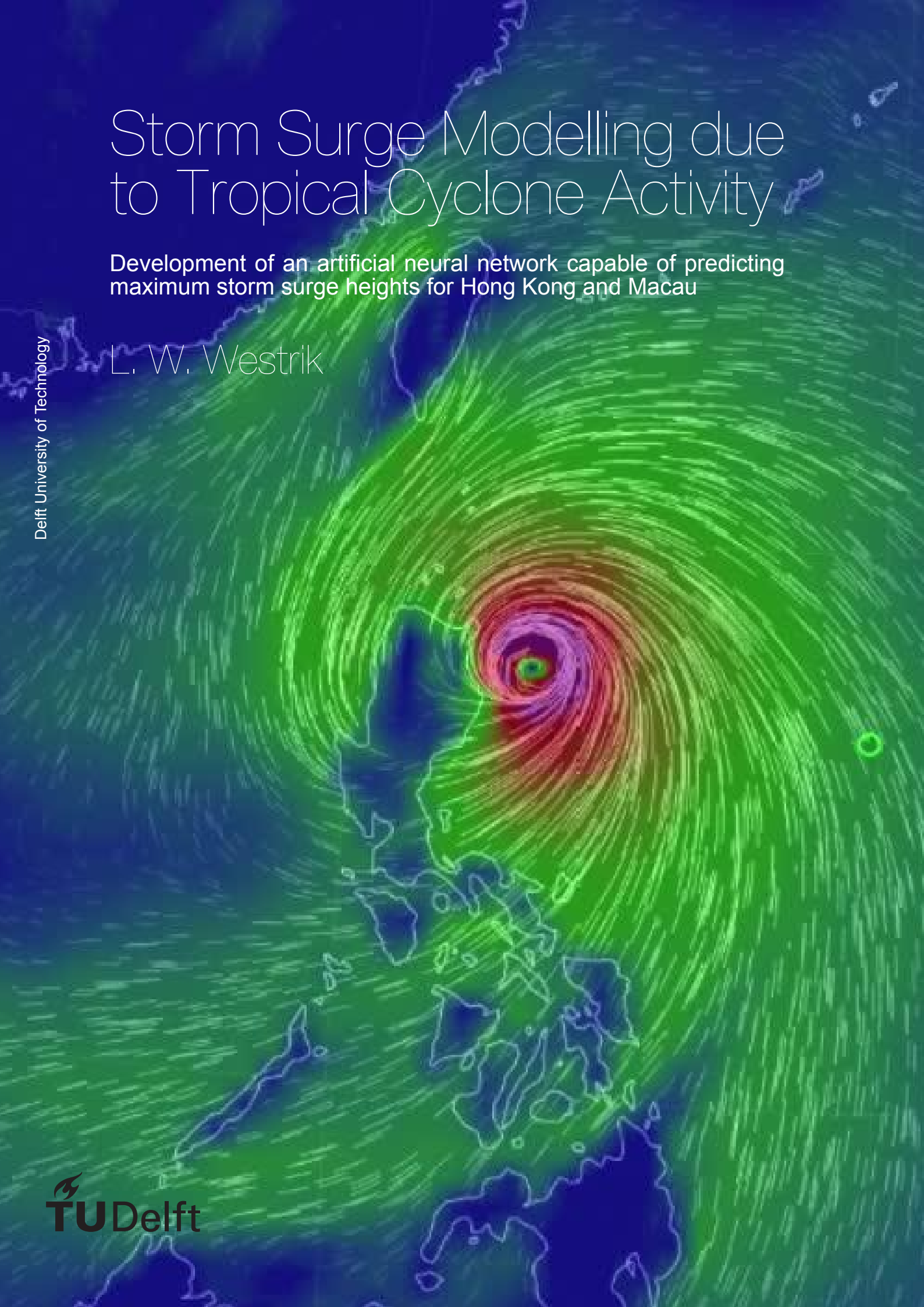
Storm Surge Modelling due to Tropical Cyclone Activity

Development of an artificial neural network capable of predicting maximum storm surge heights for Hong Kong and Macau

L. W. Westrik

Delft University of Technology



TU Delft

Storm Surge Modelling due to Tropical Cyclone Activity

Development of an artificial neural network capable of predicting maximum storm surge heights for Hong Kong and Macau

by

L. W. Westrik

to obtain the degree of Master of Science
at Delft University of Technology,
to be defended publicly on Friday November 13, 2020 at 02:00 PM.

Student number:	4250109	
Project duration:	November 18, 2019 – November 1, 2020	
Thesis committee:	Dr. ir. J. D. Bricker,	TU Delft, Committee chairman
	Dr. ir. M. A. Diaz Loaiza,	TU Delft
	Dr. ir. M. Zijlema,	TU Delft
	Prof. ir. R. Ranasinghe,	IHE Delft
	Dr. R. Meynadier,	AXA Paris

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

This project is the final part in completing my master's degree in Civil Engineering with an Hydraulic Engineering track. The project was carried out in context of a larger project between TU Delft, AXA Paris and IHE Delft that focuses on flood risk for coastal cities around the world. From the start, I knew this was going to be a challenging project. I had no prior experience with hydrodynamic modelling and neural networks. The opportunity to learn new modelling techniques and apply those to relevant problems in the real world sparked my interest. I think that hydrodynamic and machine learning techniques will become more and more important in the future. Therefore, I am glad to already have knowledge and experience on these topics. Although, it was not always easy, this final project combined with all the other courses offered in the Hydraulic Engineering track has provided me with the necessary knowledge and tools to start my professional career. Over the last few year I have been prepared to critically analyse and solve problems as an engineer. I would like to take this opportunity to thank a few people who helped my during this graduation research.

Firstly, the TU delft committee members who have supported me from the start of this project and allowed me to pursuit the topics I was most interested in. Jeremy Bricker, thank you for fulfilling the role as thesis committee chairman and giving me advice. You provided me with many ideas and solutions during the entire project. The weekly meetings with Andres Loaiza on Skype really helped to stay focused and improve my results during times of the Corona virus. I also would like to thank Andres Loaiza, as my daily supervisor. Your help and assistance was invaluable. You introduced me to hydrodynamic modelling, provided relevant data and most importantly was always open to answer questions. I also would like to thank Marcel Zijlema from the section Environmental Fluid Mechanics, for reviewing my report and scientific approach.

Secondly, I would like to thank my other committee members, Rosh Ranasinghe and Remi Maynadier. Your input during the thesis meetings has provided me with additional ideas to improve my research.

Finally, I would like to thank my mother Evelien for unconditionally supporting me throughout my time at university. Although you have limited knowledge in this field of expertise, you always offered a listening ear and motivated me to push through the tough moments for which I am very grateful.

L. W. Westrik
Delft, November 2020

Abstract

Over the recent years, flood risks and losses have been increasing for coastal cities due to climate change, subsidence, population and economic growth. Hong Kong and Macau are two cities located in the Pearl River Delta that experience a significant flood risk due to storm surges. The increased losses and risks has sparked interest around the world for efficient and accurate flood forecasting. At the moment coastal flooding events are often simulated with difficult hydrodynamic models that reproduce the physical phenoms. Over the last decades there has been more interest in other methods to forecast storm surges, namely neural networks. Other than hydrodynamic models a neural network is capable is making predictions in seconds, while the model can take hours to finish simulation. fast and accurate storm surge forecasting is of importance for disaster and evacuations management strategies and will only become more important in the future. During this research the main goal is to develop a neural network capable of prediction maximum water levels due to storm surges in case of an approaching tropical cyclone. The neural network is trained with data that is obtained from hydrodynamic simulations. A synthetic storm database is used to provide the necessary data to conduct 1000 simulations of which the results are used in the neural network.

The first step is developing a hydrodynamic model capable of accurately simulation tropical cyclone induced storm surges in Hong Kong and Macau. The model is calibrated in a way to reproduce the real world as close as possible. It accounts for the real life bathymetry, topography, tidal elevations and wind forcing. The storm surge model is validated extensively based on three historical tropical cyclones. By comparing the actual observed water level during these storms with the model output, one can proof that the model output is accurate and can be used for the synthetic simulations. The errors between the observed and simulated water level are below 20 cm, after calibrations of different physical parameters within the model.

The second step of this research is to use the validated storm surge model to run synthetic storm simulations. Instead of using historical storms who only have been recorded for 40 years, a synthetic storm database is used containing 10000 years worth of data. From that storm data, 1000 synthetic tropical cyclones are selected that come close to Hong Kong and Macau. Where it is not possible to obtain enough data for neural network training with historical storm data. The synthetic database provides the ability to produce sufficient number of samples for network training. Additionally during this research multiple Matlab tools are developed that provide a high degree of automation for synthetic model setup and data processing. These tools can be easily adapted and used for storm surge simulation around the world.

With the data obtained from the previous two steps, it is finally possible to training the neural network. In this network a total of seven input parameters (tropical cyclone track parameters) are used to estimate the maximum water level that will occur during the tropical cyclone. The input parameters considered are: latitude, longitude of TC eye, maximum wind speed, minimum eye pressure, radius of maximum winds, forward speed, forward propagation direction. Based on these TC track parameters, the network should be able to accurately predict the surge heights. During the development of the network, three different types of configurations are tested. The first one: A complex neural network capable of outputting the entire maximum water level map for Hong Kong and Macau. The second configuration only focuses on the water levels at the coastline. When considering only water level prediction for the coastline, the complexity of the NN can be significantly reduced and more configurations can be tested. The final configuration only gives the output for 10 locations of interest instead of the entire domain.

Based on extensive calibration and validation it is concluded that the neural network capable of map predictions cannot be trained sufficiently to produce accurate results. The sheer size of the output

layer makes this network very complex and is accompanied with limitations in network training and architecture. The trained map neural network, gives errors up to 1 meter. The coastline network shows significant improvements compared to the map output with much lower mean square errors. However, variability in the quality of predictions are observed. For most combinations of input parameters reasonable predictions are obtained. However, for some other input parameter combinations the predictions are poor. Additional efforts should me made to try to improve the map and coastline network for applicability to real storm surge forecasts. The complexity of the neural network with 10 locations is much simpler. This provides more training and configuration possibilities than the other network. After training, this network provides much more accurate results. Based on the track parameters, the trained network is capable of prediction the maximum water level for the 10 location with a maximum error of 30 cm. Although the neural network is trained with synthetic data, it can be used for real life storm predictions by taking the track parameters of an approaching storm. Only the tidal elevation must be added to the neural network output to be used for real storms. The estimations by the locations network are considered to be reasonable accurate but shows variability in the accuracy between combination of input parameters. The networks should be used with care.

Contents

List of Figures

List of Tables

Glossary

Artificial Neural Network Mathematical model that can be modelled and trained for performing particular tasks based on available data.

Computational Model The recreation of physical processes by means of an numerical computer simulation .

Delft3D Open source hydrodynamic modelling software capable of simulating a wide variety of flow processes.

