hat{\sigma}_{MLE}$.

Table 3.1 Maximum likelihood estimation for interarrival times, rise ratios, and peak exceedances

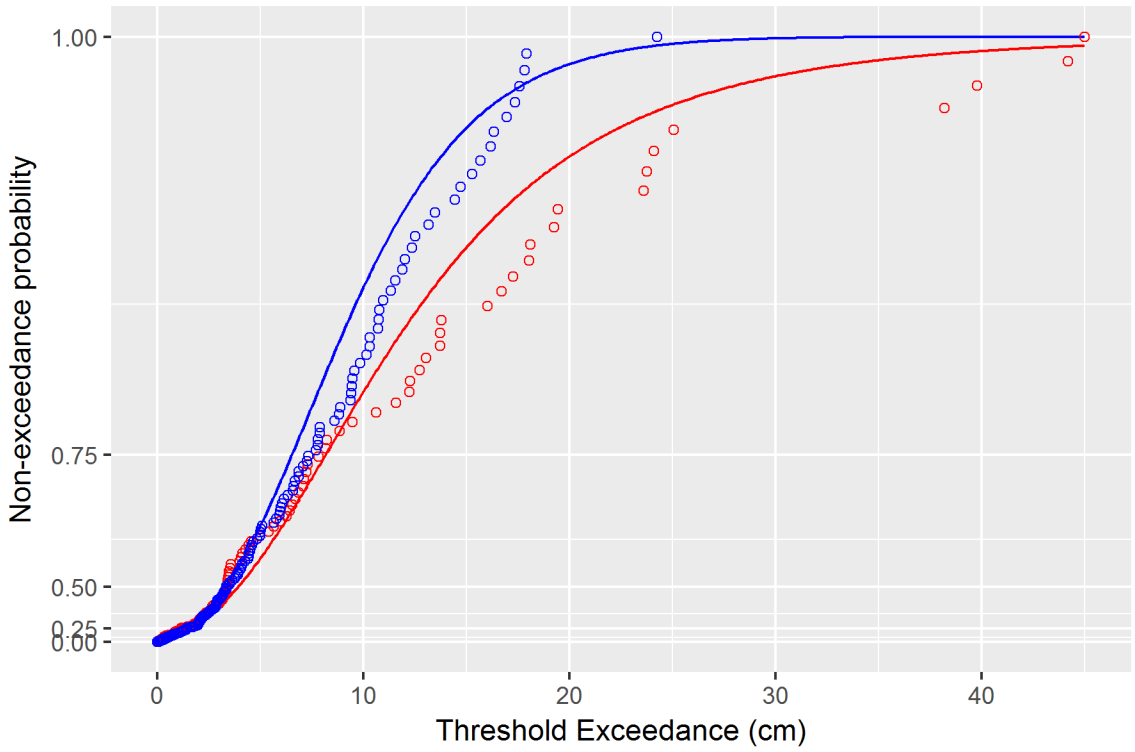

Figure 3.1 Generalized Pareto distribution (GPD) at target rate 5

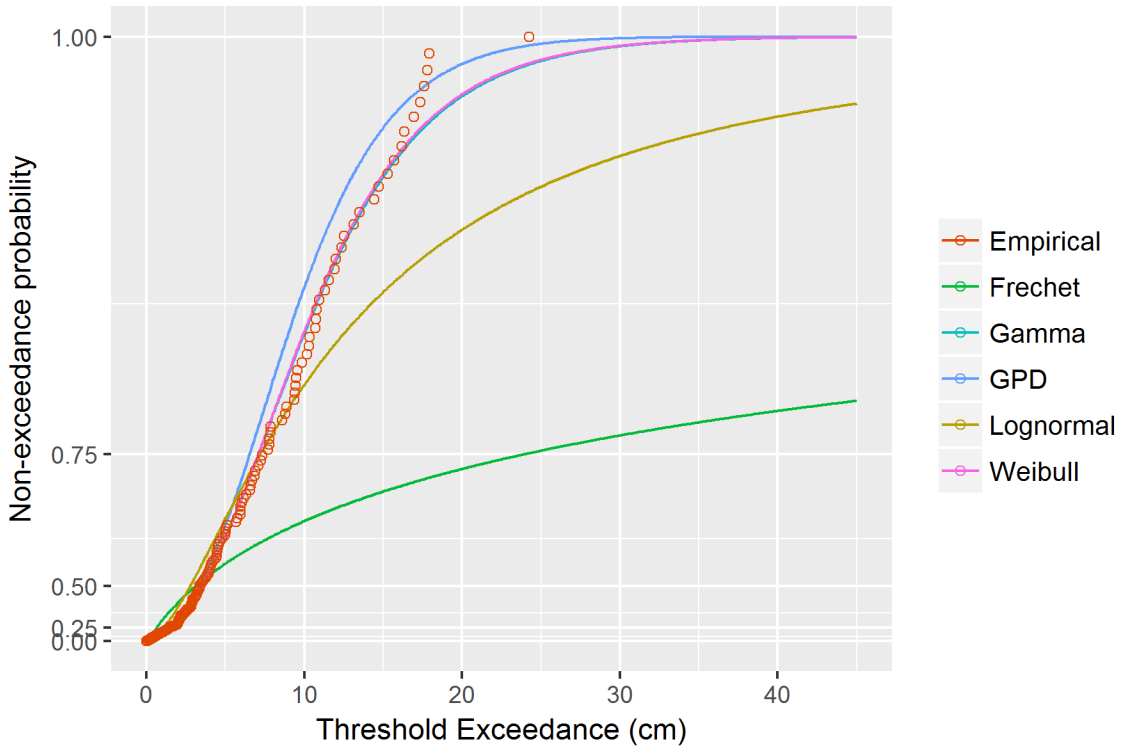

Figure 3.2 Various fits of distributions for cold season at target rate 5

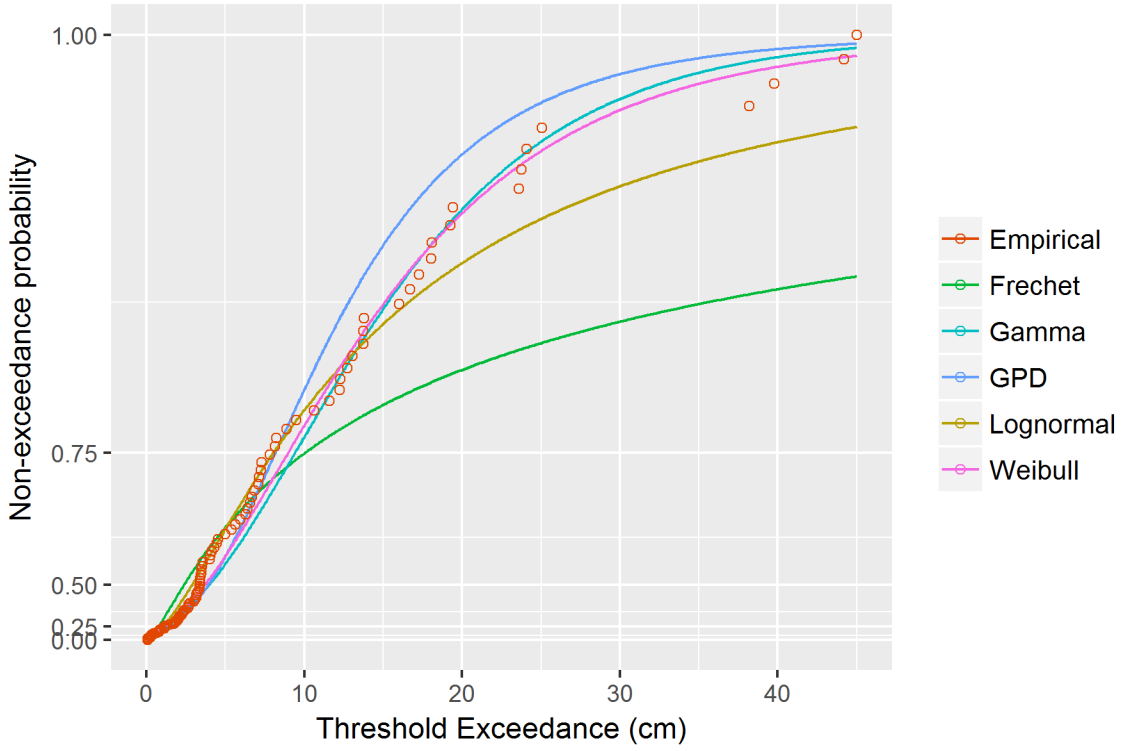

Figure 3.3 Various fits of distributions for warm season at target rate 5

Figure 3.1 shows that the GPD cumulative distribution function (estimated by maximum-likelihood) and the empirical distribution function (dots). Each dot represents the observed proportion of exceedances below a certain height in a given season (blue color: cold season, red color: warm season), while the corresponding value on the fitted line (of the same season) gives the probability that the exceedances are below that height according to the estimated GPD distribution.

Figures 3.2 and 3.3 indicate that the Gamma and Weibull distributions fitted to the data for the cold and warm seasons via maximum likelihood better than the GPD. According to the minimum square error (MSE), the Gamma and Weibull distributions provide the best fit to the data in each season and for a variety of target rates. This confirms the graphical considerations suggested by the plots.

3.1.2 Monte Carlo simulation

After determining the distribution parameters of the statistical model, a time series of total water level activity can be simulated up to any desired future horizon.

3.1.2.1 Peak Times

The first step of the simulation is the generation of the two alternating semiannual Poisson processes that regulate the peak times for each season. The simulation of the Poisson process is done, as usual, by simulating the exponentially distributed interarrival times under the estimated rates: each peak time is the sum of the interarrival times between the peaks that preceded it. According to our definition of seasons, the warm season has 183 days, and the cold season an average of 182.25 days due to leap years. Hence, to simulate N years, the Poisson process for the warm season is simulated for N times 183 days, while that of the cold season for

N times 182.25 days. Because the simulation results are contained in two vectors of peak times, the study uses a stitching procedure to build the final vector of peak times throughout the year. The final vector is made of 2*N semiannual blocks that reflect the physical alternation of seasons, and is built as follows:

- The first block is established by all the simulated warm peak times within the first 183 days;
- The second block is established by all the simulated cold peak times within the first 182.25 days, rescaled to take into account that 183 days have already been simulated for the first warm season;
- the third block is established by all the simulated warm peak times between the 184th day and the 368th day, rescaled to take into account that one year has already been simulated (i.e., first warm season and first cold season);
- and so on.

This procedure is used until all the N warm seasons and N cold seasons have been stitched together.

3.1.2.2 Rise ratios and Peak heights

Every peak time in the final vector generated using the Poisson stitching process is associated to an exceedance cluster; hence, the study can now simulate as many rise ratios and peak heights as the number of peak times in the final vector. For each exceedance cluster, its rise ratio and peak height are sampled from the Beta and Weibull distributions using the parameters that were previously estimated for the corresponding exceedance's season.

3.1.2.3 Duration and removal of overlapping clusters

For duration, the study regresses the duration variance of the peak height via linear least squares (LLS). For the duration mean, a linear regression is run using peak height except in cases of heteroscedastic data. In case of heteroscedastic data, weighted linear least squares (WLLS) is run. For instance, when the target rate is set to four storms per year, the variance of the cluster duration increases with the peak height only in the warm season. Hence, WLLS is used for the warm season data and LLS for the cold season data. For each exceedance cluster, the mean and variance of the duration is computed as a lognormal random variable at its simulated peak height.

Because the duration of an exceedance cluster depends only on its peak height and the peak heights are simulated independently, some overlapping exceedance clusters may occur. To solve this problem, all the clusters are checked for overlapping, and every time an overlap is found, one of the two overlapping clusters is randomly selected and discarded.
Note that, in general, a lower threshold will yield a higher annual rate, which in turn will result in more frequent surge events and thus more opportunities for overlap. If the threshold – or rather, the target rate – is properly selected, the removal of overlapping clusters will affect the annual storm rate only marginally.

3.1.2.4 Final outcome

The outcome of the Monte Carlo simulations is a set of exceedance clusters that span N years. Each exceedance cluster is characterized by a surge peak happening at a certain time and with a certain height, and by a rise ratio and duration. Other measures of interest (e.g., first date of the cluster, the surge level at a specific hour during a specific storm event, etc.) can be then computed using for these four parameter statistics.