Hong Kong Densely populated city inside the special administrative region of the people's republic of China. It is a former British colony, consisting of 236 smaller islands with 7.4 million inhabitants in 2019.

Macau Like Hong Kong also a special administrative region and a former Portuguese colony. The city is located approximately 70 km from Hong Kong with a population of 650,000 in 2017.

Pearl River Delta Located on the coast of the South Chinese Sea, this is one of the most densely urbanised delta of the world.

Storm Surge The rapid rise of water due to low-pressure systems like tropical cyclones. Known to cause devastating floods in coastal flood planes..

Synthetic Storm Data Artificially generated best track data of tropical cyclones. It can be used instead of historical storm data for a wide variety of purposes like storm simulations, statistics, machine learning etc.

Tropical Cyclone Rapid rotating storm system with a low-pressure center and very strong winds. Depending its locations and or strength it referred by different names like hurricane, typhoon, tropical depression..

Acronyms

AAL Anual Average Loss.

BR Bayesian Back propagation Algorithm.

CEDD Civil Engineering and Development Department Hong Kong.

CNA Chinese Meteorological Agency.

FES2014 Finite Element Solutions tide model.

GEBCO the General Bathymetric Chart of the Oceans.

GHG Greenhouse Gasses.

GRM Groupo Risk Management.

IBTrACS International Best Track Archive for Climate Stewardship.

IPCC Intergovernmental Panel on Climate Change.

JWTC Joint Warning Typhoon Centre.

LM Levenberg-Marquardt Algorithm.

ML Machine Learing.

MSE Mean Square Error.

MSL Mean Sea Level.

NN Neural Network.

NOAA National Oceanic and Atmospheric Association.

OET Open Earth Tools (Deltares toolbox).

PRD Pearl River Delta.

R50 Radius of 50 knot Wind.

RCP Repesentative Concentration Pathways.

RMSE Root Mean Square Error.

RMW Radius of Maximum Wind.

RSMC Regional Specialized Meteorological Center Tokyo (Japan meteorological agency).

SCG Scale Gradient Conjugate Algorithm.

TC Tropical Cyclone.

TPXO Global Tide Models.

WES Wind Enhancement Scheme (Delft3D Tool).

Xtide Harmonic tide clock and tide predictor.

List of Symbols

Physical Constants

ρ_a	Density of air for 30 °C	$1.1644\ kg/m^3$
ρ_w	Density of sea water	$1025\ kg/m^3$
κ	Von Kármánn constant	$0.4\ [-]$
g	Gravitational acceleration	$9.81\ m/s^2$
R_{earth}	Radius earth	$6371000\ m$

Wind modelling

γ	Veltcheva and Kawai parameter	$-$
θ	angle between observation point and direction of TC movement	$^\circ$
A	Location parameter Holland wind model	$-$
B	Shape parameter Holland wind model	$-$
f	Coriolis parameter	s^{-1}
FD	Forward direction TC	$^\circ$
FS	Forward propagation speed TC	m/s
p	Surface pressure	hPa
p_c	Central pressure	hPa
p_n	Ambient pressure	hPa
Rw	Radius of maximum winds	m
V_c	Cyclostrophic wind speed	m/s
V_g	Gradient wind speed	m/s
V_{geo}	Geostrophic wind speed	m/s
V_m	Maximum wind speed	m/s

Hydrodynamics

α	Phase angle	$-$
γ_b	Breaker parameter	$-$
η	Water level elevation	m
ν	Kinematic viscosity	m^2/s
ω	Angular frequency	rad/s
τ_{wind}	Wind shear stress	N/m^2
c	Wave propagation speed	m/s

C_d	Wind drag coefficient	–
h	Water depth	m
H	Wave height	m
$H_{1/3}$	Significant wave height	m
H_{m0}	Zero order spectral wave height	m
k	Wave number	–
L	Wave length	m
S_{xx}	Radiation stress	–
T	Wave period	s
U_{10}	Wind speed at 10m above surface	m/s
W	wind speed at water surface	m/s
E	Wave energy	J/m^2
Neural Network		
b	threshold bias	–
f	Activation function	–
e	Global error	m
t	Target vector	–
W	Weight vector	–
X	Input vector	–
Y	Output vector	–

1

Introduction

Global and mega cities are often referred as the engines of economic growth and serve as a hub within a globalized economic network. Most global cities are located on or near the coast on low lying flat areas. Often they are also located near major river mouths, providing easy access to the hinterland. The favourable locations of these cities facilitates trade and welfare growth (de Sherbinin et al., 2007). On the other hand, their location close to the sea and major rivers often puts them at greater risk for natural hazards. Due to the changing climate, subsidence, population and economic growth, the flood risks and losses are increasing in coastal cities (Hallegatte et al., 2013).

Due to the worldwide increases of flood losses, there is a growing interest in the monitoring, modelling and prediction of coastal floods. Storm surges and waves, generally generated by tropical cyclones are considered the main drivers for coastal flooding. The knowledge on flood exposure and losses of coastal cities is of increasing interest for the French company AXA, a worldwide leader in insurances. The AXA group risk management (GRM) department is collaborating with Delft University of Technology and UNESCO-IHE Delft to model, quantify and predict coastal floods around the world. This MSc thesis aims at simulating storm surges in Hong Kong and Macau based of tropical cyclones. The storm surge model will provide the basis for development of a machine learning framework that can accurately predict storm surges.

1.1. Background and motivation

Situated in the Pearl river delta on the South coast of China, Hong Kong and Macau are two densely populated areas inside the special administrative region of the people's republic of China. Hong Kong a former British colony, consists of 236 smaller islands with 7.4 million inhabitants in 2019. Macau, a former Portuguese colony is located approximately 70 km from Hong Kong with a population of 650,000 in 2017. Both Hong Kong and Macau are among the most densely populated cities around the world. Hong Kong has a land area of 1106 square kilometers and a population density of 6600 people per square kilometer. With 30.4 square kilometers of land surface, Macau's area is small compared to Hong Kong. However, the population density is significantly larger with 21,400 people per square kilometer (Kwong and Wong, 2017). With 62 million inhabitants, the pearl river delta is one of the largest megalopolis of world.

Both Hong Kong and Macau have a rich history with natural disasters like tropical cyclones (extreme wind and heavy precipitation), storm surges and earthquakes which have caused floods, storm damages and land slides (Sim et al., 2018). According to the sustainability index 2015, Hong Kong is listed as the Asian city that is most at risk from natural hazards (Arcadis, 2015). From a worldwide perspective, Hong Kong is listed as the third most city at risk. The annual tropical cyclone season is from March to early September and on average 6 to 7 tropical cyclones affect Hong Kong each year (Lam and Lam, 2005). Especially the low lying areas of Hong Kong and Macau are vulnerable to storm surges caused by tropical cyclones. In the past, storm surges induced by tropical cyclones have caused significant floods, damages and casualties in Hong Kong. The most devastating observed storm surge

occurred in 1906 and caused approximately 15,000 casualties in Hong Kong (Lau et al., 2017). Other natural hazards that Hong Kong and Macau have experienced in the past are extreme precipitation events (often during a TC), wildfires and earthquakes. Studies by Du et al. (2018), Li et al. (2018), Sepúlveda et al. (2019) conclude that the current risk for a high intensity earthquake or tsunami in the pearl river delta is low. However, the study by Li et al. (2018) also concludes that the tsunami hazard will increase significantly if sea level rise is considered.

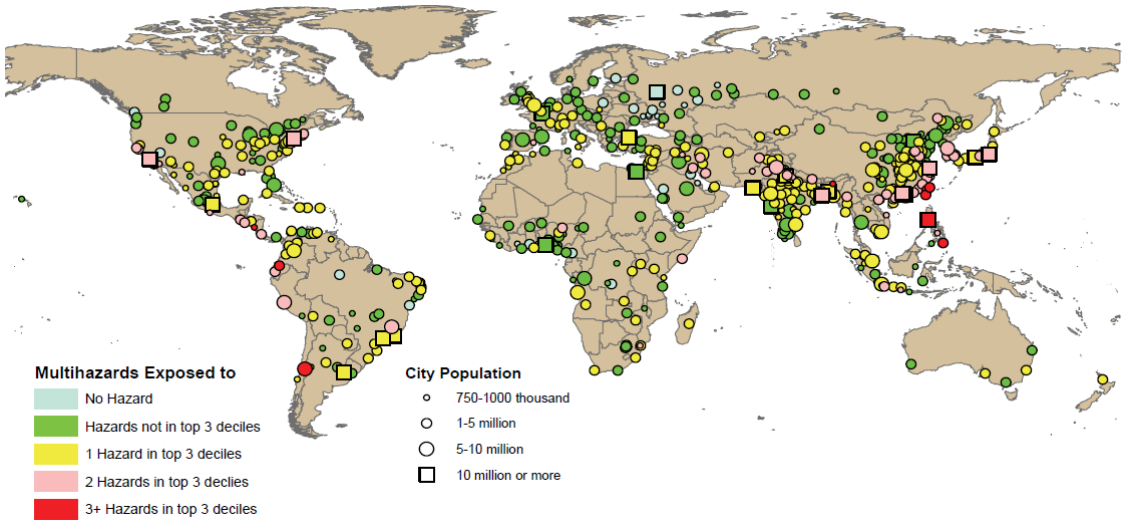


Figure 1.1: Distribution of cities by population size in 2011 and risk of natural hazards. Adapted from:United Nations (2011)

The significant natural hazard threat for Hong Kong and Macau is further underlined by a report from the United Nations (United Nations, 2011). Figure 1.1 displays the worlds largest cities and the natural hazards they are exposed to. Hong Kong and Macau are exposed to 2 hazards that fall in the top 3 deciles. This means that for two natural hazards, the exposure risk to the population is high. For example the 10th decile means that 100 % of the population is exposed to that hazard. For the Hong Kong region, flood and cyclone risks are listed in the 8-10th decile, meaning that there is a very high risk for population exposure to flood and cyclones. The multi hazard annual average loss (AAL) is often used in risk management and represents the long-term expected annual loss per year for hazards averaged over multiple years. For Honk Kong, the AAL due to storm surges is approximately 898 millions US dollars. Storm surges contribute 79% to the total AAL of Hong Kong (Preventionweb, 2019). Of all the reported losses of life during natural disasters between 1990 and 2014, loss of life due to floods account for 67.2% of all losses (EM-DAT, 2019). The AAL and loss of life statistics show, that storm surges and its associated floods are responsible for the major part of losses due to natural disasters in Hong Kong. (economic losses and loss of life). An additional trend is the rapid urbanisation of large cities. For Hong Kong it is projected that the population will grow to 9.3 million by 2050. Macau is projected to grow to 824.000 inhabitants by 2050 (United Nations, 2011). Due to the growing urban population, more people and infrastructure will be exposed to natural hazards in the future, which lead higher risks, more potential damages and loss of life in case of natural hazards.