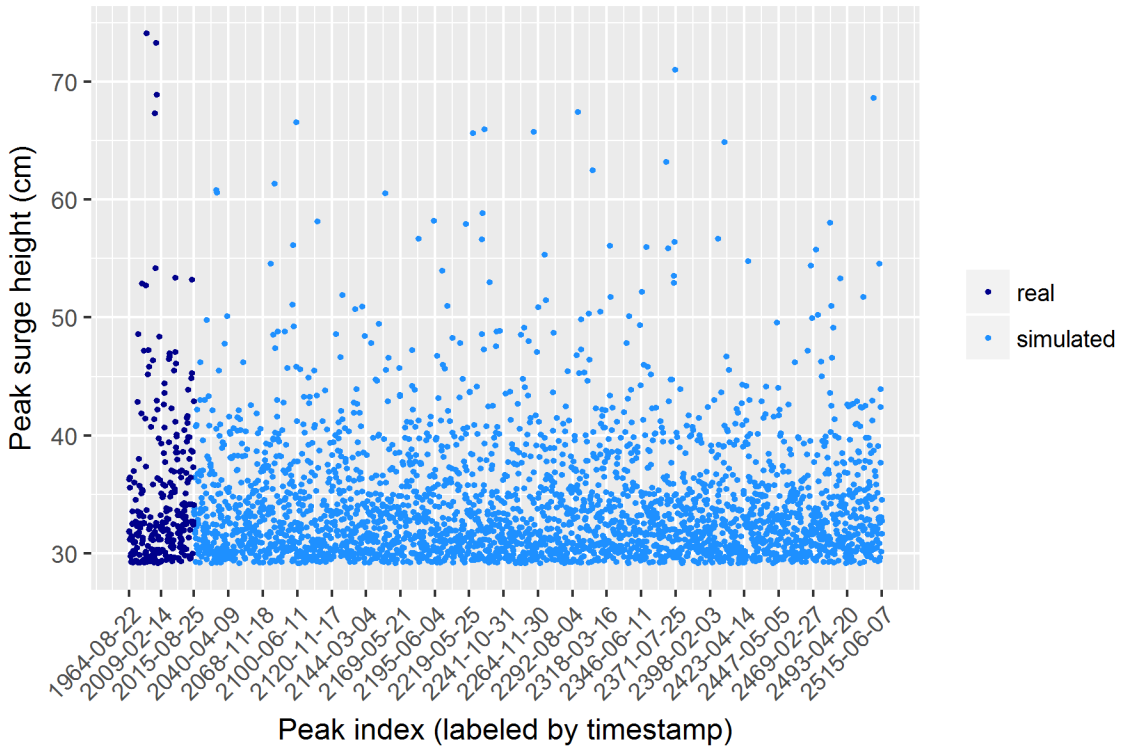

Figure 3.4 Simulation of peaks following the GPD over 500 years

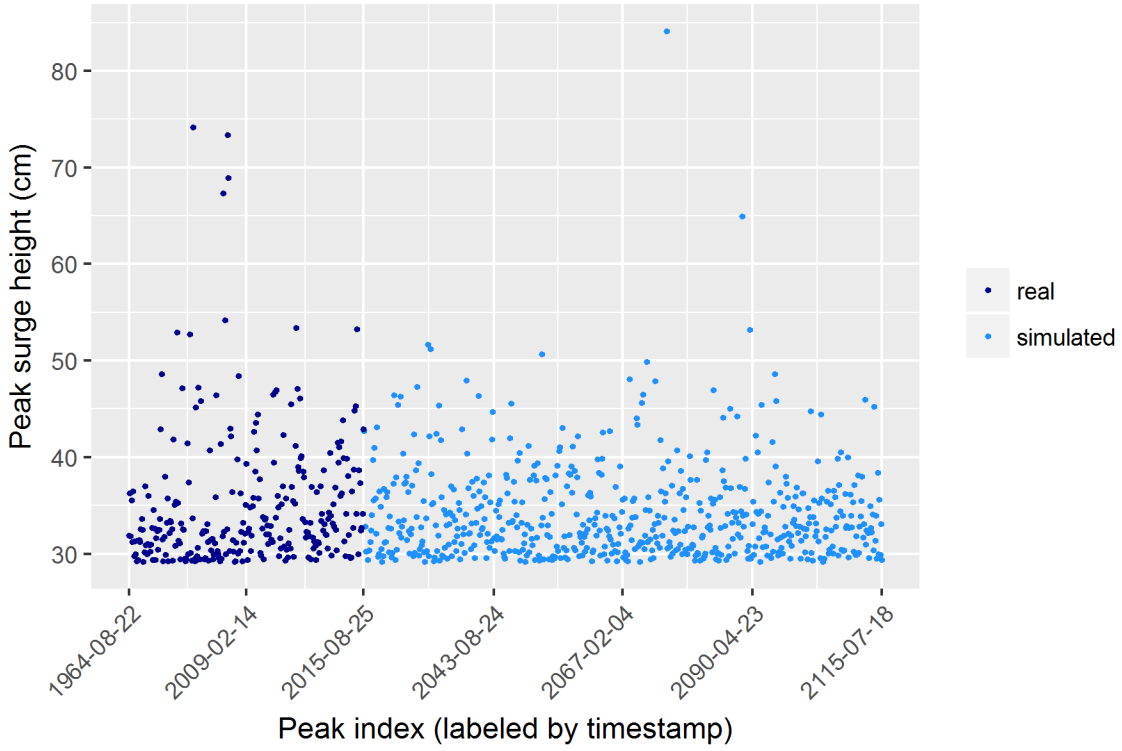

Figure 3.5 Simulation of peaks following the GPD over 100 years

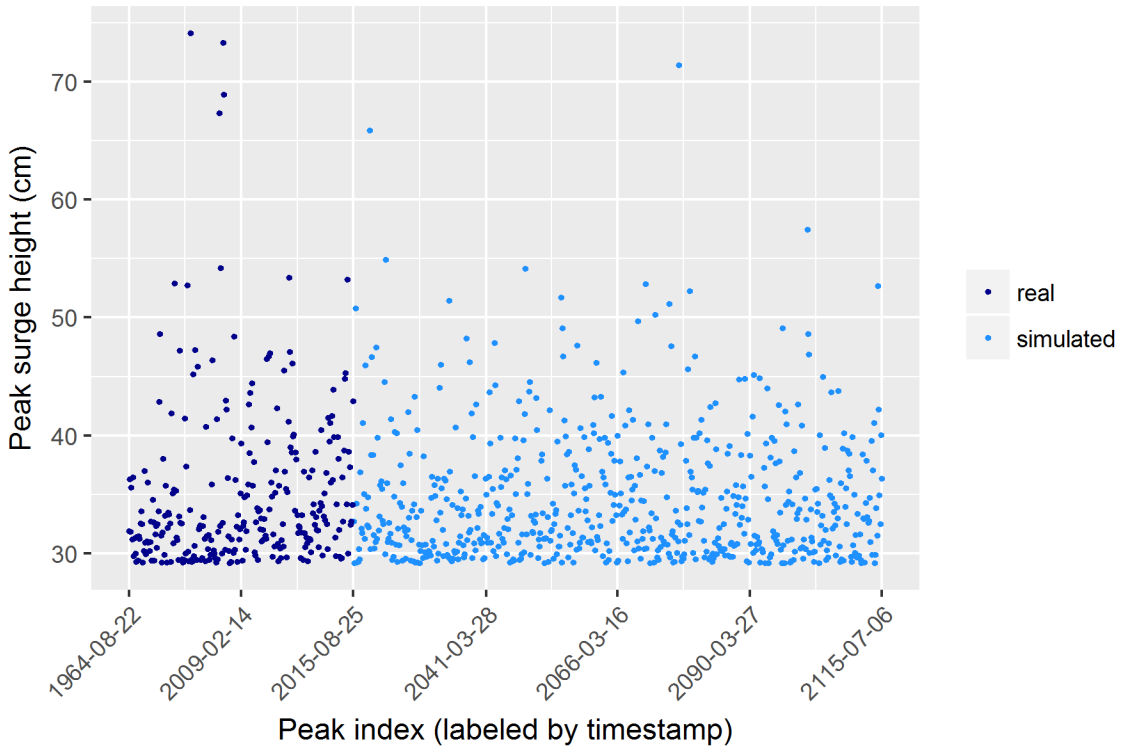

Figure 3.6 Simulation of peaks following the Gamma distribution over 100 years

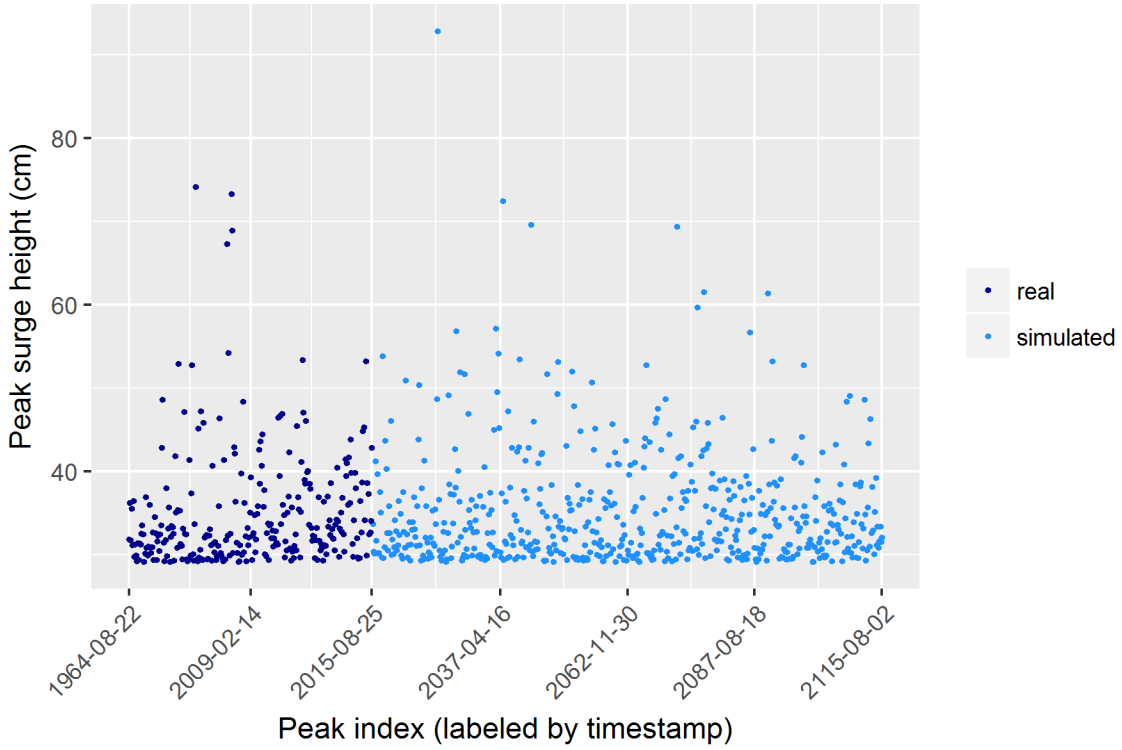

Figure 3.7 Simulation of peaks following the Weibull distribution over 100 years

After estimating the distribution parameters of the statistical model, we can simulate a time series of total water level activity up to any desired future horizon. Figure 3.4 above shows the simulation with target rate (annual rate) of 5 and 10,000 years using GPD. Also, Figures 3.5, 3.6, and 3.7 show the simulations using GPD, Gamma, and Weibull distribution with target rate (annual rate) 5 and 100 years for comparison.

In Figures 3.5, 3.6, and 3.7, Monte Carlo simulations of the peaks are compared with the observed peaks (for the given target rate and chosen threshold) over the 54 years covered by the dataset. Each simulated peak is characterized by the peak time and the peak height; the times are obtained by sampling the interarrival times from the estimated exponential distribution, while the peak heights are sampled from the estimated Weibull distribution.

For instance, assume that we want to simulate peaks for the next 100 years. Then we proceed as follows:

1. We first sample the interarrival times $T_1, T_2, \ldots$ from an exponential distribution with the rate that we previously estimated via maximum likelihood. Recall that, in general, the interarrival time T_i denotes the time elapsed between the $(i-1)$-th and the i-th simulated peak. We simulate as many interarrival time as needed to cover the timespan of interest—that is, we simulate the interarrival times until $\sum T = 100$.
2. Next, we compute the arrival times of the simulated peaks. In particular, the arrival time S_i of the i-th peak is computed as the sum of the interarrival times between all the peaks that precede the i-th—that is, $S_i = \sum_{j=1}^{i} T_j$.
3. Finally, we simulate the peak heights $H_1, H_2, \ldots$ from a Weibull distribution with the parameters that we previously estimated by maximum likelihood.

At the end of this procedure, we will have N_{sim} pairs $(S_i, H_i)_{i\geq 1}$ of peak times and peak heights that fully charaterize the simulated peaks. Note that, in the first two steps of this procedure, we are simulating a Poisson Process of peak arrivals. It should also be noted that what is described above is a simplified version of the actual procedure that we implemented, as we also took into account different seasons (cold vs. warm). However, this simplified version provides a good understanding of the main steps we followed when running Monte Carlo simulations of peaks.

3.1.3 Bootstrapping process

Given the complex relationship between exceedance clusters and the return period estimates, the study deploys bootstrapping for this last phase of the analysis. The bootstrap estimation procedure of the return period works as follows:

- 200 bootstrap simulations are created: each simulation is obtained by resampling the original input dataset through replacement of the total water level of the simulated clusters. Hence, each bootstrap simulation is a vector of total water levels with as many coefficients as the number of exceedance clusters that have been simulated.
- Each simulation is stored in a column of a matrix, in descending order, with all total water levels smaller than the N-th greatest water level of that simulation removed. As a result, the study produces a matrix with 200 columns and N rows, where each column is given by the first N water levels of a bootstrap simulation. Finally, each row of this matrix is stored in ascending order.
- The $1-\alpha$ percent confidence interval for the N/i return level can be extracted from the matrix by selecting the levels at row i with columns $200 \cdot \alpha/2$ and $200 \cdot (1-\alpha/2)$, respectively.

- In a similar but slightly more complicated fashion, the $1-\alpha$ percent confidence interval for the return period of a storm with a given total water level y can be built as follows:
    - find the first row t such that the water level on column $200 \cdot (1-\alpha/2)$ is smaller than y;
    - find the first row s such that the water level on column $200 \cdot \alpha/2$ is greater than y;
    - the confidence interval of interest is (t, s).

3.2 Estimation of Typhoon Maemi return period

For a better understanding of estimation return period, an example is provided below. Suppose that N years are simulated with 14 storms with different random peaks. To estimate the return period of a storm with peak height 29 cm, the following question must be asked: over N simulated years, how many times was a storm with peak of at least 29 cm observed? Suppose the answer is 14 time over the simulated N years. Hence, the return period of this storm is N/14 years. Similarly, to estimate the return period of a storm with peak height 73 cm, the following question must be asked: over the N simulated years, how many times was a storm with peak that of at least 73 cm observed? Suppose the answer is 1 times for the simulated N years. Hence, the return period of this storm is N/1 years.

Figure 3.8 and Table 3.2 show the above example. As seen in Figure 3.8, 14 times every N simulated years, the return period of 29 cm is $\frac{N}{14}$. Similarly, 13 times every N simulated years , the return period of 30 cm is $\frac{N}{13}$; 11 times every N simulated years, the return period of 32 cm is $\frac{N}{11}$; 10 times every N simulated years, the return period of 35 cm is $\frac{N}{10}$; and 1 time every N simulated years, the return period of 73 cm is $\frac{N}{1}$.

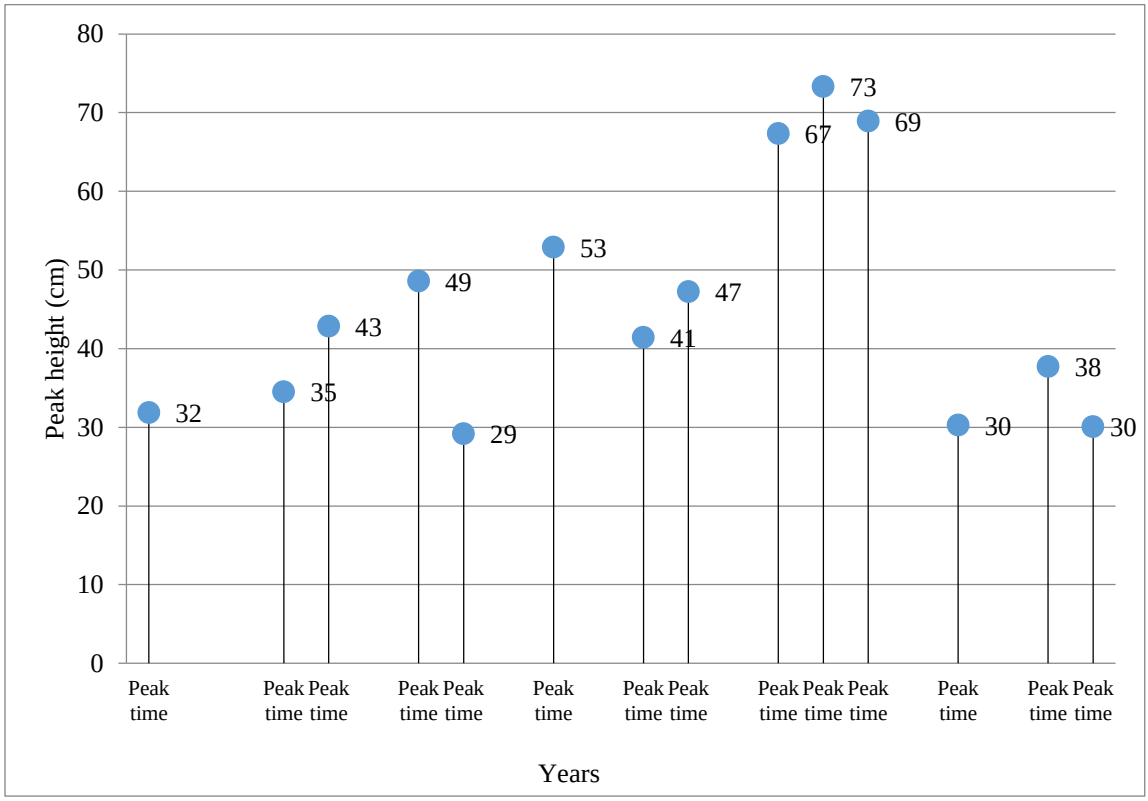

Figure 3.8 Randomly selected peak heights

Peak height (cm)	Return period (years)	Peak height (cm)	Return period (years)
29	$\frac{N}{14}$	47	$\frac{N}{6}$
30	$\frac{N}{13}$	49	$\frac{N}{5}$
32	$\frac{N}{11}$	53	$\frac{N}{4}$
35	$\frac{N}{10}$	67	$\frac{N}{3}$
38	$\frac{N}{9}$	69	$\frac{N}{2}$
41	$\frac{N}{8}$	73	$\frac{N}{1}$
43	$\frac{N}{7}$		

Table 3.2 Return period of randomly selected peak heights (N = 1000 years)

In Table 3.2, the return periods are calculated when N = 1000 years and the results shown in Figure 3.9.