Not only the rapid urbanisation but also climate change poses new problems for Hong Kong and Macau. Climate change is considered an important driver for many natural hazards. The most recent climate change report from the intergovernmental panel on Climate change (IPCC) discusses the observed and projected climate change (Yasuaki et al., 2014). The report shows that the mean annual temperature in East Asia has been increasing up to 0.2°C per decade since 1960. Precipitation trends on the other hand show great variability between regions and not significant trends can be observed on the return periods of tropical cyclones making landfall. Finally, all over the Asian region sea level rising has been observed. The magnitude of the sea level rise is specific to the ocean basin. Besides the projected climate change the IPCC report also treats the projected climate change. The assessment shows that further rising of temperatures in the 21st century is very likely. Models and simulations

shows that the mean temperature for South and Southeast Asia is projected to rise with more than 3 °C by the mid-21st century compared to the late-20th century. Furthermore, the precipitation and ocean temperatures are expected to increase under all considered scenarios. For the sea level rise the IPCC considers different scenarios that estimate a global sea level rise for the 21st century. The worst scenario predicts a global sea level rise up to 2m. Overall, the IPCC projects an increase in riverine, coastal and urban flooding leading to damages to infrastructure, livelihoods and settlements in Asia (Yasuaki et al., 2014). Furthermore, an increased risk of flood related deaths and injuries is projected with medium confidence. The people that live in low-lying coastal areas are most at risk from climate change.

According to Sim et al. (2018), Hong Kong can be considered a disaster resilience city overall. Risk mitigation strategies are developed and sufficient financial resources are allocated. However, there is still much room for improvement. Disaster management in Hong Kong focuses mainly on building resilience and emergency responses. However, a long term disaster strategy and vision is missing. Especially the response to climate change is insufficient and the government is failing to adapt an integrated approach. Part of the problem is the fact that Hong Kong and Macau have not experienced natural disasters with significant fatalities over the last 50 years. This inexperience in natural disasters and the lack of a long term vision suggests that it is unclear if Hong Kong and Macau can cope with the impacts of disasters in the future. On local scale very little can be done to reduce tropical cyclone activity and storm surges. However, the study and simulation of the impacts of TC's and storm surges for Hong Kong and Macau can lead to new insights to improve the disaster resilience.

1.2. Objective

In the previous section, the background and motivation for the study has been discussed. In summary, the natural hazards, rapid urbanisation and changing climate will increase the risk and consequences for Hong Kong and Macau in the future. From all discussed natural hazards, tropical cyclones and its associated flooding poses the most danger for Hong Kong and Macau. Therefore, this research mainly focuses on the simulation and prediction of coastal floods due to TC activity. Recently TC Hato (2017) made landfall in the pearl river delta. It was one the strongest tropical cyclones of the last decades. Hato severely impacted Hong Kong and Macau causing flooding, electricity and water shortages bringing the cities to a standstill. In Hong Kong no fatalities were registered but in Macau 10 people died and over 200 people were injured. Moreover, parts of the city were inundated due to the unprecedented storm surge heights. Track analysis and surge investigation showed that Hato had unusually rapid approach speeds (Takagi et al., 2018). Authorities did not have sufficient time to prepare and storm warning signals were not issued in time. This shows the importance for improvements in tropical cyclone and storm surge forecasting.

Presently, different numerical hydraulic models are used to simulate coastal flooding in case of a tropical cyclone. The numerical models use complex mathematical and physics processes to describe the flow of water and predict floods. Although these models have shown in the past that they can produce accurate results, their accuracy is depending on the model selection, setup, validations and analysis. Due to their complexity, the simulations often have long computational time. In case of an approaching TC, predictions on surge heights must be computed quickly. However, the more accurate the model the longer the computation time becomes. In the past decades new methods have been developed to predict floods by mimicking the mathematical and physical processes. Today, these machine learning methods like neural networks, have contributed in the advancement of prediction systems, providing better performance and cost-effective solutions (Mosavi et al., 2016). An added bonus of these ML methods is that they have a significant faster computation time. However, the development of ML methods requires historical or synthetic data to work. The data needed to develop an ML method can also be obtained by doing many simulations in a hydrodynamic model and use its results for the development of the ML method. The main objective of this research is the development of an hydrodynamic model that will be used to develop an artificial neural network that uses TC storm parameters only to accurately and fast predict storm surge heights for Hong Kong and Macau. Additionally, three sub research questions are considered in order to answer to main research question. These sub questions mostly relate back to the preliminary work that must be completed in order to start development

of the artificial neural network.

Main research question:

- Is it possible to develop a neural network using synthetic storm data that is capable of accurately prediction the maximum water levels in case of an approaching tropical cyclone?

Sub research questions:

- What is the best way to set-up and calibrate a hydrodynamic model accurately for a large number of synthetic storm simulations?
- How should a synthetic storm database be implemented to efficiently process large amounts of data for use in synthetic storm surge simulations and neural networks?
- What are the advantages and or disadvantages of using neural networks to predict maximum storm surge heights, compared to traditional hydrodynamic modelling?

Multiple research activities are defined to work systematically towards answering the main research question. These activities form the basics of the proposed methodology.

Research activities:

1. Literature study and data collection
2. Hydrodynamic model set-up in Delft3D (wind model, waves, tides)
3. Model validation with historical TC's (Hagupit 2008, Hato 2017, Mangkhut 2018)
4. Synthetic storm catalogue data preparing and handling
5. Up to 1000 storm surge simulations with synthetic TC track catalogue. Simulation results will be used for the development of the neural network.
6. Development of an artificial neural network capable of prediction maximum surge heights for Hong Kong and Macau in case op an approaching TC.

1.3. Methodology

The methodology of this research can be best described by dividing the project into three phases i.e. research initiation, Hydrodynamic processes and machine learning processes.The different activities all contribute towards answering the research questions. This section aims at describing the different activities in more detail. Figure 1.2 shows the different phases during the thesis and figure 1.3 shows a flow chart of the proposed activities.

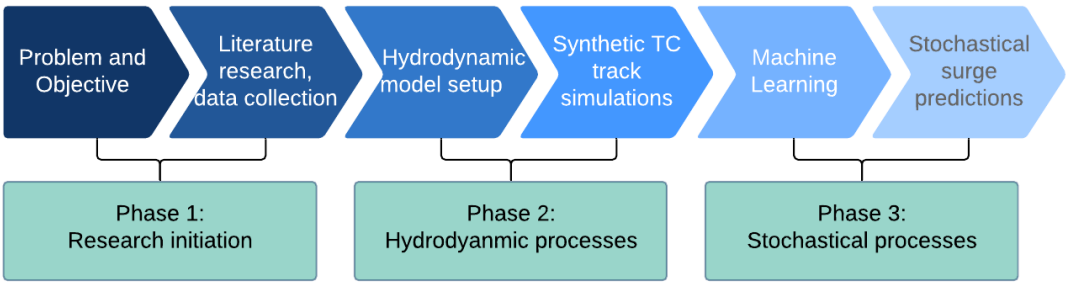


Figure 1.2: Phases Overview Thesis Research

Phase 1: Research initiation
The starting point for this research are the problem and objective statements. In sections 1.1 and 1.2 the background, motivation and objective of the research are discussed. It discusses the factors and aspects that are responsible for the conduction of this study. Chapter 2 focuses on the literature study and data collection. The literature study includes all the theoretical and physical background needed to answer the research question. Data collection involves gathering all the required data that is needed for the research. The collected data is used in phase 2 for the storm surge simulations. This includes the collection of topography and bathymetry maps for Hong Kong and Macau (provided by Delft University of Technology). The ocean tides, required as boundary conditions input in the model will be adapted from global and regional tide models (TPXO) or FES2014 model. The tidal models allows the user to include the tidal forces as harmonics in the model instead of time dependent measurements. Data on historical TC tracks are provide by the IBTrACS database. Furthermore, historical measured data from tidal and wave stations around Hong Kong and Macau are gathered from local Authorities. The data is provided by the Hong Kong Observatory, Hong Kong Marine Department and the Civil Engineering and Development Department (CEDD).

Phase 2: Hydrodynamic processes
The second phase consists of the model set-up in Delft3D. In section 2.6, different types of hydrodynamic models and software packages are discussed. For this research, the Delft3D model will be used to simulate storm surge and coastal flooding. Delft3D is a widely used and validated package around the world. The additional packages to integrate tides, waves, wind speeds and pressures in to model increase the ease of use. An additional advantage for the Delft3D model is the fact, that the Delft University of Technology thesis supervisors have a lot of experience using this software.

The accuracy of the storm surge predictions is largely dependent on the TC track data, intensity and wind fields (Bao et al., 2006). Therefore, computations of wind speeds and pressures for the input of the model are derived from the parametric Holland et al. (2010) wind model. The Holland model uses the sea level pressure and wind profiles to provide estimates of required parameters. This model is used frequently to simulate TC's and a low computational cost, which makes it suitable for a large number of simulations. The Holland wind model is extensively treated in Section 2.5. With all the gathered data, the hydrodynamic model can be set-up for estimation of surge heights and coastal water levels.

After development, the model will be validated. This can be done by hind casting a historical storm. The validation process consists of two parts. First, the model is tested with only tidal forcing and no waves or storms. The coastal water levels of the simulation are then compared with the historical tidal data from the measurement stations. By comparing the simulated and historical water levels, one can determine whether the model is capable of accurately simulating tides. The tidal validation part will be done for a historical day with very calm conditions (hardly any waves and wind). Adapting this approach, ensures that the data of the tidal stations can be compared with the model output without interference of the waves and wind. After the tidal part of the model has been validated, the second part of the validation process can be conducted. This includes a simulation with tides, waves and storm conditions combined. For this part it is necessary simulate an historical TC and compare it with the model input. TC Hato (2017) will be used for this second validation part. Again the results of the simulation will be compared with the historical tidal and wave data.

After the model has been tested and validated, one can start with the synthetic storm simulations. A synthetic TC track database is required since there are not sufficient historical TC tracks in the PRD region to conduct a correct machine learning analysis. Recently, Bloemendaal et al. (2020) developed a synthetic storm catalogue containing 10,000 years worth of storm data for all basins around the world. This database is generated from observed storm track data of historical TC's. A number of relevant storms will be selected and their track data is parametrized by a wind model. This will then be used as the main forcing in the Delft3D model to simulate many storm surges. he results of all these simulations are then used for the training of the machine learning framework in the final phase.

Phase 3: Machine Learning Method

The final phase includes the development of the neural network. As discussed previously, neural networks are a great tool for the estimation of surge heights, arrival times and flood depths at a low computational cost. The development of an accurate neural network can significantly reduce the time required for storm surge forecasting. The reduced time required for forecasting can help authorities in the future for disaster and evacuations management. Additionally, by increasing knowledge on the surge heights around Hong Kong and Macau, possible weak spots in the flood defences can be identified. This again gives authorities the possibly the most vulnerable locations, which then can be evacuated in time, when a TC approaches the area. Multiple different neural network architectures exist today that can be used for forecasting purposes. In general Bayesian Networks and artificial neural networks are used often for forecasting.