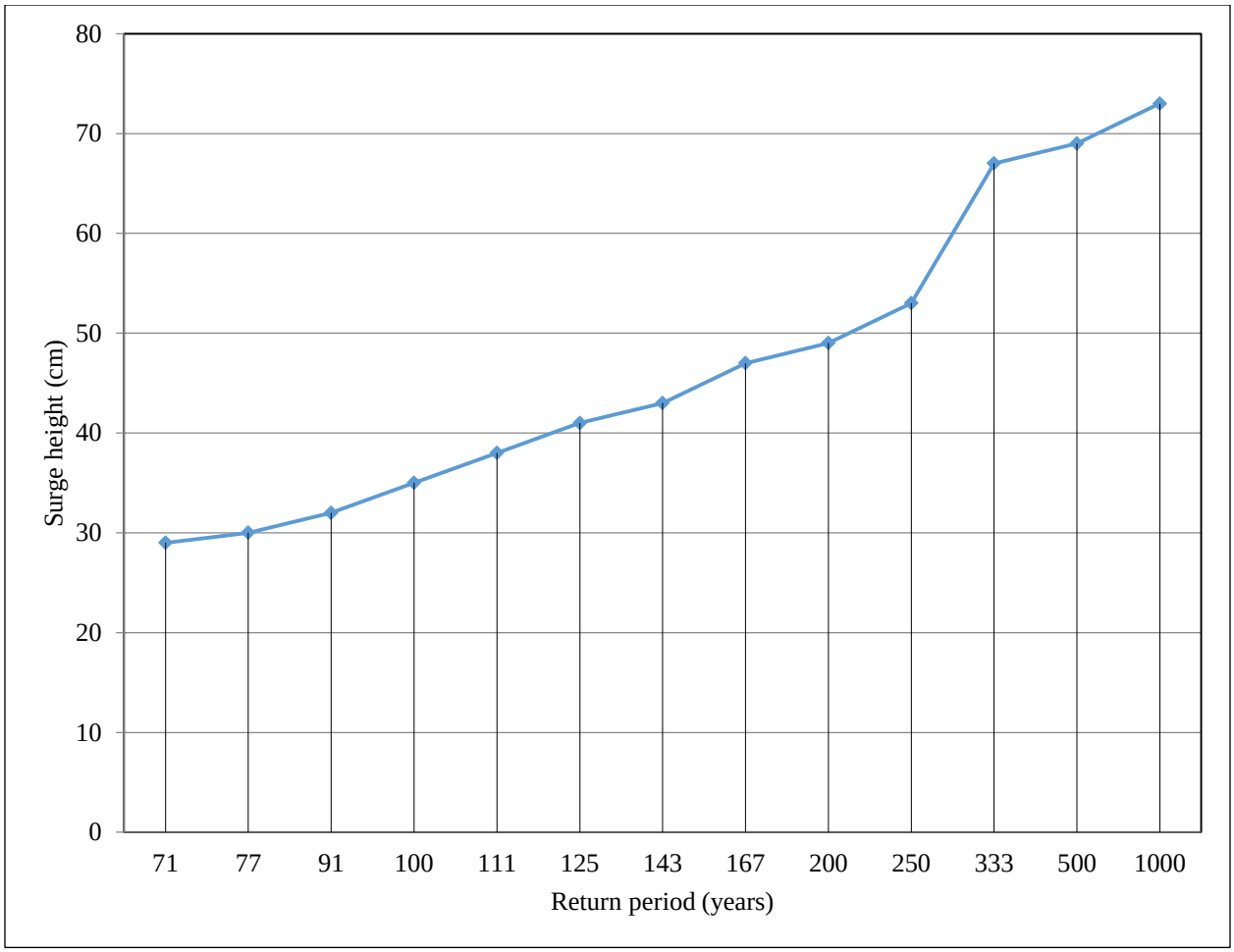

Figure 3.9 Return period of each randomly selected peak heights

The calculation for the return period estimation is based on the total water level rather than on the surge level. Hence, for each simulated storm event (i.e., exceedance cluster), the study computes the maximum total water level during that event. To do so, the study computes the total water for each hour of the surge event by adding the following components:

- the Mean Sea Level;
- the Tidal Fluctuation: this (deterministic) component is computed using the tidal components estimated with the harmonic analysis;
- the Surge level.

The maximum total water level of each simulated cluster is then computed by taking the maximum total water level among all the hourly total water levels computed for that cluster. Once sorted in descending order, each simulated water level is associated with a return period of $\frac{N}{i}$ years (i = number of storm peaks over the designated threshold).

3.2.1 Typhoon Maemi return period results at the tidal gauge station in Busan, South Korea

When simulating N=50,000 years using a Weibull distribution, the estimated 95% bootstrap confidence interval of Typhoon Maemi with a TWL of 211 cm is 389.11 years (95% CI: 342.47 - 476.2 years). Similarly, the 95% bootstrap confidence interval for the 100 -year return level is [198.29, 199.58] cm. The point estimate is 198.85 cm.

Note that only extreme events are being considered, that is, typhoons whose surge is above the threshold. If MSL, tide, and SLR are added to the threshold, the threshold is around 153 cm. Anything below is not considered "extreme" by this study. Return periods for total water levels below 100 cm were not calculated because only return periods for extreme events were considered.

As seen in Figure 3.10, 50,000 years are used to generate the return level, but only return periods up to 500 years plotted. As shown in the plots, the width of the confidence intervals is still narrow, but it is eventually increasing in time.

Figure 3.10 Return period curve for total water levels at the tidal gauge station, the city of Busan, South Korea

3.3 Conclusions

In Chapters 2 and 3, this study used tidal gauge station data in the city of Busan, South Korea, to estimate total water level return periods. Clustered separated peak-over-threshold simulation (CSPS) was used to estimate the return periods. Typhoon Maemi recorded the highest total water level (211 cm) at a tidal gauge station in South Korea's history. The return period of the total water level of Typhoon Maemi was estimated as 389 years (95% CI: 342.47-476.2 years). The results are significant because there are no existing studies estimating return periods using tidal gauge data in South Korea. The results could provide important guidelines for the insurance industry, the construction industry, and regional governments in order to obtain a better understanding of the damage caused by natural disasters.

Chapter 4

4 Fragility curves from storm-induced wind

4.1 Introduction

The severity of windstorms and the damage they cause has rapidly increased. Hurricane Katrina in 2005 devastated and paralyzed the Gulf Coast, causing 108 billion USD worth of damage, making it the costliest recorded natural disaster in U.S. history. Hurricane Ike in 2008 and Sandy in 2012 caused about 29.5 billion USD and 71.4 billion USD, respectively (Blake et al., 2007) [2]. Super typhoon Yolanda (also known as Typhoon Haiyan) in 2013 caused similar damage in South Asia. After it made landfall, Southeast Asia and Philippines recorded damages of about 2.86 billion USD. However, it should be noted that in all these storms, not all of the damage is caused by the wind. Actually in most cases, the majority of the damage is caused by the storm surge.

Governments, insurance companies, and the construction industry have developed risk management models to cope with the risk from natural disasters at regional and national levels. Such risk models for windstorm-induced loss are important to insurance companies because these models can estimate the value of potential loss from natural disasters, such as probable maximum loss, limit of liability, and exceedance of loss reinsurance. Those values are important when deciding how to designate catastrophe zones and how to share and allocate risk (Cummins et al., 1999 [9]; Kim et al., 2016 [24]).

Risk assessment models are comprised of four components: hazard, vulnerability, exposure, and damage. Hazard values represent the intensity of windstorm in terms of wind

speed. The hazard curve provides the occurrence probability of different vales of the wind speed. Exposure provides information about the property possibly affected. Vulnerability provides the damage potential as a function of the intensity of the hazard, and damage is the total loss, including liability coverage.

Fragility curves are the best way to manage financial risk from windstorms. Fragility curves, using vulnerability functions, express the correlation between damage ratios, wind speed, and damage states (Khanduri and Morrow, 2003) [19]. Since vulnerability functions are based on damage values, the quality and accuracy of damage data is critical.

4.1.1 Outline and objective

The main purpose of this research is to develop empirical fragility curves of damaged properties. An insurance company has provided loss data from Typhoon Maemi. The quality and quantity of the data allows for the construction of reliable fragility curves. Each curve is generated based on damage ratios, wind speed, and damage level (Damage I – Damage IV) using the following:

- Data collection

  Loss data on Typhoon Maemi is provided by an insurance company in South Korea and limited to the City of Busan. Wind speed (defined as 10 min sustained maximum) is provided by the Korea Meteorological Administration.

- Data sorting

  Because the original data from the insurance company is in Korean, all data is translated into English. Critical information, such as address, damage loss, value of insured property, etc., has been extracted from the dataset.

- Geographic Information System (GIS)

This research uses GIS to express the longitude and latitude of each property and construct a map of the damage state and wind speed at each property.

- Fragility curves

Empirical fragility curves as proposed by Shinozuka, et al., 2000 [49] are developed from the provided (details in the next chapter), and the range of each damage state is determined by the distribution of the damage ratio. Finally, the exceedance probability of wind speed for each damage state is calculated.

4.2 Literature review

4.2.1 Studies regarding windstorm-induced damage in South Korea

Previous research on damages caused by typhoons in South Korea mainly focused on the characteristics of typhoons. Ku et al., 2008 [29] developed a prediction model using storm-induced rainfall data, and Park et al., 2011 [40] and Park et al., 2012 [41] utilized typhoon wind field and strong wind speed to predict the area and extent of damages. Shin et al., 2013. [48] estimated typhoon paths associated with property damage based on the forward movement speed and direction of typhoons.

Kim et al., 2015 [25] identified the hurricane indicators, such as hurricane wind speed, building age, and building area, associated with damage. This study used Hurricane Ike, 2012 claim payout data from the Texas Windstorm Insurance Association. Based on the data, Kim et al., 2015 [25] established metrics to investigate the critical factors in the damage caused. Damage ratio was used as the dependent variable, and wind speed, building age, building area were selected for independent variables. Through Spearman correlation, they found maximum

wind speed and building age have a positive relationship with damage ratio. Building area, on the other hand, has a negative relationship with damage ratio. Although the findings from this study identified the most critical indicators of damage caused by hurricanes, they only used commercial building damage data and wind speed, among various hurricane characteristics. Also, only one case (Hurricane Ike) was used for the study. In 2016, Kim et al., 2016 [24] conducted similar research using the same data, but applied more independent variables in their metrics, such as appraised building value, surge zone, side of hurricane track, and distance from shore. They found that the appraised building value and the right side of the hurricane track have a positive relationship with damage ratio; surge zones and distance from shore line have a negative relationship with the damage ratio. However, the study, like Kim et al, 2015 [25], focused solely on commercial property damage and Hurricane Ike.

Another approach to predict and estimate damage caused by typhoons was investigated by Kim et al., 2018 [23]. They used the tracks of typhoons in the Korean Peninsula to estimate property vulnerability. Critical typhoons were categorized into Type 1 and Type 2, according to their path and location of landfall. Typhoons that made landfall on the west coast on the Korean Peninsula were categorized as Type 1, while typhoons that made landfall on the southern coast were grouped as Type 2. Rainfall, radius, maximum wind speed, forward movement speed, and number of land areas with decreased slope due to development were deployed as independent variables, using loss ratio as the dependent variable. The variables were applied to Type 1 and Type 2 to compare which variable was the most critical factor in damages cause by typhoons. Kim et al., 2018 [23] concluded that the extent of damage of Type 1 typhoons was correlated, in descending order, with maximum wind speed, number of land areas with decreased slope, and rainfall. However, the critical damage indicators of Type 2 typhoons were correlated, in

descending order, with forward movement speed, rainfall, maximum wind speed, and radius. Although Kim et al., 2018 [23] found that landfall location could affect the extent of damages caused by typhoons, their study did not measure the relationship between each damaged property and indicator.

4.2.2 Fragility curves

Insurance companies recognize the importance of catastrophe loss modeling in quantifying and allocating risk within their portfolios. Usually, vulnerability models include information related to damage functions, such as building history, floors, occupancy classes, locations, building codes, etc.

Watson and Johnson, 2004 [55] compared and evaluated the combination of windstorm damages, vulnerability, and risk. Historically, wind damage data as well as the inventory and structural characteristics of properties have been critical factors in developing vulnerability functions or fragility curves. These fragility models are empirical and fit the regression curve to the historical damage data as a function of damage levels. These models reflect the vulnerability of specific types of structures and are widely used in the insurance industry.

The exceedance probability for each damage ratio is expressed as a function of wind speed, and the different curves indicate different damage levels. Fragility curves show that the probability of exceeding a particular damage state is 0 % up to a certain critical wind speed (v_{critical}), at which point a certain threshold determined by the damage ratio has been reached. Above that threshold, the probability of exceeding a particular damage state increases until reaching 100% exceedance probability (meaning the property has totally collapsed) at a certain maximum wind speed (v_{maximum}) (Lopeman and Deodatis, 2015) [33], [34], [35], [36]. From the

literature review, the damage ratio (the cost of the damage as a percentage of appraised value of the property) as a function of wind speed is used as the main variable. Using wind speed and direction, the curves indicate storm intensity by specific type of construction and geographical characteristics of the region. Ultimately, fragility curves provide useful tools for predicting the possibility of reaching or exceeding different damage states (Hwang, 2013) [17].

4.2.3 Various characteristics of typhoons related to the extent of damage (Possible causes of damage caused by typhoons)

The severity of typhoons can be measured by maximum wind speed, peak wind duration, wind radius, forward motion speed, and motion direction (Dunion et al., 2003 [12]; Vickery et al., 2006 [53]; Burton, 2010 [5]). There are several studies regarding the extent of damage to properties based on various characteristics. Among those characteristics, some researchers have argued that maximum wind speed and maximum wind radius are the most critical in estimating the damages from typhoons (Watson and Johnson 2004 [55]; Vickery et al., 2006 [53]; Burton, 2010 [5]). Others (Watson and Johnson, 2004 [55]; Rego and Li, 2009 [46]) have posited that forward motion speed is the most important variable since that variable can affect flood volume. Slow moving typhoons can cause more property damage due to heavy flooding than fast moving typhoons (Rego and Li, 2009) [46]. According to Choi and Fisher, 2003 [6], rainfall is another key variable in determining the extent of damage, while built environment, frequency, and magnitude have all been considered main factors as well for windstorm damage assessment (Khanduri and Morrow, 2003; Kim et al., 2016 [24]; Huang et al., 2001 [16]).

The National Oceanic and Atmospheric Administration (NOAA) coordinates a hurricane observation system which provides the Hurricane Research Division (HRD) real time wind

analysis. Wind analysis contains collected wind data from observation stations based on period of 4-6 hours during hurricanes. The collected wind data includes maximum wind speeds, forward movement, hurricane direction, as well as wind speed direction and duration (Dunion et al., 2003 [12]; Powell and Houston, 1998 [42]; Powell et at., 1998 [43]; Powell et al., 2010 [44]). This wind analysis can be applied to gather information on hurricane intensity and can be combined with data from GIS.

Built vulnerability is also regarded as a critical factor in the damage caused by typhoons (Khanduri and Morrow, 2003; Kim et al, 2016 [24]; De Silva et al., 2008 [11]). Khanduri and Morrow, 2003 suggested high-rise buildings could incur more damage than other structures. Also, flood control facilities (Brody et al, 2008) and the decreased slope of mountains due to development are associated with the extent of damages caused by typhoons (Dai et al. 2002 [10]; Zhai et al. 2007 [58]; Cui et al. 2009 [8]; Ayalew and Yamagishi 2005 [1]).

Another important factor in assessing damage is property position in relation to the location of the hurricane. Usually, the right side of hurricanes cause more damage than the left side in the northern hemisphere (Keim et al., 2007 [21]; Noel et al., 1995 [39]). The reason for this difference is that wind intensity and direction in combination with forward movement and counterclockwise rotation causes higher wind speed and higher storm surges on the right side, thus causing more damage.

4.2.4 Line of business

Appropriate fragility curves are determined by type of line of business. Line of business is a well-known term in the insurance industry and is used to group properties or buildings according to with similar accounting and property policies. Within the field of civil engineering, the term groups buildings according to structure, use, and financial features. Three common lines of business within the field are industrial, commercial, and residential (Kim et al., 2017 [26]). Most of the literature focuses on residential due to the abundance of loss data compared to other lines of business. This study combines all three common lines of business in order to produce a robust dataset.

4.2.5 Construction typology and contents

Empirical fragility curves can be established based on the historical damage recorded for specified regions. They do not take into account construction types, materials, loads, etc. For instance, masonry walls and structure frame walls would have different damage levels at the same wind speeds. In order to account for those variables, empirical fragility curves can be combined with engineering-based ones. For example, construction typologies such as materials, structure types, load types, etc., can be categorized into several lines of business within a region. If each line of business has its own damage ratio, that damage ratio could have its own curve representing the vulnerability of each construction typology. Comparing the vulnerability models with similar construction typology may prove significant. (Lopeman, 2015) [37]. However, an analysis of each construction typology would require more data and is therefore beyond the scope of this study as such data was not avoidable.

Interior contents could also be considered when creating fragility curves. When insured properties are damaged by high wind speed, not only outer structures but also interior contents may be damaged. However, while relevant literature does consider the interior contents as part of the total damage cost, some researchers do not consider it as damage losses since it is very difficult to gauge and analyze. Moreover, at low wind speeds, damages to interior contents are not regarded as damage because interior contents are not exposed to storm winds, only to the possibility of the flooding. Therefore, they are not included in the calculation of the damage ratio in this study.

4.3 Windstorm-induced damage

4.3.1 Damage data from an insurance company in South Korea

Investigating damage losses by wind speed is crucial for the insurance industry in deciding risk factors and insurance premiums. Primary insurance and reinsurance companies should prepare a plan for unexpected damages from extreme natural disasters, such as Typhoon Maemi in 2003, Hurricane Katrina in 2004, and the Thailand flood in 2011. Insurance companies usually designate catastrophe zones and determine the risk factors for each zone. This work can help companies reduce their losses and distribute risk when extreme events happen.

Vulnerability functions or fragility curves are important tools in designating these catastrophe zones. However, lack of data on damage loss is an obstacle in deploying these methods. The most accurate and reliable data are the claim payouts and values of insured properties. Those values can be used for calculating damage ratios and susceptibility to extreme events. Claim payouts are determined by engineers and adjusters from the insurance companies after the extreme event.

Detailed information regarding the insured properties, such as materials, heights, number of floors, age, and location, is critical to evaluate the damage losses. However, it is very difficult to collect property information for developing countries since they usually don't have a systematic structure to collect the information. Also, inaccurate and insufficient property information hinders the government's or insurance companies' ability to plan for extreme disasters. Thus, developing fragility curves is used in this study to supplement the limited information on damaged properties.

As shown in Figure 4.1, the orange border indicates the boundary of Busan, and the yellow line shows the track of Typhoon Maemi. The damage ratios for each insured property are showed in Figure 4.2 with differently sized circles indicating different damage extents.

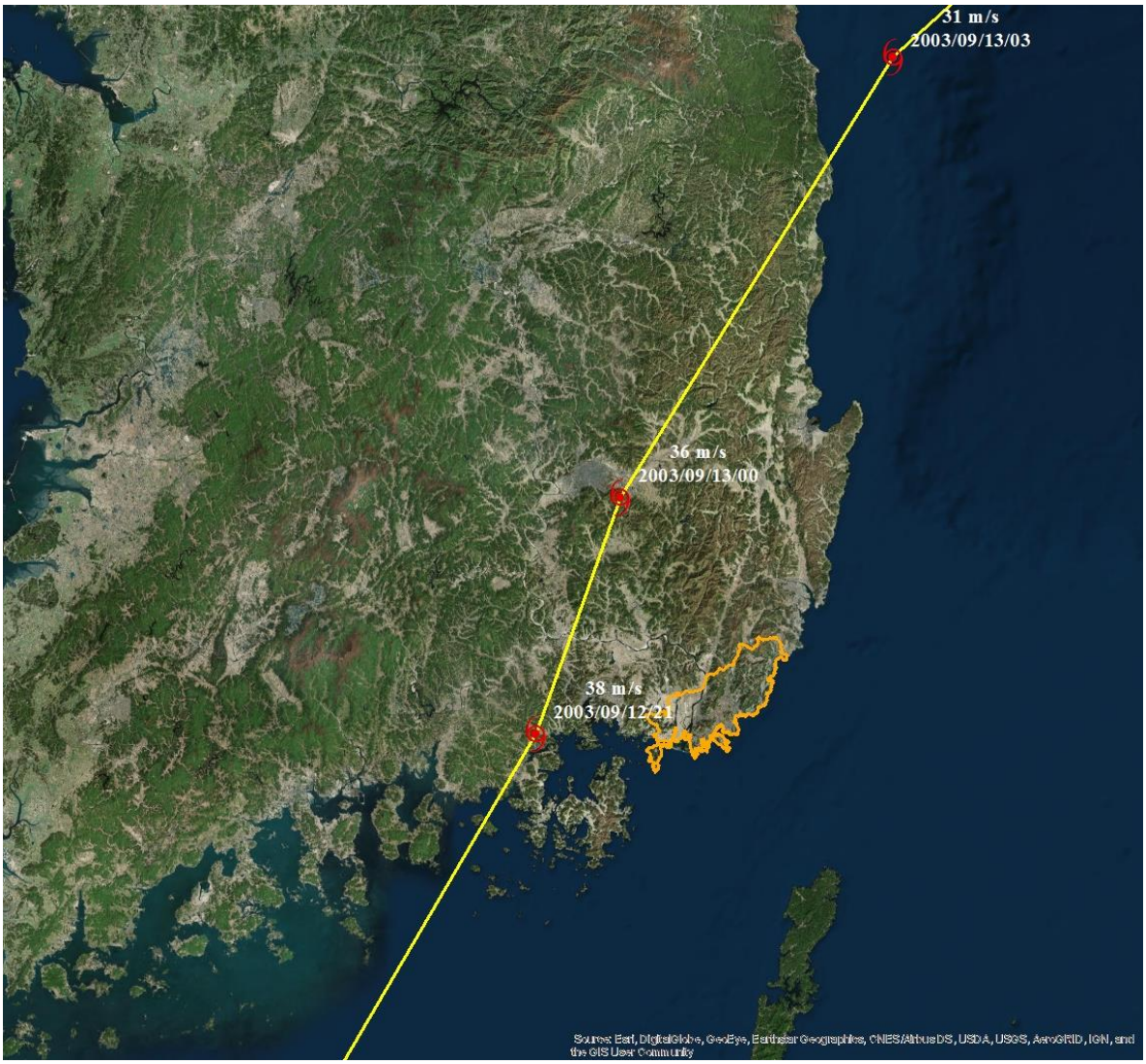

Figure 4.1 Track of Typhoon Maemi

4.3.2 Damage ratio

This study develops a statistical model to construct fragility curves for wind-induced damage. Each curve provides the exceedance probability of the each damage state at a particular wind speed. To establish the curves, property damage information caused by Typhoon Maemi is utilized. The main purpose of the curves is to provide the probability of certain damage levels as

a function of wind speed. The damage ratio is calculated using the claim payout divided by the property's appraised value:

$$\text{Damage Ratio} = \frac{\text{claim payout}}{\text{Appraised value of property}}$$

4.3.3 Geographical information system (GIS) based spatial damage distribution

This study combines the damage ratio (i.e., claim payouts divided by appraised values) provided by the insurance company for each insured property with wind speed at each property. Wind speed at the location of each property is calculated by ArcGIS using the inverse distance weighted (IDW) tool, producing a dataset that includes maximum sustained wind speed and direction, as well as gridded data, image data, and GIS shape files—all essential information when constructing a wind map in ArcGIS. Figure 4.3 indicates the wind map created by ArcGIS. In order to produce the wind map of the damage, the wind speed at each damage property is expanded and interpolated using the maximum wind speeds (10 min. sustained) at each wind observation station. Different wind speeds in Figure 4.3 are indicated by color (from green to red) and contour. Each property's location (based on the longitude and latitude provided by the insurance company) and the extent of damage is marked on the map.

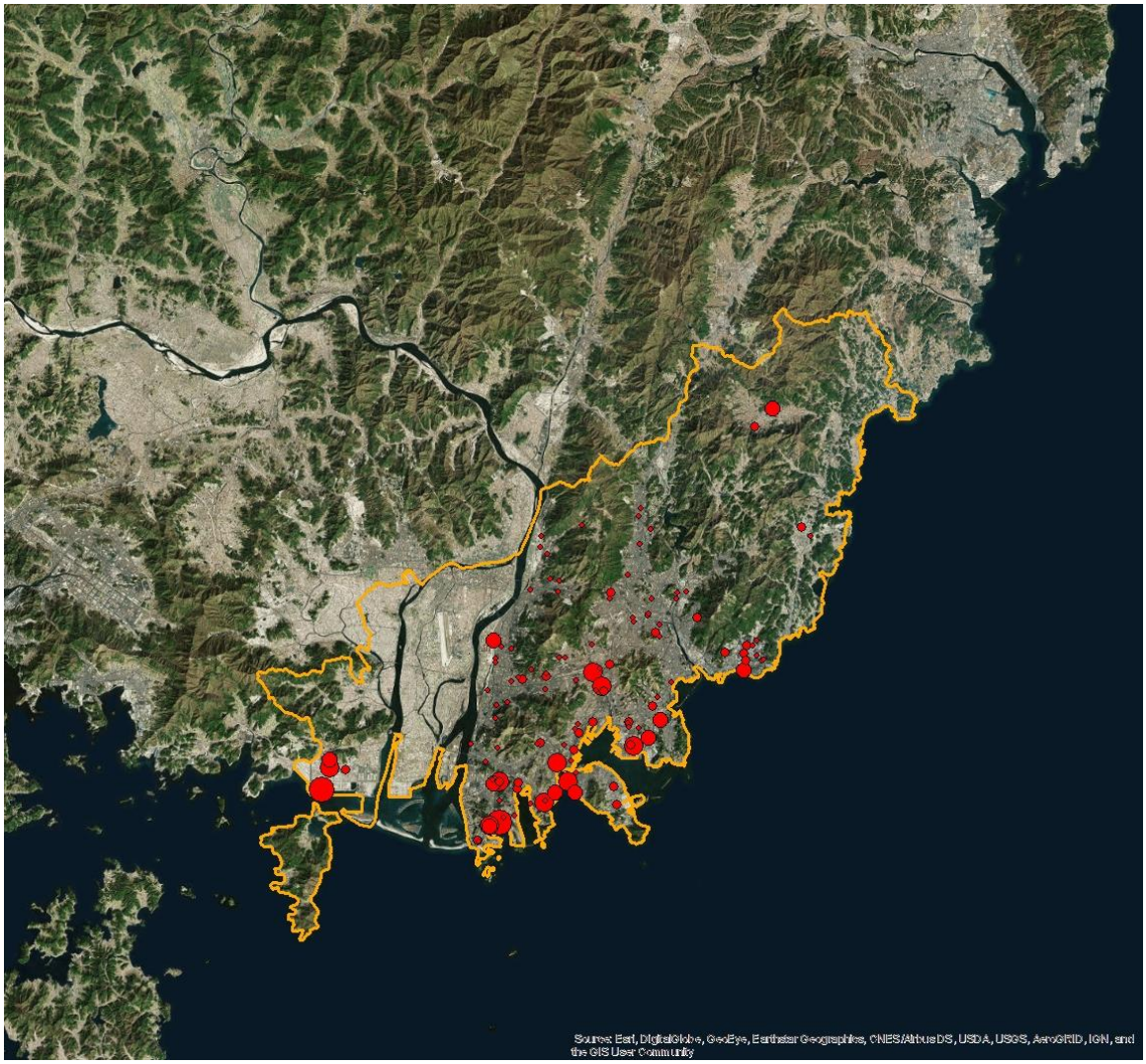

Figure 4.2 Spatial distribution of damages

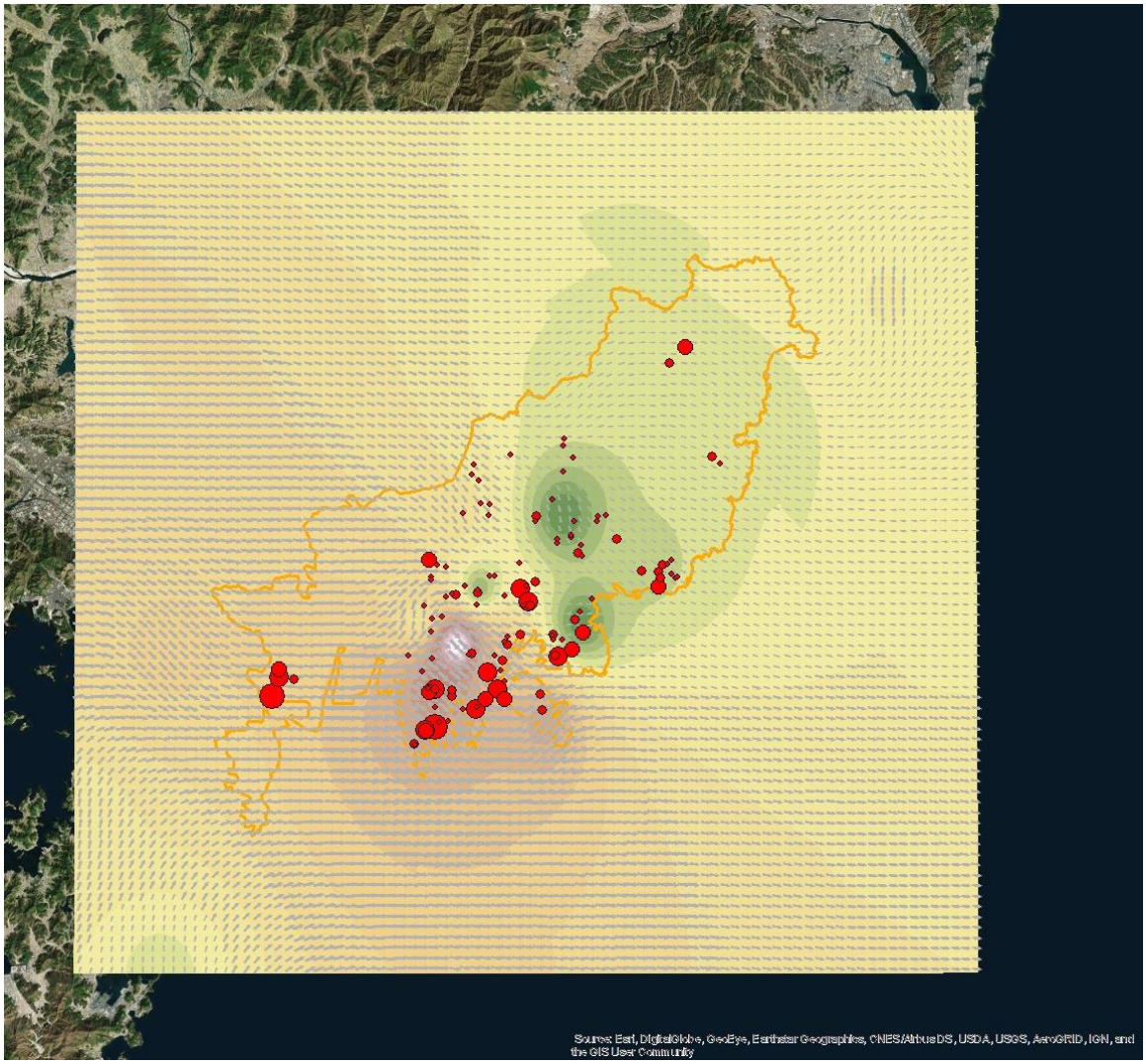

Figure 4.3 Wind speed map with spatial distribution of damages

4.4 Empirical fragility curves

This study uses damage data from the insurance company and is limited to the City of Busan, South Korea, during Typhoon Maemi. Based on that data, fragility curves are established to describe the vulnerability of each property.