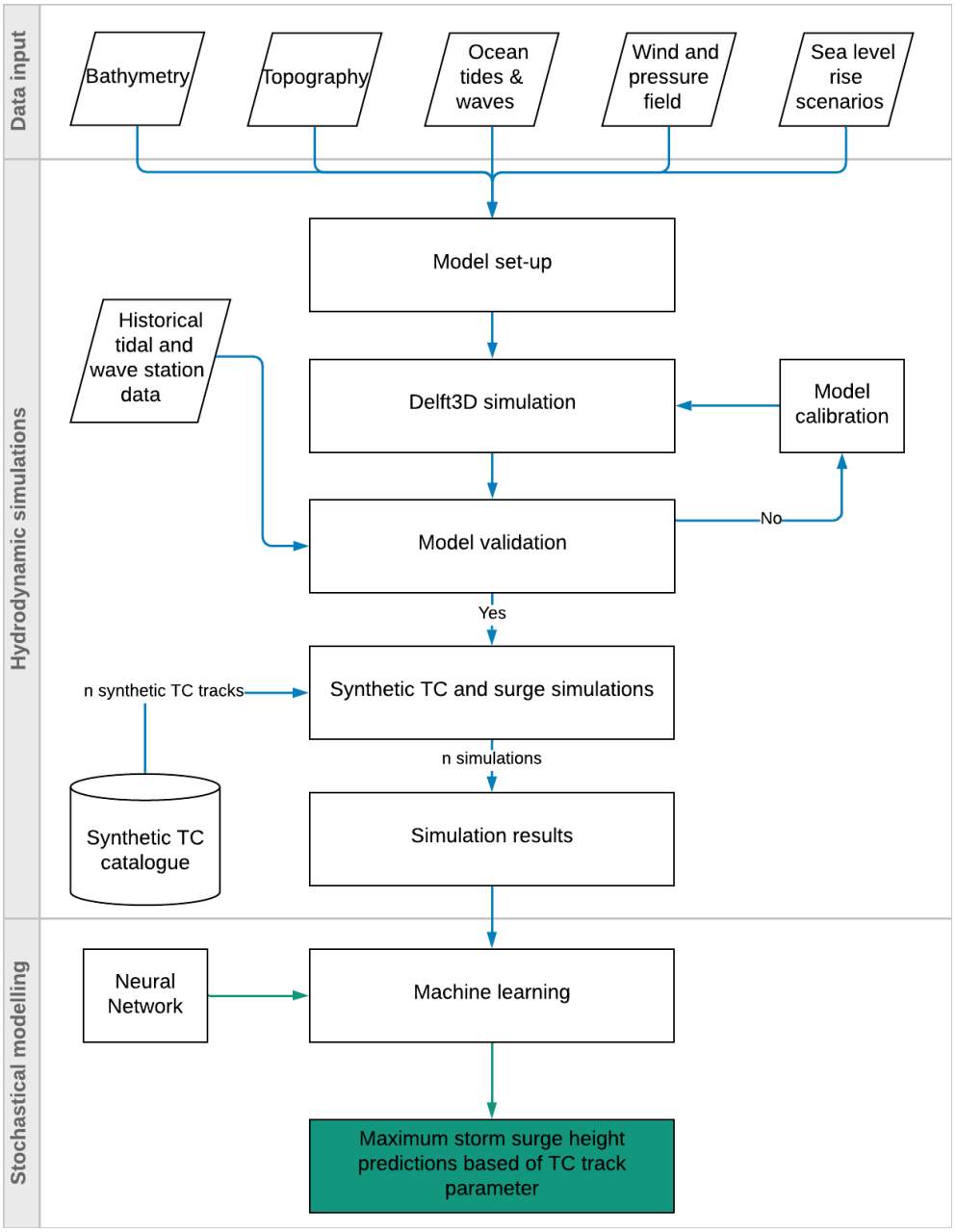


Figure 1.3: Flow chart of thesis project

1.4. Readers Guide

The section is written with the aim to give the readers insight in the structure of this thesis research. Chapter 2 focuses on reviewing the fundamental literature that is used during this research. It treats the physical process associated with storm surges and tropical cyclones. Furthermore, the background of hydrodynamic and neural network modelling is discussed. Chapter 3 aims at setting up and validating of the hydrodynamic storm surge model. All aspects that are relevant for model setup is treated in detail. The final product is an validation model capable of prediction storm surge height for tropical cyclones. Chapter 4 contains all steps to efficiently handle the synthetic storm data and make it suitable for synthetic storm surge simulations. In chapter 5 the results of the synthetic simulations are used for the development of a neural network capable of predicting maximum surge height for TC track parameters. The network is tested and validated extensively to ensure the production of accurate results. Chapter 6 discusses the results of the hydrodynamic and synthetic modeling. Finally, chapter 7 discusses the conclusion of this study. Additionally, recommendations for future research possibilities are given.

2

Literature

This chapter aims at describing all the theoretical and physical backgrounds required for this research. Furthermore, this chapter also present the collected relevant data on astronomical tides, waves and wind that will be used for the coastal flooding model. First, The formation and structure of TC's will be discussed. This is followed by a theoretical explanation of storm surges, waves and tide processes. The final part of this chapter focuses on the numerical hydrodynamic models that are used in the study. Different models will be discussed and their advantages and limitations. A review on the different methods is required to correctly choose the methods to be used during the study. This chapter often refers to appendix A that contains additional theory and background information to supplements the topics discussed in this chapter.

2.1. Tropical cyclones

Tropical cyclones are large scale warm-cored rotary storms that form over warm ocean waters in tropical regions and are driven by heat transfer from the ocean (Emanuel, 2003, Montgomery and Farrell, 1993). These storms are characterised by a low pressure system, strong winds, heavy rain and thunderstorms. The extreme winds and rain accompanied by storm surges can have devastating effects on coastal regions in the tropics. The regions in which these storm form are around the equator and is bordered by the tropic of Cancer ($23.5°N$) and the tropic of Capricorn ($23.5°S$). Depending their location in the world, tropical cyclones are named differently. In the Atlantic ocean they are called hurricanes and in the Pacific ocean they are called Typhoons. In the Indian ocean they are simply called tropical cyclones or cyclones. Although named differently, these storms all refer back to the same type of storm. The classification of tropical cyclones is defined according their maximum wind speed at an altitude of 10 m, averaged over 10 minutes (Bell, 2010). Three different intensity classes can be distinguished: Tropical depressions with maximum wind speeds of 17 m/s, tropical storms with wind speeds ranging from 18 to 32 m/s and finally tropical cyclones with a maximum wind speeds over 33 m/s. This classification depends on the time averaged wind speed, however gusts can be significantly higher. For hurricanes, a more extensive classification is used (Saffir-Simpson scale) with five different intensity categories.

2.1.1. Tropical Cyclone structure

As stated by Kepert (2010), tropical cyclones can vary significantly from one to another, and from day to day, in intensity, size, boundary layer structure, spiral banding, eye structure and degree of symmetry. These complex and different parts have been researched extensively in the past decades. However, physical changes processes responsible for structure and intensity changes are still not understood well today. To increase knowledge, data from these storms has been gathered by means of airplanes, sensors, radar, satellites and numerical simulations (Emanuel, 2003). Wang and Wu (2004) describes the TC structure by defining three components: storm-scale structure, inner-core and spiral rainbands. The storm-scale structure represents the whole cyclonic circulation with sustained winds larger than 15 m/s. The inner-core includes the eyewall and eye of the storm. Figure 2.1 shows an schematic of a TC structure with it main components.

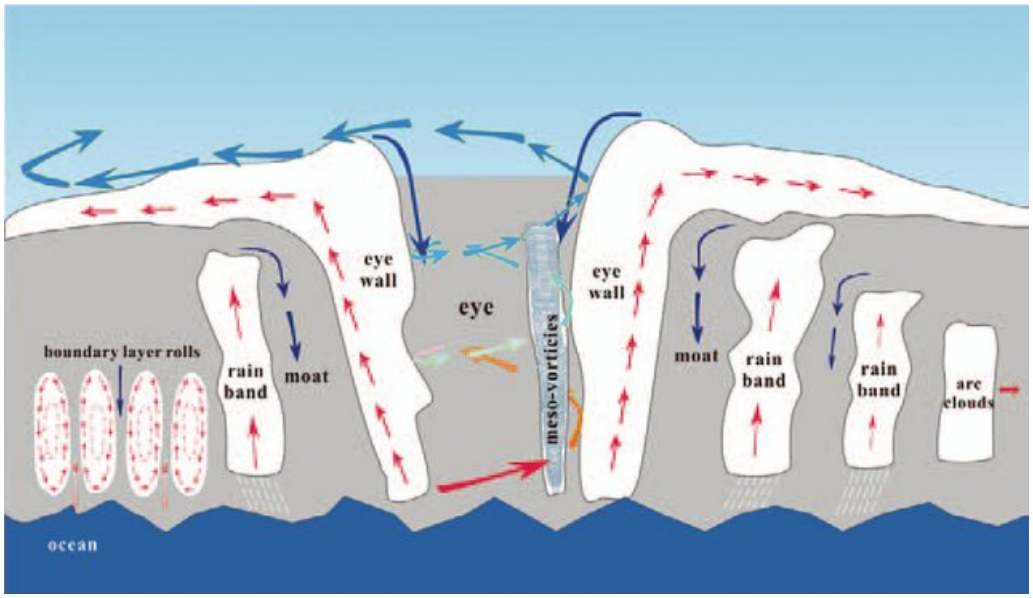


Figure 2.1: Schematization of TC structure. Adapted from:Li et al. (2013)

The TC center is called the eye and has relatively calm conditions compared to the rest of the storm system. Storm eyes can often be easily seen in case of well-developed storms with high intensities. For less intense storms, the eye is usually less defined. The eyewall is a ring of very deep convective cloud with thunderstorms extending from the edge of the eye to 20-50 km outwards (Wang and Wu, 2004). Usually the eyewall has the highest winds and precipitation. Another component of the TC structure are the spiral rainbands. The spiral rainbands are located outside the inner-core of the storm and often produce extreme precipitation. The rainbands of a storm system can extend up to 150 km from the eye. According to Wang and Wu (2004) the dynamics in the TC core are believed to be the key in structure and intensity changes. TC's can vary significantly in size.

2.1.2. Formation

The formation of tropical cyclones is often referred as cyclogenesis. Today, this field of research is still not completely understood. However, there is general consensus regarding the climatological conditions in which TC's can form. According to Tory and Frank (2010) the following five conditions can be distinguished that are favorable for the formation of TC's:

- Sea surface temperature above 26.5 °C coupled with relatively deep ocean mixed layer. In general the TC uses the heat of the ocean to gain ins tensity and strength. The warm ocean water is evaporated and included in the developing TC. Eventually, the water vapor condensates to form clouds.