4.4.1 Minimum square error (MSE) method

This study is interested in exploring the relationship between damage to buildings caused by typhoons and the corresponding wind speed, in particular, how the probability of exceeding a certain damage threshold varies with the maximum wind speed. The study achieves this goal by estimating the fragility curves against the wind speed from a dataset of more than 120 observations of damage ratios and related maximum wind speeds. Following the literature (for example, Shinozuka et al., 2000 [49]), this research models the analytical form of fragility curves by using the cumulative distribution function of the lognormal distribution and then estimates four distinct fragility curves with damage thresholds ranging from low damage to more severe lvels. Finally, this study uses two different criteria for the estimation: the first is based on the minimum square error (MSE), while the other on the likelihood function. These two approaches are described below.

Assume that this study observes n damage ratios $r_1, r_2, \dots, r_n$ and related (maximum) wind speeds $w_1, w_2, \dots, w_n$. These ratios are separated in different groups based on their wind speeds. For each group, the fraction of ratios that exceed a given threshold is computed and then used as an observation of the fragility curve of interest at that point.

All damage ratios are assigned according to their wind speed in the wind interval. The wind intervals are 10 – 15 m/s, 15 – 20 m/s, 21 – 25 m/s, 25 – 30 m/s, and above 30 m/s. Each wind speed interval has four different damage levels: 0 – 0.003 (damage level I), 0.003 – 0.006 (damage level II), 0.006 – 0.009 (damage level III), and above 0.009 (damage level IV). For each damage state within each wind speed interval, a damage ratio is counted. The percentage of the number of damage ratios equal to or larger than each damage state in each wind interval is then calculated. The below Table 4.1 and Figure 4.4 shows the results.

Wind speed interval (m/s)	10-15 m/s	15-20 m/s	20-25 m/s	25-30 m/s	Above 30 m/s
Total number of damages	8	50	30	35	3
Number of damage ratios equal to or larger than damage level I	8	50	30	35	3
Number of damage ratios equal to or larger than damage level II	6	34	19	26	3
Number of damage ratios equal to or larger than damage level III	3	18	12	19	2
Number of damage ratios equal to or larger than damage level IV	3	13	10	14	2
Percentage of damage ratios at damage level I (%)	100	100	100	100	100
Percentage of damage ratios at damage level II (%)	75	68	63	74	100
Percentage of damage ratios at damage level III (%)	38	36	40	54	67
Percentage of damage ratios at damage level IV (%)	38	26	33	40	67

Table 4.1 Number of damage ratios and percentage of damage ratios in each different wind speed interval

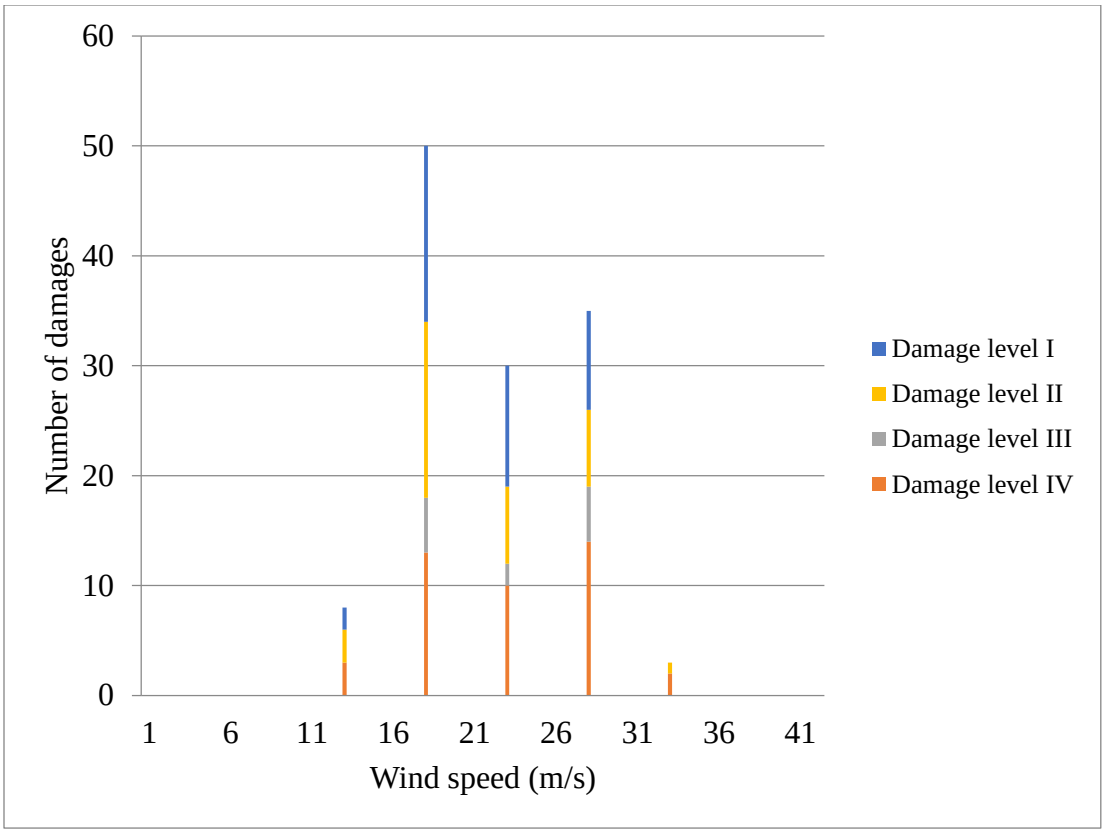

Figure 4.4 Number of damage ratios for each wind speed interval, damage state indicated by color

For example, the study assigns all damage ratios whose related wind speed is between 15 and 20 meters per second (m/s) to damage state II. By dividing the number of damage ratios equal to or larger than damage state II in the wind speed interval 15 – 20 m/s by the total number of damage ratios within the wind speed interval 15 – 20 m/s, the study calculates that 0.68 of these ratios are above the threshold for damage state II, 68 % indicating the possibility of exceeding damage state II in the wind speed interval. The study can then use 68 % as the percentage of damage ratios when calculating the fragility curve for damage state II at an average maximum wind speed of 17.5 m/s (that is, the midpoint between 15 and 20 m/s). The same method applies to other three curves (damage state I, damage state III, and damage state IV) with their respective wind speed interval and probability of exceeding each damage state.

Figure 4.5 show the plotting of the exceedance probability of each damage stats in the each wind speed interval.

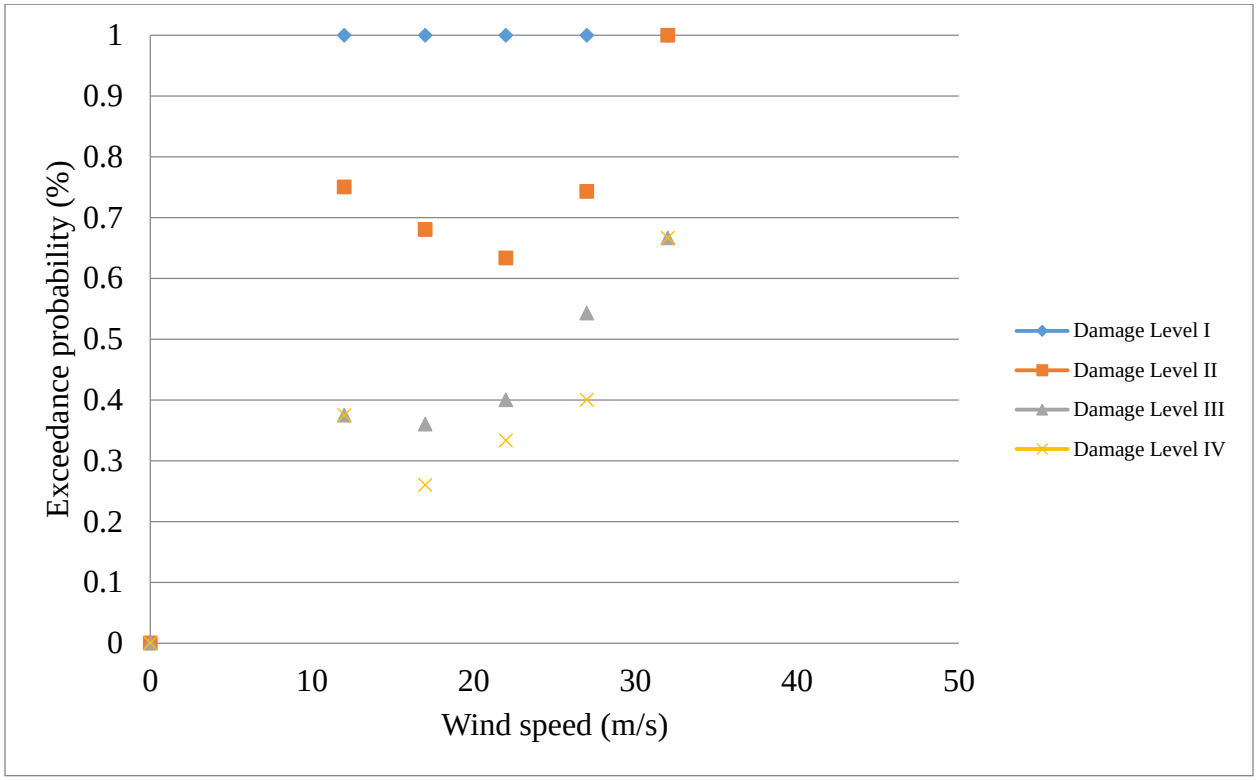

Figure 4.5 Exceedance probability of damage levels in the wind speed intervals

Following this approach, this study produces m realizations $p_1, \dots, p_m$ (with $m < n$) of the fragility curve for given wind speeds $\bar{w}_1, \bar{w}_2, \dots, \bar{w}_m$. These realizations are estimated by finding the lognormal curve that minimizes the MSE with respect to $p_1, \dots, p_m$. Because each lognormal curve is univocally identified by two parameters, μ and σ, this amounts to finding the values of those parameters that minimize the MSE:

$$\sum_{j=1}^{m} \left(F_{\mu,\sigma}(\bar{w}_j) - p_j\right)^2,$$

where $F_{\mu,\sigma}(\cdot)$ is the lognormal cumulative distribution function with parameters:

$$F_{\mu,\sigma}(\cdot) = \Phi\left[\frac{\ln(\cdot) - \ln(\mu)}{\sigma}\right]$$

The resulting algorithm is described in more detail in Table 4.2 below:

Algorithm: *Mean square error (MSE) estimation of fragility curves*
Input: Data: $(r_1, w_1), \dots, (r_n, w_n)$. Damage thresholds: $\bar{r}_{\text{damage state I}}$, $\bar{r}_{\text{damage state II}}$, $\bar{r}_{\text{damage state III}}$, $\bar{r}_{\text{damage state IV}}$. Representative wind speed per group: $\overline{w}_1, \overline{w}_2, \dots, \overline{w}_m$. (midpoint values) **Output:** $\mu_{\text{MSE}}, \sigma_{\text{MSE}}$.
For each damage threshold $\bar{r}$, do the following: 1. Assign each damage ratio x_i to a group based on its wind speed w_i. 2. For each group j, compute the fraction p_j of damage ratios that are above $\bar{r}$. 3. Find the values of μ_{MSE} and σ_{MSE} that minimize the mean squared error of the fragility curve $F_{\mu,\sigma}(\cdot)$ from $p_1, \dots, p_m$ and at the representative wind speed values $\overline{w}_1, \overline{w}_2, \dots, \overline{w}_m$. 4. Plot the fragility curve $F_{\mu_{\text{MSE}}, \sigma_{\text{MSE}}}(\cdot) = \Phi\left[\frac{\ln(\cdot) - \ln(\mu_{\text{MSE}})}{\sigma_{\text{MSE}}}\right]$, and return μ_{MSE} and σ_{MSE}.

Table 4.2 Mean square error (MSE) estimation of fragility curves

As shown in Figure 4.6 below, the fragility curves calculated using MSE method have limitations when it comes to explaining the correlation between curves. First, the curves overlap at low wind speeds. At the same time, the fragility curves show the exceedance probability of each damage state, except for damage state I, increasing continuously beyond the maximum

wind speed of 200 m/s as shown in Figure 4.7. This means that there would be a possibility that some properties may not collapse at the wind speed of 200 m/s.