- A deep surface based layer of conditional instability. This basically means that there must be a pre-existing disturbance presence which can grow into a TC.

- Enhanced values of cyclonic low level absolute vorticity. The easiest way to describe vorticity is simply as a clockwise or counter clock wise movement.

- Organized deep convection in an area with large scale mean ascent and high to mid level humidity. TC need the humidity in troposphere to sustain the thunderstorm and clouds. A low humidity will cause the dry air the start eating away the clouds.

- Weak to moderate vertical wind shear. Vertical wind shear is defined as the difference in wind speed and directions at two different heights in the atmosphere. In general TC have problems developing in area with high wind shear, meaning that the disturbance are displaced way from the TC centre.

Figure 2.2 displays a schematic diagram of a TC lifecycle and shows the intensity changes over time. The formation typically start with warm ocean temperatures, relatively low surfaces pressures and a cluster of thunderstorms (initial disturbance) (Pielke and Pielke, 1997). Air is attracted towards the center of the low pressure system (convergence). The Coriolis effect deflects the wind creating vorticity. The warm sea level temperature provides the heat for evaporation and adds heat to the lower atmosphere. The to the low pressure spiraling air is forced up, since is has no way to go. If the troposphere's humidity is favorable (mid to high humidity), clouds can develop up to top of the troposphere. The stratosphere acts as an barrier for upward moving heat and moisture. When the stratosphere is reached, the air spreads diverges. The diverging air caused an additional pressure drop at the surface, leading the extra air convergence which again feeds the thunderstorms and clouds. Due to the small windshear the thunderstorm stay concentrated, allowing the system to further grow. By now the system is called a tropical disturbance (cluster of thunderstorms and weak wind circulation). If conditions stay favorable, the disturbance can keep developing and grow into a TC. As the system increases in strength it becomes more difficult for air to reach the centre. This is cause by the increase in speed when the air spirals towards the center. As a result the air is spun out of the inwards spiral and spiral bands are formed. Due to the inflow of air at low level, the spiral bands can further grow. The spiral bands also contribute more heat and moisture to the center enhancing the storm. When the winds speeds reach sufficient speeds, it can no longer reach the center, resulting in relatively calm area (storm eye). The tropical cyclone has reached 'maturity' and can travel of thousands of kilometers. The final stage, the decay can be caused by multiple factors. Increase in windshear, movement of dry air into the center or disruption in the outflow. However there are usually two factor that cause the decay: travelling over cool water and making landfall. Cool water temperature prevent the evaporation and heat convergence, stopping the ability for the TC to sustain itself. As the TC makes landfall, it loses its water source. As a consequence, the storm starts to decrease in intensity.

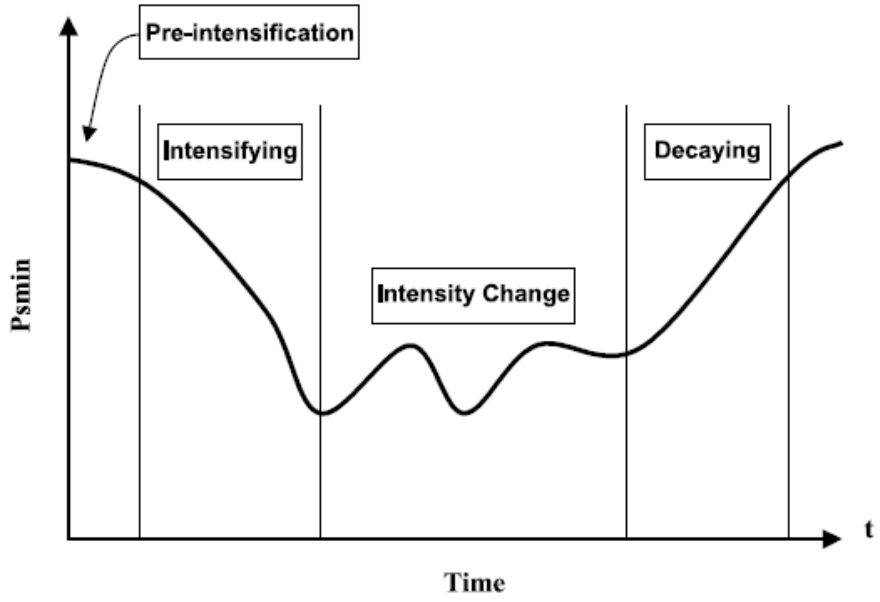


Figure 2.2: Lifecylce of a TC. Adapted from:Wang and Wu (2004)

2.2. Storm surge

Storm surges can be defined as the increase in sea surface elevation due to a TC moving towards the coast. The significant increase of sea surface elevation can cause devastating coastal floods with many fatalities. Multiple factors determine the height of the storm surge. The surge is generated by atmospheric forces like, drag of the wind on the sea surface and variations in the atmospheric pressures. Not only the intensity of the TC but also, the topography, bathymetry, forward speed, central pressure and angle of approach are influencing the height of the surge (Ellis and Sherman, 2015). The storm surge represents the total marine inundation associated with the storm (Nott, 2015). Other components that can be distinguished are: The tides, wave setup, wave actions and wave run-up. The contribution of waves and tides towards the sea level rise, is discussed in detail in sections 2.3 and 2.4.

2.2.1. Storm surge processes

Harris (1964) lists five main processes that are responsible for the storm surge generation.

- The **pressure effect** causes the sea level to rise due to low atmospheric pressure. The relation between pressure and sea level rise is 11 millibar drop in atmospheric pressures will cause approximately 1cm of sea level rise.

- **Direct wind effect (wind setup)**. The surface winds generate surface currents with a 45°angle. This deviation from the wind angle is caused by the Coriolis effect. The Coriolis effects forces the flow to the right of their motion in the Northern Hemisphere and to the left in the Southern Hemisphere. The same effect occurs deeper below the sea surface but with lower speeds, creating a spiral effect. As a consequence, the resulting net water transport is perpendicular to the wind direction. This effect is called the Ekman spiral. The same theory implies that for shallow water, the net transport must be parallel to the wind direction (no spiral effect). However, Water flow near the coast is parallel to the depth contours. The main effect that causes the sea level rise is called wind setup. Wind blowing on a lake has the tendency to drop water levels at upwind shore and increase water level at the downwind shore. Due to the Ekman spiral effects, the wind set-up is proportional to the water depth. The same effect holds for coastal areas and estuaries. The shallower the water, the higher the wind setup effect will be. The theoretical formulation to calculate the wind setup can be found in appendix A.1.

- **Coriolis effect** (Earth's rotation). Surge effects can be further amplified when the currents are perpendicular to the shoreline due to the Coriolis effect.

- **Wave effect**. The wind generated waves that follow the same direction of the wind (windwaves). The effect of wind waves are minimal at deeper water. However, towards the coast where the water is shallower, the wave will increase in steepness and eventually break. The breaking waves transport significant amounts of water towards the coast (wave run-up and overtopping).

- **Rainfall effect**. Finally, the extreme precipitation during a TC cause surface run-off towards rives and estuaries, increasing the water level even more.

Figure 2.3 shows the different components of the storm surge and their increasing effect towards on the water level towards the shoaling coast. Especially the wind setup causes a significant increase in the water level towards the shoaling coast line. A fairly recent study by Irish et al. (2008) investigates the influence of TC size and other parameters on the peak surge height. With the use of numerical simulations it can be concluded that the surge heights are mainly influenced by the storm intensity, bottom slope and storm size. The results indicated that the storm size plays a key role in the surge generation in coastal areas, especially in case of a severe storm with shallow slopes.

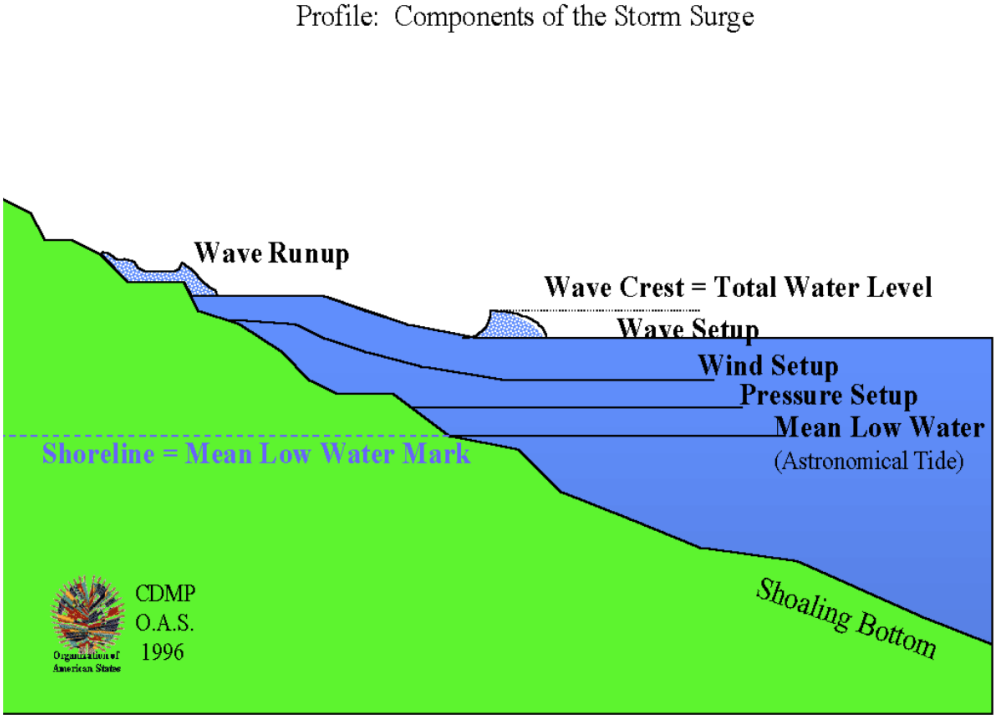


Figure 2.3: Storm surge profile. Adapted from OAS (1996)

2.3. Waves

Wind generated gravity waves are the major supplier of energy to the coastal system. They form in open oceans by wind and have in general an irregular character. Therefore, they are called irregular or random. The short term variations can be best described in a statistical way by taking average parameters (Holthuijsen, 2007). There are two methods the characterise these irregular wave records, by direct analysis or spectral energy density analysis. Both these methods result in a wave height parameter. The wave height can be expressed in multiple forms i.e. the mean wave height, significant wave height and root mean square height. For practical purposes, the significant wave height is used most often and is defined as the average wave height of the highest 1/3 of the waves in the sample set.

$$H_{1/3} = H_{m0} = \frac{1}{N/3} \sum_{j=1}^{N/3} H_j \tag{2.1}$$

Ocean waves are generated by local wind fields. At the generation area, the new waves are steep and short-crested and have different travel directions and travel speeds. Location dependent parameters that influence the wave characteristics are the wind field, fetch and local water depth. The Navier-Stokes and continuity equations are used to describe the motion of waves. Difficulties arise when trying to solve these equations. To overcome this problem, the non-linearity's are neglected to derive the linear wave theory. With the linear wave theory it is possible to describe a variety of wave characteristics like: wave speeds, height, velocity, length, period etc in shallow, deep or transitional water depths. According the linear wave theory, the propagation speed of a wave is dependent on the water depth i.e. the shallower the water depth the slower the propagation speed is. Sea waves (short and irregular) vary considerably from swell waves (long and fairly regular). This can be explained by the fact that wave field disperses since the different harmonics travel at different speeds depending their frequency. This phenomenon is called frequency dispersion. Furthermore, the wave fields spreads out due to different directions of propagation (directional dispersion). As a result, only a long, fairly regular

wave field remains (swell). More detailed information and theoretical background of the linear wave theory can be found in appendix A.2.

2.3.1. Wave transformations

Wave transformation occurs when waves propagate from deep into shallow water depths. Different processes causes the wave height, length and direction changes until the wave finally breaks and its energy is dissipated. The processes that causes these wave transformations are: shoaling, refraction and wave breaking.

- **Shoaling.** The waves slow down due to the decreasing water depth. As a consequence the waves increase in height. The wave height increase due to shoaling is limited by dissipation due to wave breaking.
- **Refraction.** This phenomenon occurs when waves approach underwater contours. The wave crests in the deeper parts will propagate faster than in the shallower parts. The wave crest turns towards the depth contours
- **Wave setup** During a TC, waves generated by storm winds are propagating towards the coast. These wave not only carry energy towards the coast but also momentum. When the waves reach coastal areas they can induce wave setup. Wave setup is defined as the increase in water level above the still water level due to momentum transfer by waves that are in the surf zone. This setup is primarily present in and near the coastal surf zone due to the shallow water depth. A more detailed theoretical background on wave setup can be found in appendix A.3.
- **Wave breaking.** Wave breaking occurs when a wave becomes to steep. The bed slope and bottom friction both have a influence on the breaking process. Depending the slope the waves break in different types (collapsing, surging, plunging and spilling breakers).

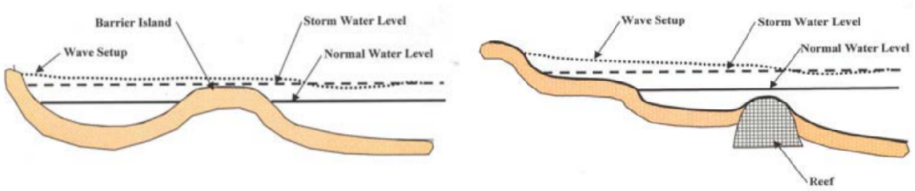


Figure 2.4: Example of beach profiles that can cause wave setup in front of the coast. Adapted from Dean et al. (2005)

2.4. Tides

Tides are the rise and fall of the sea water level caused by so called tide generating forces. The tide generating forces are caused by the gravitational pull of the moon and the sun on the water in the oceans (Bosboom and Stive, 2015). Tides can be best observed along the coast and are either diurnal or semi-diurnal. At most places around the world, tides are semi-diurnal, which means that the tidal period is approximately half a day. A diurnal tides has a period of approximately 1 day, which means that one high and one low water can be observed each day. The tidal range is depending on the location around the world. In this section the theory behind the generation of tides is explained. Furthermore, the tidal situation in the South Chinese Sea is analysed.

2.4.1. Generation of the tide

The first theory on the generation of tides was explained by Isaac Newton and is called the equilibrium theory of tides. In this theory, Newton assumed that the tide-generating forces cause an instant response of the ocean waters. The theory furthermore assumed, that there is no friction and the earth is entirely covered with water. To explain the tide generation theory, the approach from Bosboom and Stive (2015) is used. The tide-generating forces find their origin in the gravitation pull of the moon and sun on the ocean water bodies. The gravitation pull can be best described as the attraction that the earth exerts on an object. For the tidal generation the gravitational pull caused by the moon and sun are of importance. The moon and earth move in a circular orbit around a common center of gravity. The same holds for the sun and earth system. The revolving motion around the common centers of gravity implies that the attraction forces act as centripetal forces. A centripetal force is a force that acts on an object to make it follow a circular path. The centripetal force maintains the motion of the earth around the center of gravity of the earth-sun or earth-moon system. The gravitational pull between two bodies is directly proportional to the product of the masses of the bodies and inversely proportional to the square of the distance between them. The gravitational pull of the sun is approximately 2 times larges than that of the moon. Although the gravitational pull of the sun is larger than the moon, the sun only accounts for approximately 30 % of the tidal amplitudes. This can be explained by the fact that the tide is generated by a different effect, namely differential pull. The differential pull is the difference between the gravitational pull on ocean water masses that are located a different distances from the moon and the sun. For example the gravitational pull directly under the moon is bigger than the gravitational pull on the opposites side of the planet (greater distance from the moon). The differential pull can be calculated for all locations of the earth. Since the differential pull is responsible for the tidal generation, it is referred as the tidal force.

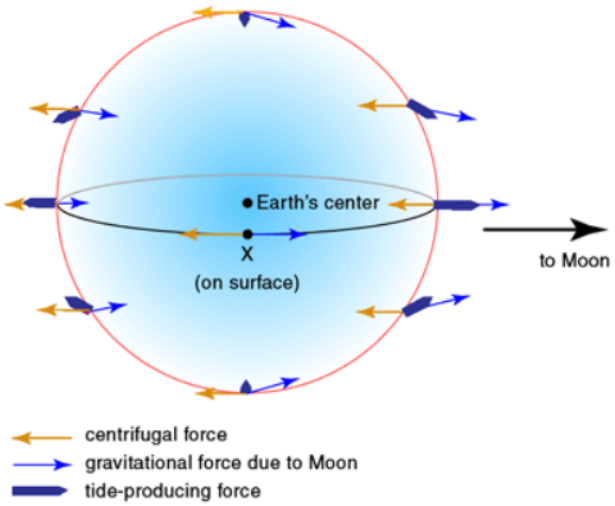


Figure 2.5: Horizontal component tidal force.

The effect of the horizontal (tangential) forces is shifting the water to the side of the earth facing the

sun or moon and to the opposites side of the earth (tidal bulges). The rotation of the earth around the axis shown in figure 3.4 forces the earth to rotate underneath the tidal bulges, producing two low and two high waters (semi-diurnal tide) on earth everyday (on the equator). We have seen that both the moon and sun are responsible for the tidal motion on the earth. The rotation of the earth around the sun and the moon around the earth with different periods implies that the differential pull at a location on earth changes over time. Tides are reinforced when the moon,sun and earth are in one line (spring tide). One the other hand neap tide occurs when the sun and the moon are 90 degrees out of phase. In the previous section it was assumed that the sun and moon are directly above the equator of the earth. However, in real life the orbits of the moon and sun are not in the equatorial plane all the time. The theory behind this is not in the scope of the thesis, but it implies that at the equator the daily inequality is zero and it increases with the latitude. At higher latitudes, the daily inequality becomes so large to there is only one high and low water per day (diurnal tide). Another effect that influences the tides on the earth is seasonal varying declination of the sun. This effect is responsible for the seasonal variability in the height of the tides. The influence of the moon, sun, daily inequality and seasonal effects on the tides can be described by tidal constituents. Three main categories are used for the tidal constituents: the semi-diurnal constituents, diurnal and long period constituents. The constituents all have their own equilibrium amplitude and period.

2.4.2. Propagation of the tide

In the previous section the background of the tidal theory was given. However, this theory was made under the assumption that the entire earth is covered with water without any land masses. The in real life existing land masses act as an barrier for the water. The water cannot move through the land masses but rather along the land masses. The propagation speed of a tidal wave can be calculated with the dispersion relation ship. A tidal wave is considered a long wave since the wave length is much bigger than the water depth (L»h or h/L<1/20) Since the tidal wave is a long wave, the relationship reduces to equation 2.2.

$$c = \sqrt{gh} \tag{2.2}$$

where:

c = wave propagation speed	$[m/s]$
g = gravitation acceleration	$[m/s^2]$
h = water depth	$[m]$

Equation 2.2 shows that the wave propagation speed is dependent on the water depth. For a shallow water depth the propagation speed is also lower. The wave propagation equation is derived from simplified shallow water equations and can be used for tide propagation in open oceans. It neglects friction, advection, diffusion and short wave effects. For deep oceans, a small tidal level and current velocities are observed. In shallower water the tidal amplitude is large and current velocities are fast. Friction, shapes and depth of the ocean all influence the propagation of the tides. The tidal motion is also influence by the Coriolis effect. The movement of the tides are def elected by the Coriolis effect and land masses, as a consequence rotary movements are formed in ocean basins, bays and seas. The movement of these systems is counter-clockwise in the Northern Hemisphere and opposite and the Southern Hemisphere. In those so-called amphidromic systems, the tidal waves propagates around a node with an anti node rotating at the basin edge. The node is the point that has no vertical displacement, while the anti nodes have maximum vertical displacement.

2.4.3. Tidal Prediction

Since the tides are generated by astronomical forces, the water level elevation due to the pure tide signal can be predicted accurate. This does not include the influence of wind and waves on the water level. The water level elevation due to the astronomical tides can be described with formula 2.3. The tidal constituents i.e. the phase angle and amplitude are location specific and can be determined by observations (Satellite or physical observations. When the astronomical tides are known for a certain location it can be used to predict the tide over a long time. Today, the tidal constituents are known for all oceanic waters. Although being it with varying resolutions.

$$\eta(t) = a_0 + \sum_{n=1}^{N} a_n cos(\omega_n t - \alpha_n) \tag{2.3}$$

where:

$\eta(t)$ = tidal level	$[m]$
a_0 = mean level	$[m]$
a_n = amplitude of astronomical component number n	$[m]$
ω_n = angular velocity of astronomical component number n	$[1/hr]$
α_n = phase angle of astronomical component number n	$[-]$

The tidal data required for the study can be collected by means of using tide models. Today multiple solutions exist that provide tidal levels and constituents for oceanic waters all around the world. The software contains tidal models that include, bathymetry/depth grid (m), elevations (m) and transport m^2/s. It is capable of predicting tides at given times and locations, and also features historical tidal levels. Additionally, the software allows the used to extract the harmonic constants of the tide that is required as input in the hydrodynamic model for the boundary condition forcing. There are different tidal models that provide both worldwide and regional solutions. For the storm surge simulation the focus lies on the regional tidal models. Three different tidal models are tested during the hydroynamic flow model setup. These tidal models are the TPXO 8 and TPXO 9 (regional South Chinese Sea models) and the FES2014 global model. These models are discussed in more detail in B.1.