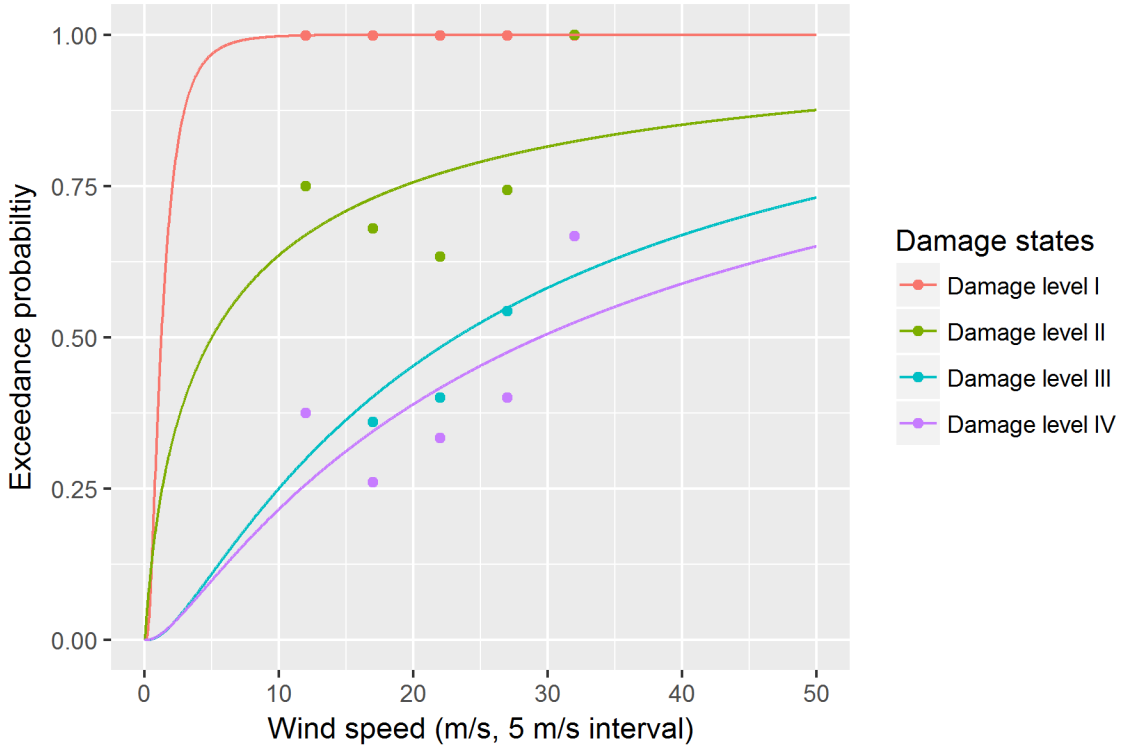

Figure 4.6 Fragility curves simulated by Minimum Square Error (MSE) at wind speed interval 0 – 50 m/s

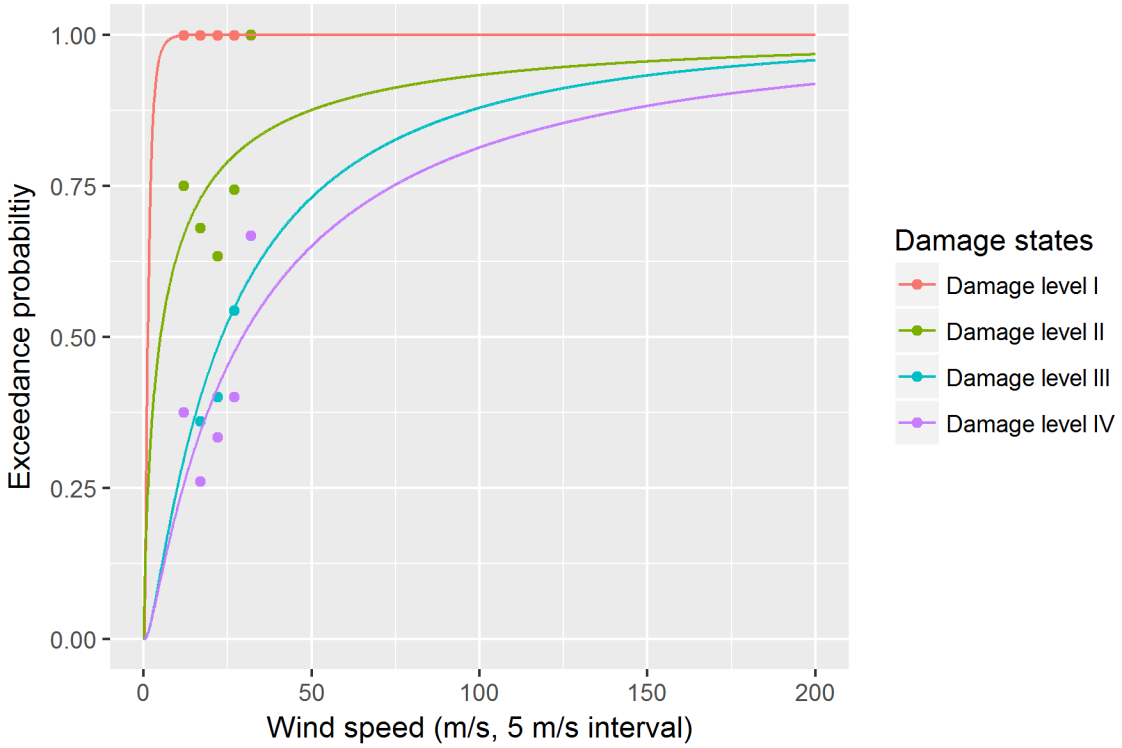

Figure 4.7 Fragility curves simulated by Minimum square error (MSE) at wind speed interval 0 – 200 m/s

4.4.2 Maximum likelihood estimation method (MLE)

Given a family of distributions indexed by some parameters, this study uses maximum likelihood estimation (MLE), a method for selecting parameter values under which data are most probable. The same damage state intervals used in MSE are used in MLE. A median of each damage state interval is selected as the threshold: 0.0015 for damage state I (0 – 0.003), 0.0045 for damage state II (0.003 – 0.006), 0075 for damage state III (0.006 – 0.009), and 0.01 damage state IV (above 0.009). For the purposes of this study, MLE involves choosing the parameters that define the lognormal shape of the fragility curve by maximizing the likelihood function given the data. To do that, a probability model for the data generation process is specified. In particular, assume that, for each damage, the study observes n boolean outcomes $x_1, \dots, x_n$ that indicate whether the damage is at least moderate or not. Hence,

$$x_i = \begin{cases} 1, & i\text{-th damage is at least moderate;} \\ 0, & \text{otherwise.} \end{cases}$$

Assume that this study also observes the maximum wind speed $(w_1, w_2, \dots, w_n)$ associated with each damage. Following Shinozuka et al., 2000 [49], the likelihood function can be built as:

$$L(\mu, \sigma | x_1, \dots, x_n;\ w_1, w_2, \dots, w_n) = \prod_{i=1}^{n} \left[F_{\mu,\sigma}(w_i)\right]^{x_i} \cdot \left[1 - F_{\mu,\sigma}(w_i)\right]^{1-x_i}$$

where $F_{\mu,\sigma}(\cdot)$ is the fragility curve under the lognormal assumption, as previously defined.

The study estimates each fragility curve by finding the values of the lognormal parameters μ_{MLE} and σ_{MLE}that maximize the likelihood function above for the given data $(x_1, w_1), \dots, (x_n, w_n)$. For example, this could be achieved by solving numerically the resulting first order conditions:

$$\frac{d}{d\mu} L(\mu, \sigma | x_1, \dots, x_n;\ w_1, w_2, \dots, w_n) = \frac{d}{d\sigma} L(\mu, \sigma | x_1, \dots, x_n;\ w_1, w_2, \dots, w_n) = 0.$$

In practice, the study directly optimizes the likelihood via the Limited-memory Broyden-Fletcher-Goldfarb-Shanno Algorithm, which is already implemented in R. As with the MSE approach, this study repeats this process for each level of damage so as to obtain four different fragility curves. Assuming this study observes the damage ratios $r_1, r_2, \dots, r_n$ together with the wind speeds $\mathrm{w}_1, \mathrm{w}_2, \dots, \mathrm{w}_n$, the algorithm summarized in Table 4.3 below implements this method:

Algorithm: *Maximum likelihood estimation (MLE) of fragility curves*
Input: Data: $(r_1, w_1), \ldots, (r_n, w_n)$. Damage thresholds: $\bar{r}_{\text{damage state I}}$, $\bar{r}_{\text{damage state II}}$, $\bar{r}_{\text{damage state III}}$, $\bar{r}_{\text{damage state IV}}$. **Output:** $\mu_{\text{MLE}}, \sigma_{\text{MLE}}$.
For each damage threshold $\bar{r}$, do the following: 1. Set $x_i = 1$ if $r_i > \bar{r}$ and $x_i = 0$ if $r_i \leq \bar{r}$, for $i = 1, \ldots, n$. 2. Define the likelihood function $L(\mu, \sigma \mid x_1, \ldots, x_n;\ w_1, w_2, \ldots, w_n)$. 3. Find the values of μ_{MLE} and σ_{MLE} that optimize the likelihood L via the L-BFGS-B under the constraint that $\sigma \in (0, \infty)$ and for the given data $(x_1, w_1), \ldots, (x_n, w_n)$. 4. Plot the fragility curve $F_{\mu_{\text{MLE}}, \sigma_{\text{MLE}}}(\cdot) = \Phi\left[\frac{\ln(\cdot) - \ln(\mu_{\text{MLE}})}{\sigma_{\text{MLE}}}\right]$, and return μ_{MLE} and σ_{MLE}.

Table 4.3 Maximum likelihood estimation (MLE) of fragility curves

Using the above algorithm by Shinozuka et al., 2000 [49], this study simulates and plots the fragility curves as shown in Figure 4.8 below.

We first test Shinozuka's MLE method for fragility curves on simulated data so as to assess the validity of our implementation. To generate the synthetic data, we first obtain simulated wind speeds that are equally-spaced at plausible intervals. Next, we generate the related damages from a linear model with normally distributed errors. Given its simplicity, the linear model provides a convenient way to generate damages that increase with wind speed. Finally, we rescale the simulated damages so as to obtain damage ratios that are consistent with the observed ones. The full procedure is described below:

- Compute $\widetilde{w}_i = 13\ m/s + 0.1 * i$ for $i = 0, \dots, 30$ so as to obtain simulated wind speeds that are equally-spaced at plausible intervals $[13\ m/s, 16\ m/s]$.
- Generate the related damages from the linear model $\widetilde{D}_i = \widetilde{w}_i + \varepsilon_i$, where ε_i is an error that follows a standard normal distribution. Hence, for each i, first sample a random draw ε_i from a standard normal distribution and then add it to $\widetilde{w}_i$ to compute the related damage $\widetilde{D}_i$.
- Finally, generate the simulated damage ratios by standardizing the simulated damages as follows:
    - Compute the standard deviation σ_{obs} of the observed damage ratios (that is, the damage ratios from the non-simulated data) and find the smallest damage ratio DR_{min}
    - Compute the standard deviation σ_{sim} of the simulated damages $\widetilde{D}_0, \dots, \widetilde{D}_{30}$ and find the smallest simulated damage $\widetilde{D}_{min}$
    - For each i, generate the i-th simulated damage ratio $\widetilde{DR}_i$ as:

$$\widetilde{DR}_i = \frac{\sigma_{obs}}{\sigma_{sim}} \cdot \left(\widetilde{D}_i - \widetilde{D}_{min}\right) + DR_{min}.$$

The normalization in the last step of the procedure above ensures that the simulated damage ratios preserve the observed variance of the real damage ratios, which vary from 0 to 1. After simulating the data as described above, we classify the simulated damage ratios into four damage levels based on the following thresholds: below .005 for Level I, between 0.005 and 0.015 for Level II, between 0.015 and 0.025 for Level III, and between 0.025 and 0.035 for Level IV. For each damage level, we then use Shinozuka's MLE approach to fit a lognormal fragility curve to the data $(\widetilde{w}_i, DR_i)_{i \geq 1}$. The resulting fragility curves capture the simulated data as desired and are plotted in Figure 4.8.

Finally, we run the same fit using the real wind speeds w_i and by setting more suitable thresholds for the damage ratios. The curves that we fit on the (fully) empirical data are plotted in Figures 4.9, 4.10, 4.11, 4.12, and 4.13.

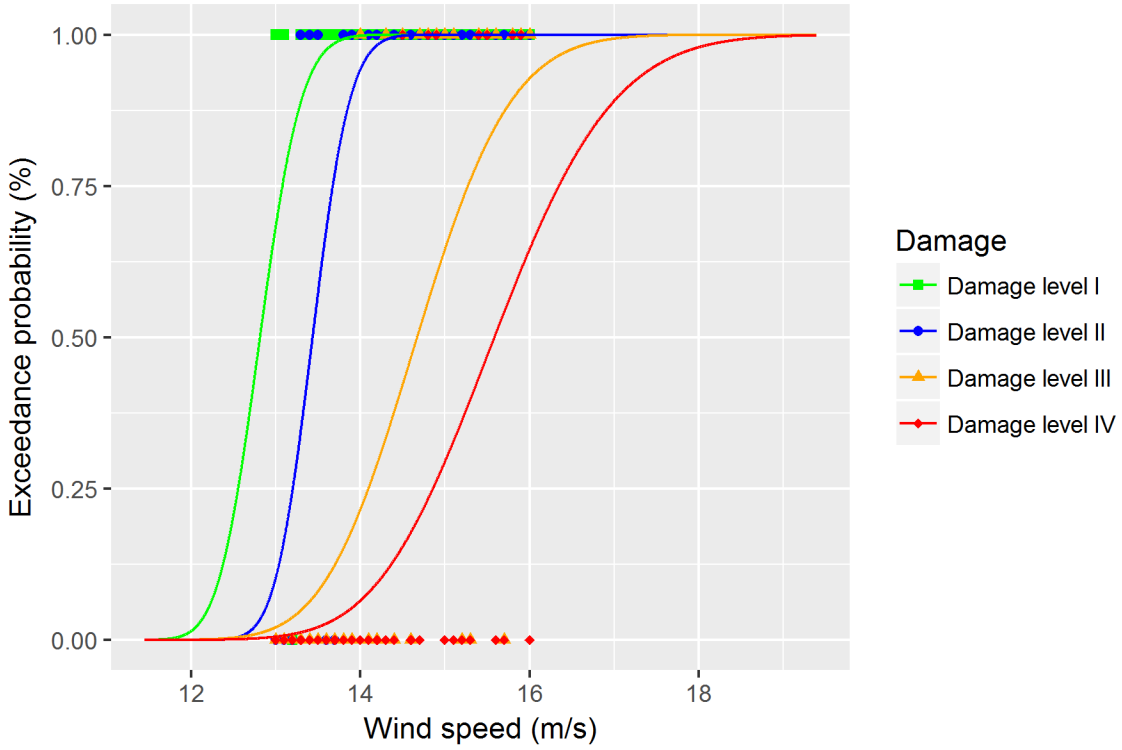

Figure 4.8 Fragility curves simulated according to Shinozuka et al., 2000 method [49]

4.5 Discussions

Fragility curves for Typhoon Maemi

Typhoon Maemi was the costliest typhoon in the history of South Korea, followed by Typhoons Kompas and Bolaven in terms of recorded damage and loss of life. The damage done by these three storms was 85.7 % of the total damage and 66 % of the total number of casualties from all typhoons in South Korea's history combined (Kim et al., 2017 [26]). Although these three typhoons impacted the entire country, the southern part of the Korean Peninsula was especially damaged. The city of Busan, in particular, suffered 35% of the country's total damage costs and 29% of the total amount of losses caused by Maemi, Kompas, and Bolaven.