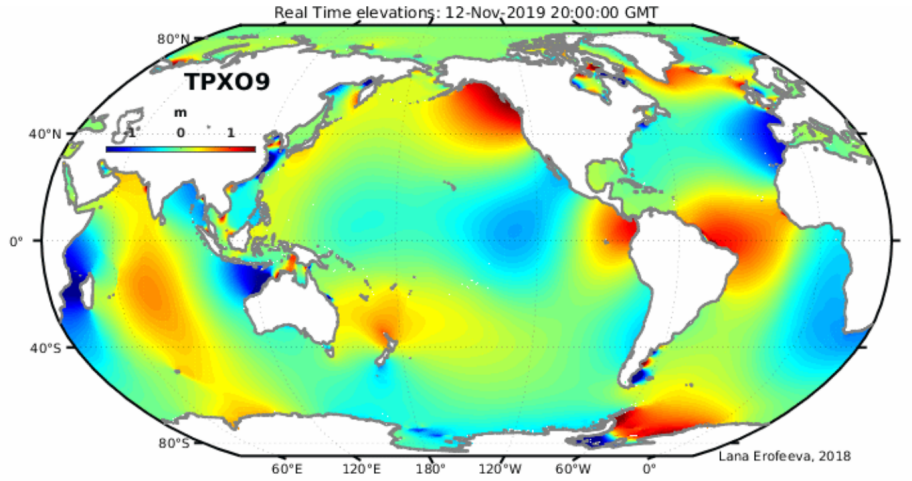


Figure 2.6: Example of global tide model (TPX09). Adapted from:Egbert and Erofeeva (2002)

The second source of tidal data comes from tidal stations around the waters of Hong Kong and Macau. This historical water level data is needed to validate the model. By comparing the simulated water levels of a hydrodynamic simulation with the historical water levels, one can determine if the model is capable of producing accurate results. The tidal stations are operated by local authorities and provide historical tidal data. For validation purposes an effort was made to collect the data of as many station as possible. Data from three different tidal stations in Hong Kong was obtained. On the other side for Macau no tidal data could be collected.

2.5. Wind model

Today, multiple parametric wind models exist that can be used as forcing input in hydraulic models that estimate surges and wave actions. The wind field, that is required as input for the hydraulic model can be estimated by different methods. One option is to reanalyse historical TC data sets. However, these methods are often to coarse to represent the wind field accurately, leading to underestimations of maximum winds (Ruiz-Salcines et al., 2019). An alternative is to reconstruct the wind field based on a limited set of parameters. These parametric wind models are often used due their efficiency, simplicity and low computational cost. The low computational cost makes these parametric models especially usefull for studies that require the modelling of a large number of TC's. Parametric wind models use a wind profile to derive the wind distribution around the centre of a storm Dima and Desflots (2010). The choice of the wind profile is important for computing realistic wind speed in tropical cyclones.

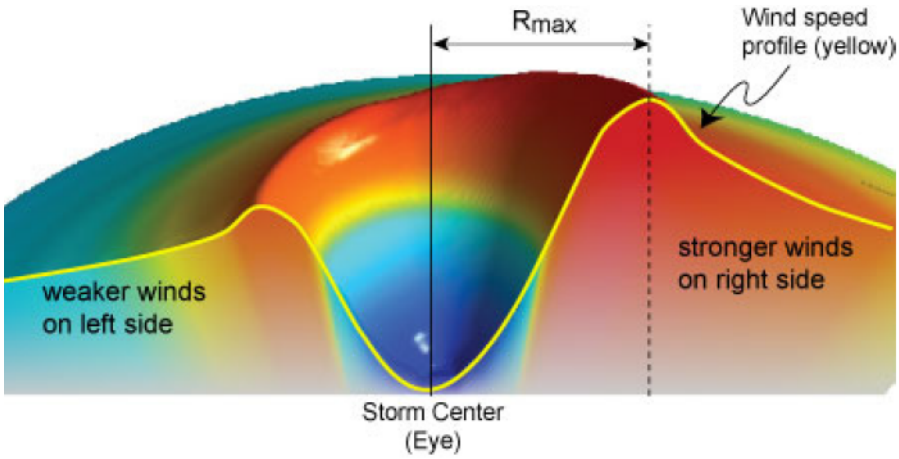


Figure 2.7: schematic cross section of Tropical cyclone wind field with wind speed profile in Northern Hemisphere. Adapted from: Dima and Desflots (2010)

The basis of most parametric wind models is the gradient wind equation. This equation describes the balance between three forces in tropical cyclones: pressure gradient forces, centrifugal forces and Coriolis forces. The gradient wind equation is derived from the continuity and momentum equation, more information on the derivation can be found in Holton and Hakim (2012).

$$V_g = -\frac{fR}{2} \pm \left(\frac{f^2R^2}{4} + fRV_{geo}\right)^{1/2} \tag{2.4}$$

Where:
V_g = Gradient wind speed
R = Radius of curvature
f = Coriolis forces
V_{geo} = Geostrophic wind

The Holland (1980) wind model is a modification to the gradient wind equation. Holland improved the work by Schloemer (1954) who suggested that the radial surface pressure profile can be approximated by an rectangular hyperbola which can be differentiated to derive wind speeds. Holland (1980) improved this work by describing the rectangular hyperbola for the pressure field in a different form with two scaling parameters A and B.

$$p(r) = p_c + (p_n - p_c)\exp - \left(\frac{A}{r^B}\right) \tag{2.5}$$

Where:

$p(r)$ = Surface pressure at distance r from storm center

p_c = Central pressure

p_n = Ambient pressure

A = Location parameter

B = Holland pressure profile parameter

By taking equations 2.5 and using the gradient wind equation, the wind profile can be described by equation 2.6, where V_g represents the gradient wind at radius r, f is the Coriolis parameter and ρ the air density.

$$V_g = [\frac{AB(p_n - p_c)exp(-A/r^b)}{\rho r^B} + \frac{r^2 f^2}{4}]^{1/2} - \frac{rf}{2} \tag{2.6}$$

According to Holland (1980), the Coriolis force is small compared to the pressure gradient and centrifugal forces in the region of maximum winds. The winds in these regions are described by equation 2.7 and is called the cyclostrophic wind speed.

$$V_c = [\frac{AB(p_n - p_c)exp(-A/r^b)}{\rho r^B}]^{1/2} \tag{2.7}$$

By setting the derivative of V_c to 0, the radius of maximum winds (RMW) is obtained. The RMW value is dependent on the scaling parameters A and B, it is independent of the central and ambient pressure

$$R_w = A^{1/B} \tag{2.8}$$

By substituting equation 2.8 into 2.7, the maximum wind speed is obtained:

$$V_m = C(p_n - p_c)^{1/2} \tag{2.9}$$

where:

$$C = (B/\rho e)^{1/2} \tag{2.10}$$

The symbol e in equation 2.10 represents the base of natural logarithms. The empirically derived parameter C has been used extensively for estimating the RMW in the past. The approach from Holland allows to derive Tropical cyclones pressure and wind speed profiles from pressures observation taken during TC's. The parameter B defines the shape of the pressure profile and A determines the location relative to the origin. The model requires the parameters A and B to be derived from observed tropical cyclones to let the wind and pressure profile fit the real tropical cyclone. Figure 2.8 shows the influence of parameter B on the wind and pressure profile. It is clearly shown that higher values for B will generate sharper wind speed profiles with higher peak wind speeds.

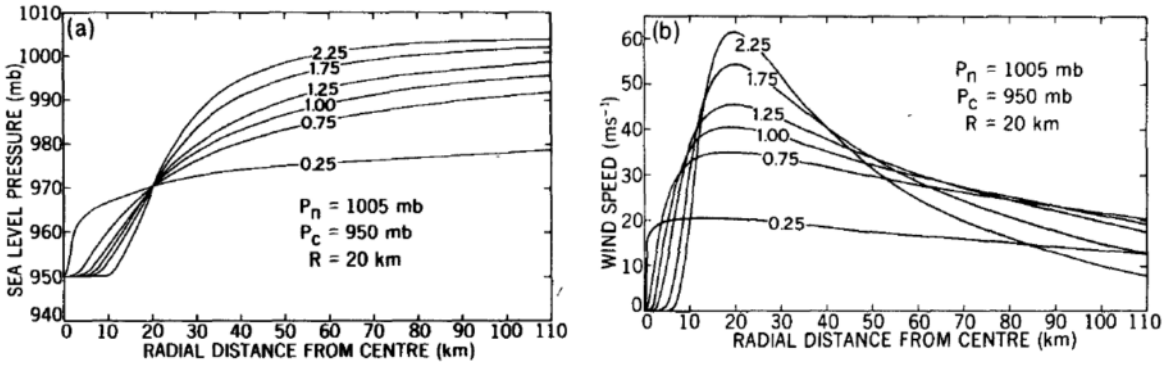


Figure 2.8: Effect of parameter B on the radial wind and pressure field of Tropical cyclones. Adapted from: Holland (1980)

Equation 2.8 shows, that the RMW is independent from the air pressure and defined entirely by the scaling parameters A and B. Technological developments of the past decades have improved the predictions and observations of TC's significantly. This has led to new estimations of RMW directly from observations or new methods to estimate the RMW. If the RMW is known, parameters A and B can now be rewritten and expressed in a different form.

$$A = R_w^B \tag{2.11}$$

$$B = \frac{\rho e V_{max}^2}{p_n - p_c} \tag{2.12}$$

This shows that A and B can be determined easily if the RMW is known. Today some TC agencies provide values for RMW in their track data. Other investigations suggested the estimation of the RMW based of the eye pressure or radius of 50 knot winds. By substituting equations 2.11 and 2.12 in 2.6 the gradient wind equation can be expressed in terms of the RMW.

$$V_g(r) = \sqrt{(R_w/r)^B V_{max}^2 exp(1 - (R_w/r)^B) + r^2 f^2/4)} - \frac{rf}{2} \tag{2.13}$$

Over the recent years, multiple studies into the estimation of parameter B and RMW have been conducted to improve the wind profile derived from the parametric model. Lajoie and Walsh (2008) suggests to estimate the RMW from satellite cloud data by analysing the radius of the storm eye. The study by Vickery and Wadhera (2008) proposes the estimate the RMW from the eye pressure. The same study also looks into a new determination of the shape parameter B.