Conducting vulnerability analysis of insured properties according to wind storm damage in the city of Busan, this study develops fragility curves for four different damage levels (Level I, Level II, Level III, and Level VI). As mentioned earlier, fragility curves provide the probaility of reaching or exceeding different damage levels as a function of wind speed. To create those curves, the following data was obtained:

- The value of the insured property
- The claim payout of the damaged property
- Wind speeds of Maemi, 2003 at all property locations
- Definition of damage levels based on wind speed

The fragility curves follow the likelihood function according to Shinzuka et al.. 2003 as written below:

$$L = \prod_{i=1}^{N} [F(a_i)]^{x_i} [1 - F(a_i)]^{1-x_i}$$

$F(\cdot)$ represents the fragility curve for a given level of damage; a_i is the wind speed value in which i represent the number of damages; x_i = 1 or 0 depending on whether or not the damage ratio exceeds or reaches the state of damage under a_i; and N = total number of damage ratios. Under the current lognormal assumption, $F(a)$ takes the following analytical form:

$$F(a) = \Phi\left[\frac{\ln\left(\frac{a}{c}\right)}{\zeta}\right]$$

in which a represents the wind speed; and $\Phi[\cdot]$ = standardized normal distribution function. The two parameters c and ζ in the above equation are computed as c_e and ζ_e satisfying the following equations to maximize ln L and hence L:

$$\frac{d\ln\ L}{dc} = \frac{d\ln L}{d\zeta} = 0$$

The curves show that the probability of each damage level rests at 0, but once each damage level meets the threshold of the damage ratio, the exceedance probability as a function of the wind speed moves asymptotically toward a probability of 100%. The figures below (Figure 4.9, 4.10, 4.11, and 4.12) show the exceedance probability of four different damage levels as a function of wind speed (wind speed limited to 200 m/s):

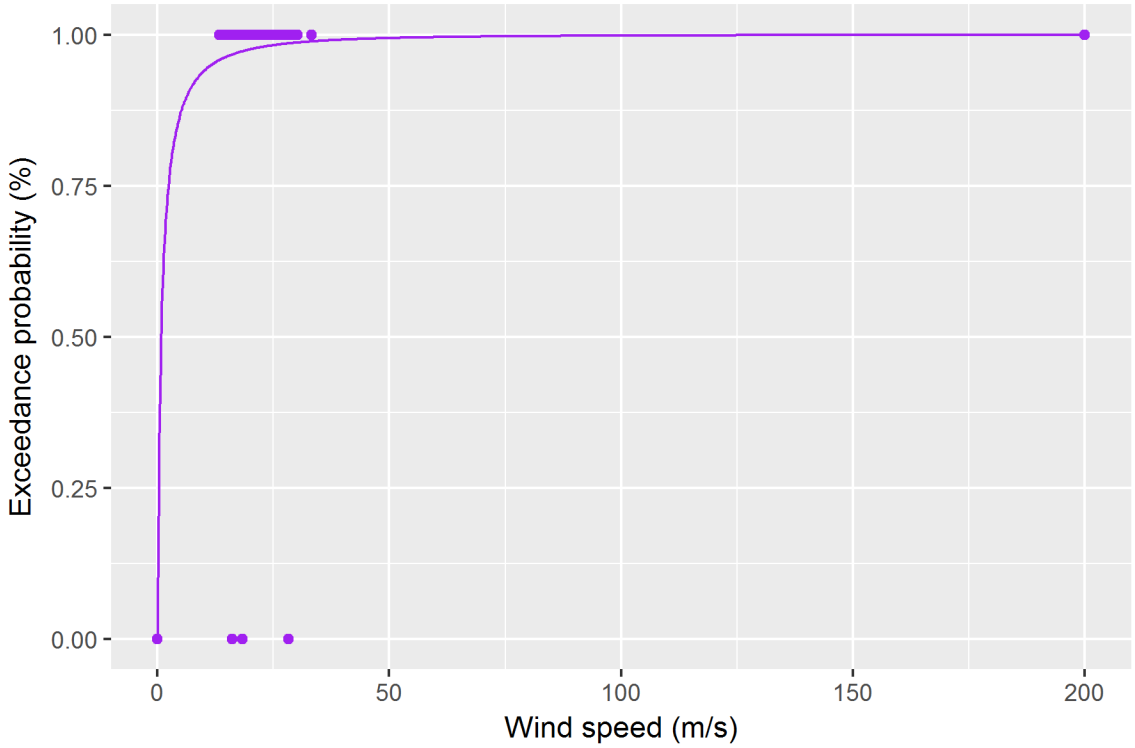

Figure 4.9 Fragility curve - Damage level I

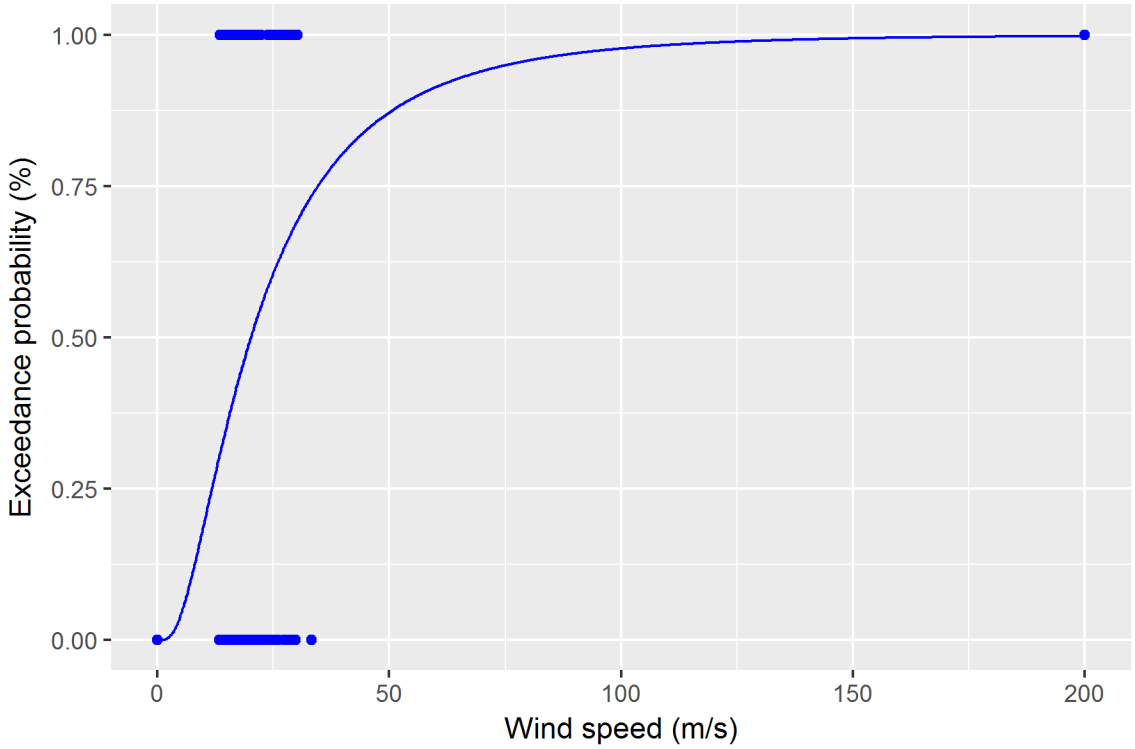

Figure 4.10 Fragility curve - Damage level II

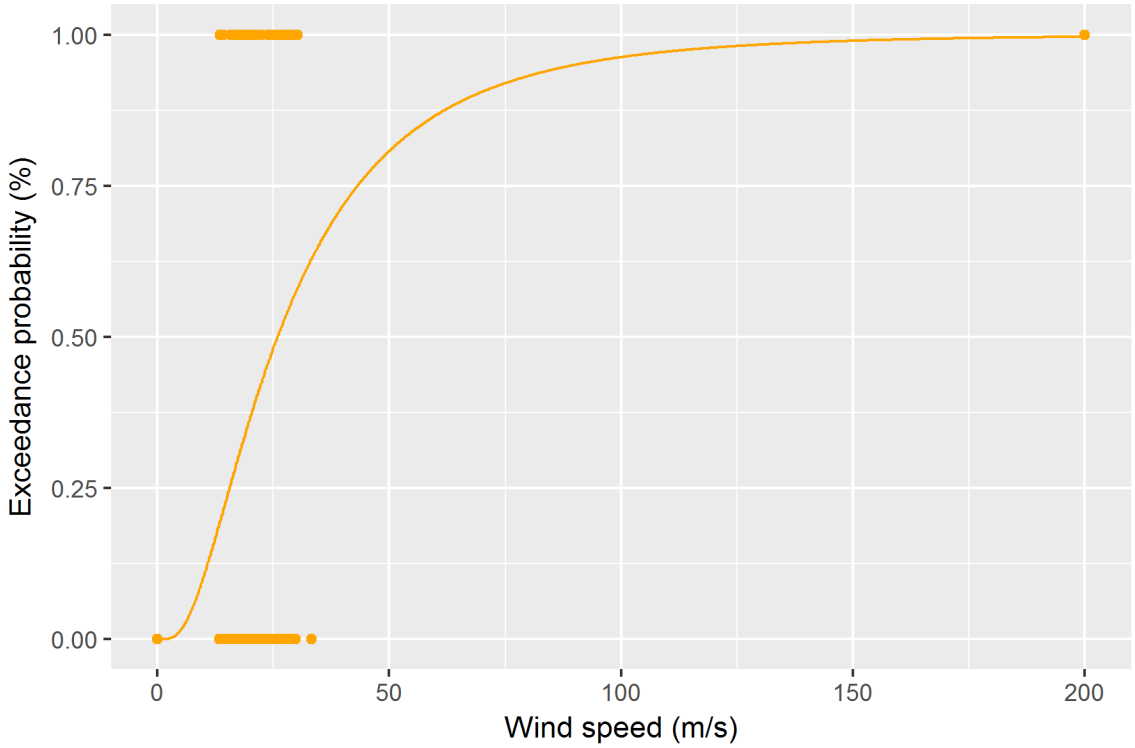

Figure 4.11 Fragility curve - Damage level III

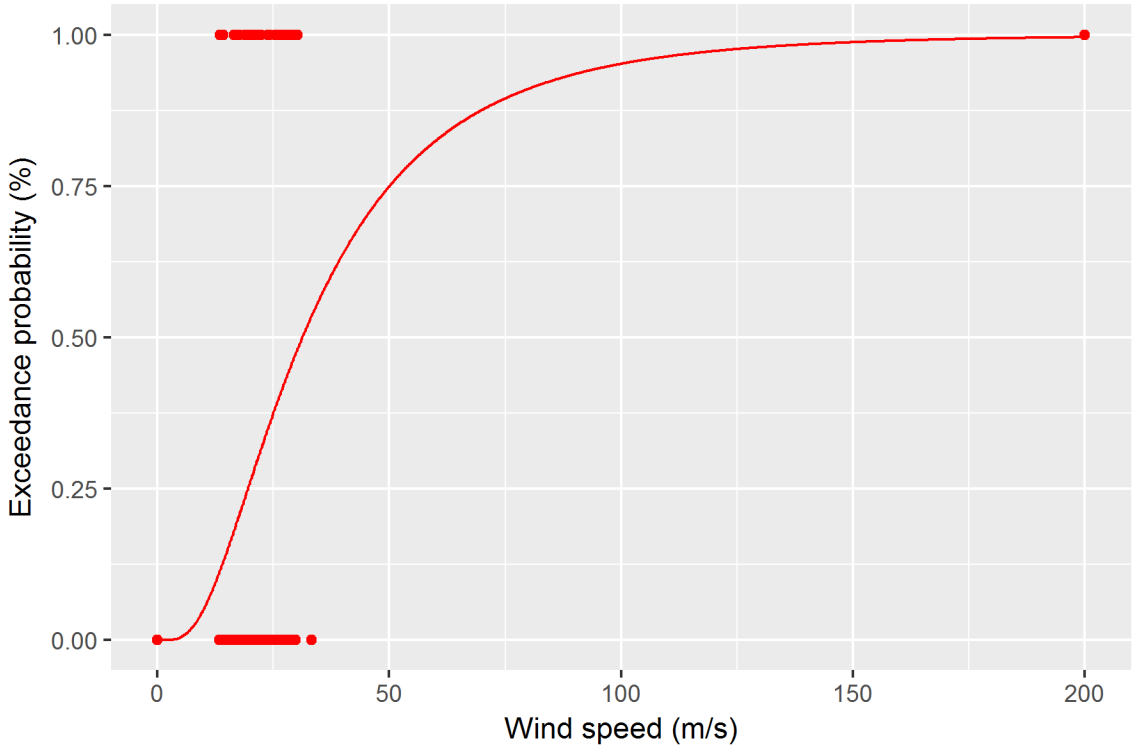

Figure 4.12 Fragility curve - Damage level IV

Figure 4.13 below shows that the fragility curves from damage state I to damage state IV. The curve of damage state I indicates that the probability of at least damage state I (as known as the most minor damage) at each wind speed a_i. Each dot plotted at the top of the horizontal axis means x_i = 1 for exceeding or reaching the state of damage, and each dot plotted at the bottom means x_i = 0 for not exceeding or reaching the state of damage under a_i.

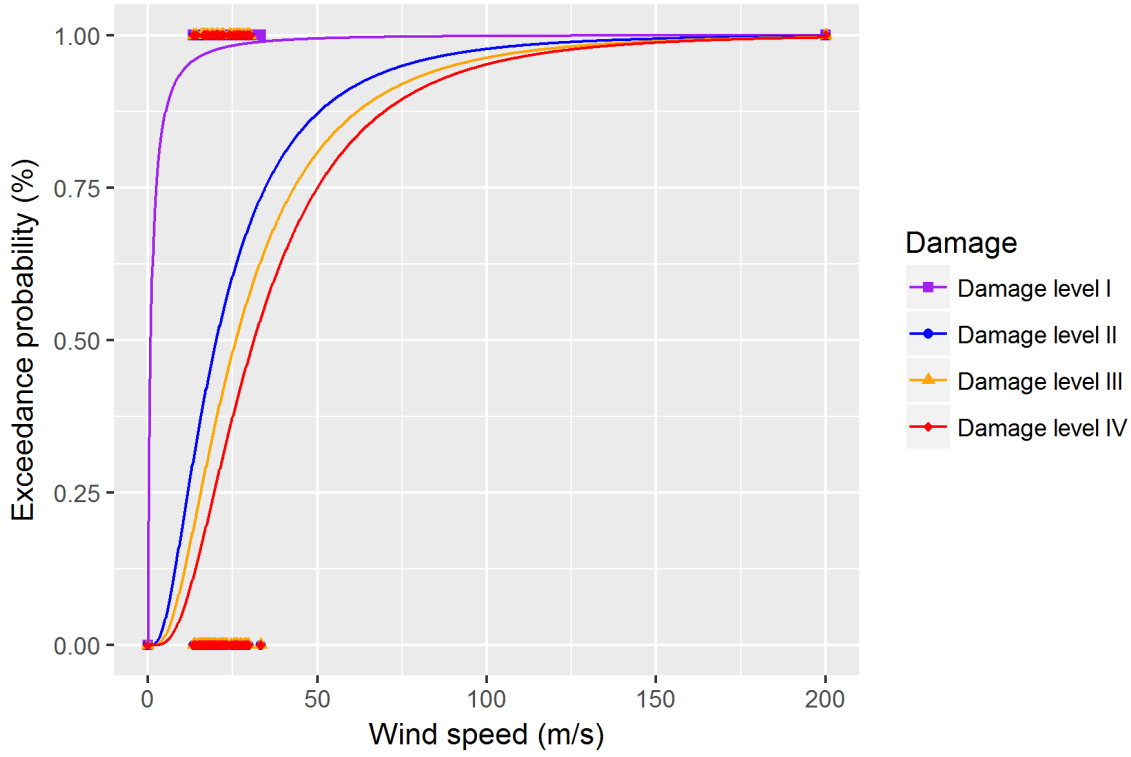

Figure 4.13 Fragility curves from damage level I to damage level IV

Chapter 5

5 Conclusions

To effectively plan for extreme events, governments and the insurance industry need accurate estimates not only for the return period of storms, but also the vulnerability of infrastructure to those storms. This dissertation focuses on the damages related to the surge height and wind speed of South Korea's most severe typhoon, Typhoon Maemi 2003. It develops two new approaches for quantifying natural disaster risk assessment and management. First, applying clustered separated peaks-over-threshold (CSPS) to hourly based tidal gauge data, the study estimates the return period of total water level for Typhoon Maemi. Second, it uses wind speed-induced damage data provided by an insurance company in South Korea to develop fragility curves. Ultimately, the exceedance probability of the fragility curves are shown to predict the damage levels of properties caused by wind speed.

5.1 Contributions

In South Korea, much research has been conducted by simulating storm surges based on the track of typhoons and other meteorological data such as wind profile, fluid dynamics, and atmospheric pressure. However, estimating the return period of storm surges has not been investigated, nor has statistical analysis been applied to tidal gauge data. The present study addresses that gap in the literature by using tidal gauge data collected at the city of Busan, South Korea, during Typhoon Maemi. The storm's tidal components are separated and their relationship with threshold selection, clustering, surge modeling, and seasonal separation

analyzed. Those features offer the opportunity to study other aspects of natural disaster risk management. The study inputs data on those features into the CSPS developed by Lopeman, 2015 [37]. But whereas Lopeman [37] used GPD distribution to estimate return period on Hurricane Sandy, this study uses a Weibull distribution (with N equal to 50,000 years simulation) which proves a better fit with the data on Maemi. Using the Monte Carlo simulation, return periods are extrapolated in order to overcome the limited data with regards to time period and deployed bootstrapping to generate confidence intervals on return period calculations. The study finds that the return period of total water level (211 cm) is 389.11 years with 95% confidence interval (342.47 – 476.2 years). The 100 year return period is 198.85 cm and the 95% bootstrapping confidence interval is 198.29 – 199.57 cm.

Previous research regarding windstorm-induces damage focused on variables related to particular storm features and possible damage indicators of typhoons. Much of this research used hypothetical data in constructing simulations. This study uses a new set of variables such as wind speed, threshold, and damage ratios in the form of actual loss data provided by an insurance company in South Korea. Applying an exceedance probability approach developed by Shinozuka et al. [49], the study constructs fragility curves for Typhoon Maemi.

5.2 Future research

By using historical data and statistical methodology, this two-part study suggests practical applications to cope with unexpected natural disasters. The first part focuses on predicting return periods of extreme events. While there is existing research on return period for hurricanes, no research on return periods currently exist for typhoons in South Korea. Building on the various methods and results from Lopeman and Deodatis [33], [34], [35], [36] and other, the present study proposes CSPS combined with probability distribution models as a possible new approach to estimating return periods. Because CSPS applies to only this one case study in a specified region, the method developed may not be applicable to different countries with different tidal components. If more cases around the world were collected, a more comprehensive statistical analysis would be possible. In particular, generating return periods of total water level would be possible where the hourly tidal gauge data is available. Meteorological data such as storm track, rainfall, direction, pressure, fluid dynamic, etc., should also be considered to better estimate accurate return periods.

While part one of the study does compare the fit of other distributions such as Logmormal, Frechet, Gamma, and GPD, it only uses the Weibull distribution. In the future, return periods models combining all possible distributions and comparing the results should be studied. Also, threshold selection should be studied more. Although this study does use an algorithm to determine threshold selection, more elaborated algorithms for target rate and threshold section should be considered in future research.

Part two of the study focuses on fragility curves and exceedance probability of damage levels, but does so only for one typhoon. Therefore, the curves cannot account for all damage levels related to wind speed. More case studies are needed. Also needed are curves that combine

construction typology and line of business with damage ratios. This two-part study offers two models of extreme events in order to highlight the importance of tidal gauge data and damage ratios for estimating return periods of storm surges and fragility curves. Because more accurate prediction of return period and exceedance probability can help prevent property damage and loss of human life, combining these two models should be the goal of future research.

References

[1] Lulseged Ayalew and Hiromitsu Yamagishi. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, 65(1), 15–31, 2005.

[2] Eric S. Blake, Edward N. Rappaport, and Christopher W. Landsea. The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts). *National Oceanic and Atmospheric Administration, National Weather Service, National Hurricane Center*, Miami, Florida, 2007.

[3] Esther Bommier. Peaks-Over-Threshold Modelling of Environmental Data. *Department of Mathsmatics, Uppsala University*, 2014.

[4] Jae-Young Chun, Kwang-Ho Lee, Ji-Min Kim, and Do-sam Kim. Inundation Analysis on Coastal Zone around Masan Bay by Typhoon Maemi. *Journal of Ocean Engineering and Technology*, 22(3), 8-17, 2008.

[5] Christopher G. Burton. Social vulnerability and hurricane impact modeling. *Natural Hazards Review*, 11(2), 58–68, 2010.

[6] Onelack Choi and Ann Fisher. The impacts of socioeconomic development and climate change on severe weather catastrophe losses: Mid-Atlantic Region (MAR) and the US. *Climatic Change*, 58(1), 149–170, 2003.

[7] Stuart Coles. An introduction to statistical modeling of extreme values. *Springer*, 2001

[8] Baoshan Cui, Chongfang Wang, Wendong Tao, and Zheyuan You. River channel network design for drought and flood control: A case study of Xiaoqinghe River basin, Jinan City, China. *Journal of environmental management*, 90(11), 3675–3686, 2009.

[9] J. David Cummins, Christopher M. Lewis, and Richard D. Phillips. Pricing excess-of-loss reinsurance contracts against cat as trophic loss. *University of Chicago Press*, 93–148, 1999.

[10] F.C. Dai, C.F. Lee, and Y.Y. Ngai. Landslide risk assessment and management: An overview. *Engineering Geology*, 64(1), 65–87, 2002

[11] Dakshina G. De Silva, Jamie Brown Kruse, and Yongsheng Wang. Spatial dependencies in wind-related housing damage. *Natural Hazards*, 47(3), 317–330, 2008.

[12] Jason P. Dunion, Christopher W. Landsea, Samuel H. Houson, and Mark D. Powell. A reanalysis of the surface winds for Hurricane Donna of 1960. *Monthly Weather Review*, 131(9), 2003

[13] Timothy M. Hall and Adam H. Sobel. On the impact angle of Hurricane Sandy's New Jersey landfall. *Geophysical Research Letters*, 40(10), 2312–2315, 2013.

[14] Dong-Soo Hur, Gyeong-Seon Yeom, Ji-Min Kim, Do-Sam Kim, and Ki-Sung Bae. Storm Surge Characteristics According to the Local Peculiarity in Gyeongnam Coast. *Journal of Ocean Engineering and Technology*, 20(3), 45-53, 2006.

[15] Dong-Soo Hur, Gyeong-Seon Yeom, Ji-Min Kim, Do-Sam Kim, and Ki-Sung Bae. Estimation of Storm Surges on the Coast of Busan. *Journal of Ocean Engineering and Technology*, 20(3), 37-44, 2006.

[16] Z. Huang, D.V. Rosowsky, and P.R. Sparks. Hurricane simulation techniques for the evaluation of windspeeds and expected insurance losses. Journal of Wind Engineering and Industrial Aerodynamic, 89(7), 605–617, 2001.

[17] Yunji Hwang. Stochastic analysis of storm-surge induced infrastructure losses in New York City. *Ph. D dissertation, Columbia University in the City of New York*, 2013.

[18] Intergovernmental Panel on Climate Change (IPCC). Contribution of working group I to the fourth assessment report of the intergovernmental panel on Climate change. *Cambridge University Press*, 1-996 2007.

[19] A.C. Khanduri and G.C. Morrow. Vulnerability of buildings to windstorms and insurance loss estimation. *Journal of Wind Engineering and Industrial Aerodynamics*, 91(4), 455–467, 2003.

[20] Yoon-Koo Kang. Patterns of Water Level Increase by Storm Surge and High Waves on Seawall/Quay Wall during Typhoon Maemi. *Journal of Ocean Engineering and Technology*, 19(6), 22-28, 2005.

[21] Barry D. Keim, Robert A. Muller, and Gregory W. Stone. Spatiotemporal patterns and return periods of tropical storm and hurricane strikes from Texas to Maine. *Journal of Climate*, 20(14), 3498–3509, 2007.

[22] Hyeon-Jeong Kim and Seung-Won Suh. Improved Hypothetical Typhoon Generation Technique for Storm Surge Frequency Analysis on the Southwest Korean Coast. *Journal of Coastal Research, Special Issue*, 85, 516-520, 2018.

[23] Ji-Myong Kim, Kiyoung Son, and Young-Jae Kim. Assessing regional typhoon risk of disaster management by clustering typhoon paths. *Environment, Development and Sustainability*, 2018.

[24] J. M. Kim, P. K. Woods, Y. J. Park, and K. Son. Estimating the Texas Windstorm Insurance Association (TWIA) claim payout of commercial buildings from Hurricane Ike. *Natural Hazards*, 84, 405-424, 2016

[25] Ji-Myong Kim, Paul K. Woods, Young Jun Park, Taehui Kim, and Kiyoung Son. Predicting hurricane wind damage by claim payout based on Hurricane Ike in Texas. *Geomatics, Natural Hazards and Risk*, 7(5), 1513-1525, 2015.

[26] Ji-Myong Kim, Taehui Kim, and Kiyoung Son. Revealing building vulnerability to windstorms through an insurance claim payout prediction model: a case study in South Korea. *Geomatics, Natural Hazards and Risk*, 8(2), 1333-1341, 2017.

[27] Tae-Yun Kim and Kwang-Woo Cho. Forecasting of Sea-level Rise using a Semi-Empirical Method. *Journal of the Korean Society of Marine Environment and Safety*, 19(1), 1-8, 2013.

[28] Korea Hydrographic and Oceanographic Agency. *Korea Real Time Database for NEAR-GOOS*

[29] Hyesun Ku, Sungsu Lee, and Yong-Kyu Lee. Statistical model and characteristics of typhoon-induced rainfall around the Korean peninsula. *Journal of the Korean Society of Hazard Mitigation*, 8(5), 45-51, 2008.

[30] Jong-Chan Lee, Jae-Il Kwon, Kwang-Soon Park, and Ki-Cheon Jun. Calculations of Storm Surges, Typhoon Maemi. *Journal of Korean Society of Coastal and Ocean Engineers*, 20(1), 93-100, 2008.

[31] Ning Lin, Kerry Emanuel, Michael Oppenheimer, and Erik Vanmarcke. Physically based assessment of hurricane surge threat under climate change. *Nature Climate Change*, 2(6), 462–467, 2012.

[32] N. Lin, K. A. Emanuel, J. A. Smith, and E. Vanmarcke. Risk assessment of hurricane storm surge for New York City. *Journal of Geophysical Research*, 115(18), 2010.

[33] Madeleine Lopeman, George Deodatis, and Guillermo Franco. A Critical Comparison of Windstorm Vulnerability Models with Application to Extra-Tropical Cyclones in Northern Europe. *International Conference on Structural Safety and Reliability (ICOSSAR)*, New York, NY, 2013.

[34] M. Lopeman, G. Deodatis, and G. Franco. Clustered Separated Peaksover-threshold Simulation—Estimation and uncertainty quantification of Hurricane Sandy's return period. *Engineering Mechanics Institute Conference*, Stanford, California, 2015.

[35] Madeleine Lopeman, George Deodatis, and Guillermo Franco. Extreme storm surge hazard estimation in lower Manhattan. *Natural Hazards*, 78, 335-391, 2015.

[36] Madeleine Lopeman, George Deodatis, and Guillermo Franco. Storm surge hazard estimation on the United States' Atlantic coast using the clustered separated peaks-

over-threshold simulation (CSPS) method. *International Conference on Risk Analysis*, Barcelona, Spain, 2015.

[37] Madeleine Lopeman. Extreme Storm Surge Hazard Estimation and Windstorm Vulnerability Assessment for Quantitative Risk Analysis. *Ph. D dissertation, Columbia University in the City of New York*, 2015.

[38] National Typhoon Center. Typhoon technical Book. *Korea Meteorological Administration, National Typhoon Center*, 2011.

[39] Jill M. Noel, Andy Maxwell, William J. Platt, and Linda Pace. Effects of Hurricane Andrew on cypress (taxodium distichum var. nutans) in south Florida. *Journal of Coast Research*, 21, 184–196, 1994.

[40] J. G. Park, W. S. Jeong, E. B. Kim, and J. S. Kim. A study on possible regional disaster prediction for the accompanying high winds in the storm. *Asia-Pacific Journal of Atmospheric Sciences*, 10(1), 306–307, 2011.

[41] J. G. Park, W. S. Jeong, Kim, J. S. Kim, and E. B. Kim. The trend of damage cost during Typhoon period affected in the Korean Peninsula. *Asian Journal of Atmospheric Environment*, 5, 152, 2012.

[42] Mark D. Powell and Samuel H. Houston. Surface wind fields of 1995 hurricanes Erin, Opal, Luis, Marilyn, and Roxanne at landfall. *Monthly Weather Review*. 126(5), 1259–1273, 1998.

[43] Mark D. Powell, Sam H. Houston, Luis R. Amat, and Nirva Morisseau-Leroy. The HRD real-time hurricane wind analysis system. *Jounal of Wind Engineering and Industrial Aerodynamics*, 77 and 78, 53–64, 1998.

[44] Mark D. Powell, Shirley Murillo, Peter Dodge, Eric Uhlhorn, John Gamache, Vince Cardone, Andrew Cox, Sonia Otero, Nick Carrasco, Bachir Annane, and Russell St. Fleur. Reconstruction of Hurricane Katrina's wind fields for storm surge and wave hindcasting. *Ocean Engineering*, 37(1), 26–36, 2010.

[45] Valentina Radic and Regine Hock. Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise. *Nature Geoscience*, 4, 91-94, 2011.

[46] Joao Lima Rego and Chunyan Li. On the importance of the forward speed of hurricanes in storm surge forecasting: A numerical study. *Geophysical Research Letters*, 36(7), 2009.

[47] Michiel Schaeffer, William Hare, Stefan Rahmstorf, and Martin Vermeer. Long-term sea-level rise implied by 1.5°C and 2°C warming levels. *Nature Climate Change*, 2, 867-870, 2012.

[48] H. Shin, W. Lee, G. Byun, J. Kim, K. Jang, and J. Lee. Evaluation of the numerical models' Typhoon track predictability based on the moving speed and direction. *Asia-Pacific Journal of Atmospheric Sciences*, 10, 372–373, 2013

[49] Masanobu Shinozuka, M.Q. Feng, Jonghoon Lee, and Toshihiko Naganuma. Statistical Analysis of Fragility Curves. *Journal of Engineering Mechanics*, 126(12), 1224-1231, 2000.

[50] Alex Stephenson. Harmonic Analysis of Tides, *CRAN*, 2017

[51] Carl Scarrott and Anna Macdonald. A review of extreme value threshold estimation and uncertainty quantification. *Statistical journal*, 10(1), 33-60, 2012

[52] S. A. Talke, P. Orton, and D. A. Jay. Increasing Storm Tides in New York Harbor, 1844-2013. *Geophysical Research Letters*, 2014.

[53] Peter J. Vickery, Peter F. Skerlj, Jason Lin, Lawrence A. Twisdale Jr., Michael A. Young, and Francis M. Lavelle. HAZUSMH hurricane model methodology. II: Damage and loss estimation. *Natural Hazards Review*, 7(2), 94–103, 2006.

[54] Pickands. Statistical inference using extreme order statistics. *The annuals of statistics*, 3(1), 119-131, 1975

[55] Charles C. Watson Jr. and Mark E. Johnson. Hurricane loss estimation models: Opportunities for improving the state of the art. *American Meteorological Society*, 85(11), 1713–1726, 2004.

[56] Jong Joo Yoon and Sang Ik Kim. Analysis of Long Period Sea Level Variation on Tidal Station around the Korean Peninsula. *Journal of Coastal Research*, 12(3), 299-305, 2012.

[57] Chris Zervas. NOAA Technical Report NOS Co-OPS 067: Extreme Water Levels of the United States 1893-2010. *Technical report, National Oceanic and Atmospheric Administration, National Ocean Service, Center for Operational Oceanographic Products and Services*, Silver Spring, MD, 2013.

[58] Guofang Zhai, Teruki Fukuzono, and Shaburo Ikeda. Multi-attribute evaluation of flood management in Japan: A choice experiment approach. *Water and Environment Journal*, 21(4), 265–274, 2007.

[59] H. Zhong, P. H. A. J. M. can Gelder, P. J. A. T. M. van Overloop, and W. Wang. Application of a fast stochastic storm surge model on estimating the high water level frequency in the Lower Rhine Delta. *Natural Hazards*, 73, 743-759, 2014.