Veltcheva and Kawai (2002) proposes a new method to estimate the wind and pressure distribution. The structure of the tropical cyclone changes when approaching land due to an increase in friction of the air system with the land surface. This distortion effect is not captured in the original pressure distribution equation (equation 2.5) where the pressures field is distributed symmetrically with the radial distance form the TC centre. Veltcheva and Kawai (2002) includes the distortion effect into the pressure field equation. The TC pressure field is now dependent on the angle between the observation point and the direction of the TC movement:

$$p(r, \theta) = p_c + (p_n - p_c) \exp - \left(\frac{r(\theta)}{r} \right) \tag{2.14}$$

To include the effect of pressure distortion for the determination of the wind field, the gradient wind equation is modified.

$$U_g = V_c(\sqrt{\gamma^2 + 1} - \gamma) \tag{2.15}$$

$$\gamma = 0.5(\frac{FS * sin\theta}{V_c} + \frac{Vc}{Vgeo}) \tag{2.16}$$

Equation 2.16 is used to calculate the parameter γ. This parameter is dependent on the forward speed of the TC, cyclostrophic and geostrophic wind speeds and the parameter θ. Where θ is the angle between the observation point and direction of typhoon movement.The cyclostrophic and the geostrophic wind speeds at distance r form the TC centre are given by equation 2.17 and 2.18.

$$V_c = \sqrt{\frac{r}{\rho} \frac{dp}{dr}} \tag{2.17}$$

$$V_{geo} = \frac{1}{\rho f} \frac{dp}{dr} \tag{2.18}$$

The forward direction of the TC can be calculated by taking the Four-quadrant inverse tangent function (atan2) of the latitude and longitude difference between time steps. The output of this function is given in radian values between $-\pi$ and $+\pi$. These values can than be transformed to values in

degrees and to the correct reference plane. The output of equation 2.19 also contains negative values in degrees. To overcome this problem 360 is added to all negative values, as a results all FD values now range between 0 and 360° indicating the forward direction of the TC.

$$FD(t) = -(atan2((lat(t) - Lat(t-1)), Long(t) - Long(t-1)) * cosd(Lat(t))) * 180/pi) + 90; \tag{2.19}$$

$$FS(t) = \sqrt{\frac{(Lat(t) - Lat(t-1)) * R_{earth} * \pi/180)^2 + ((Lon(t) - Lon(t-1)) * cosd(Lat(t)) * R_{earth} * \pi/180)^2}{((day(t) - day(t-1)) * 24 * 3600(hour(t) - hour(t-1)) * 3600}} \tag{2.20}$$

The comparison of this method to the older methods, that don't include the distortion of the pressure field showed that a better agreement of the wind was obtained for the new method. Furthermore, the corrected wind and pressure field were tested on a storm surge model. The results showed that the new method improves the estimation of the storm surge in the model. The original Holland model and the discussed modifications/ improvements allow, the pressure and wind speeds to be calculated. By varying the equations for different distance from the center of the storm, the entire radial profile of the storm can be computed. This radial wind profile can than be adapted on a moving spiderweb grid, which can be used in the hydrodynamic model as TC wind forcing.

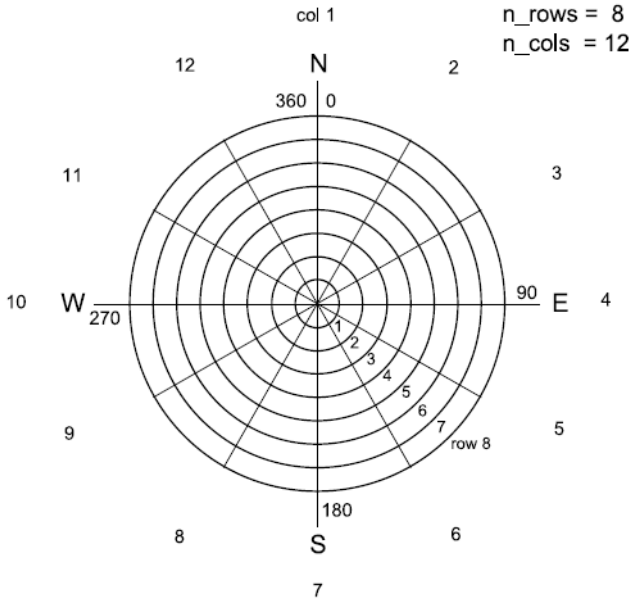


Figure 2.9: Spiderweb grid. adapted from DELFT3D flow manual.

The spiderweb grid allows for space varying wind velocity, direction and air pressure input in the hydrodynamic model. The grid cell sizes are dependent on the number of radial bins, tangential bins and total radius of the grid. The larger the number of radial and tangential bins, the higher the resolution of the grid cells. The location of the eye of the spiderweb is given by the latitude and longitude. The spiderweb grid can be generated for different time steps moving along with the TC.

2.6. Flood models

As discussed in chapter 1, floods have caused significant damages and loss of life all over the world in the past. Due to the existing risk of floods, many different models have been developed to understand, predict and assess flood events. The hydrodynamic processes that exist in coastal ares are generated by multiple drivers. The drivers of these processes are currents, tides, waves, surges etc. All these drivers are forces that drive the fluid motion in the coastal area. For the modelling of these coastal hydrodynamic processes, different types of models can be used depending the topic of the investigation. Four main types of models can be identified, namely the physical, empirical (data driven), hydrodynamic, and conceptual models (Teng et al., 2017).

2.6.1. Physical models

Physical and laboratory experiments are still used today to better understand the effects of tides, currents and waves on hydraulic structures like beaches, dunes, dikes, harbors, locks and other flood defences. Conducting physical or laboratory experiments allows for better understanding on the physics of coastal processes. Furthermore, the use of physical models is well establishes method for testing and designing of hydraulic structures. The possibility to test how a hydraulic structure will perform before it is constructed gives engineers the possibly to gain knowledge on its performance. For example wave flumes are used to test the proposed design of flood defences. By scaling down the design, one can reproduce the 'real world' in a controlled environment. For example, by creating a scale model of a proposed dike improvement in a wave flume, it is possible to investigate the dynamic wave actions, run-up and overtopping. The data collected during the tests can then be used to validate, or improve the design. Physical models are often not suited for studies into coastal floods and therefore won't be considered in this thesis.

2.6.2. Hydrodynamic models

The previously discussed coastal processes can all be described by mathematical expressions. However, these processes are often complex and vary in time due to the changing conditions. These processes are also space dependent. Changes in bathymetry, bottom roughness, and coastline shapes all have an effect on the fluid motion. These complex processes all influence each other. For example changes in the water level will also change the wave characteristics at that location. Traditionally these processes have been studied with physical and empirical models. However, the reproduction of these complex and unsteady processes often requires the development of expensive facilities and model representations. The development in computer science has made it possible to develop (often numerical) models that can simulate the coastal processes ranging from small to very large study areas. By using hydrodynamic models, engineers can gain insight in the interaction between the complex coastal processes, which can be used in the design, planning and disaster management for coastal areas. It must be noted that numerical models are also often used for the simulation of fluid and sediment transport in rivers and lakes. It must be noted that the use of hydrodynamic models is complex and the accuracy of the results are dependent on different factors like boundary conditions, model parameters and numerical methods. The development of these models is therefore a specialised task, and requires a good understanding of the physical processes that occur.

Today, a wide variety of numerical models exist that can be used for different applications. The hydrodynamic models can be classified in three classes (1D,2D and 3D).

- **1D models**. The most simple model is one-dimensional and can be used for flood plain flow in the down-valley direction. The governing equation for 1D channel flow can be derived form the mass and momentum equation between two cross sections. This yields the St. Venant or shallow water equations. The shallow water equations can be solved numerically with with use of boundary and initial conditions (Teng et al., 2017). Modelling software that uses these 1D equations are: MIKE11,ISIS and SOBEK.
- **2D models**. This model represents the flood plain in a two-dimensional field. This approach uses the depth average velocity obtained by integrating the Navier-Stokes equations over the flow depth. This method can be solved numerically to calculate predictions of water depth and the two components of the depth-averaged velocity. 2D models are often applied to flows with

large aerial extent compared to the depth and for large lateral variations in velocity. They are well suited for the computation of overbank flood flows in compound channels, tides, tsunamis or dam break (Asselman, 2009). Examples of 2D modelling software are: Mike 21, SOBEK-OF and Delft-3D.

- **3D models**. The most complex model is the 3D model which requires the three-dimensional Navier Stokes equations to be used. The modelling of vertical turbulence, vortices and spiral flow for large scale floods (Teng et al., 2017). However, due to the computational cost for the model, there is a trade off between the cell size, domain size and complexity of used schemes. The 3D models give as output the inundation extent, water depth and the u,v,w velocities for each cell. It is mostly used for coastal applications where 3D velocity profiles are important. Examples of 3D models are: Delft-3D and Delft-3D FM.

The Sea Lake and Overland Surge from Hurricanes (SLOSH) is a well known storm simple surge model for sea, lake and overland surges caused by hurricanes. The model was developed by the US National Weather Service. SLOSH is a 2D model that can estimate surge heights from hurricane data. The model is computationally efficient and fast but that comes at a cost of losing accuracy. SLOSH also lacks the possibility the include tides, waves, precipitation and rivers. Empirical wind profiles are used as input to compute wind speeds at the coastline and surge heights. The model is reasonably accurate but much simpler than for example Deflt3D and TELEMAC.

TELEMAC-MASCARET is an open source suite for the field of free surface flow. It includes packages for 1D, 2D and 3D flow computations. The 2D module uses a finite elements method to solve the St. Venant equations with a triangular mesh. The main fields of application are free surface maritime and river hydraulics like the sizing of port structures, effects of building immersible dikes or dredging, impact of construction works in rivers etc. The 3D module is mostly used for topics related to the marine environment like tides, currents, wind and air pressure. Overall, the TELEMAC package is validated and used frequently by professionals.

Delft3D is an open source 2D and 3D hydrodynamic model developed by Deltares in the Netherlands. It has many areas of applications for Tides and wind driven flow (storm surges), river flows, deep lakes and reservoirs and the simulation of tsunamis, hydraulic jumps and flood waves. The Delft3D suite consists of various modules that are integrated and linked with each other. The modules are divided in to D-flow, D-waves, D-water, D-morphology, D-particle tracking and D-sediment. Additionally, visualisation and other tools are provided with the package. The D-wave package uses a modified version of the SWAN model, which is developed by Delft University of Technology. The SWAN model is used for the computation of short-crested-wind-generated waves in coastal regions and inland waters. The entire Delft3D suite is well tested and validated. Currently, Deltares is developing the successor of Delft3D. The new Delft3D Flexible Mesh package can be used for hydrodynamic simulations on unstructured grids in 1D, 2D or 3D. The software is currently available for users, but is still in development. The new version has the same functionalities as it predecessor and has the same areas of application. The main progress in the flexible mesh version is the possibility to use grids that consist of triangles, quads, pentagons and hexagons besides the curvilinear grids of the older version. This addition provides improved modelling flexibility and ease in setting up and modifying grids. Furthermore, 1D and 2D grids can be combined and it features new user and graphical interfaces. The different Delft3D FM modules have been extensively tested, but the suite as a whole still needs testing and validation. Furthermore, the integration of the wave package still has some errors. A conceptual description and theory behind DELFT3D can be found in appendix A.4.

2.6.3. Empirical, data driven models

Empirical data driven models try to find a relationship between a set of inputs and outputs. Statistical models, are also included in this category. A difference compared to the physical and hydrodynamic models is that data driven models do no use any physical parameters or equations to solve the problem. An advantage of these models is that they are generally fast to run, making them interesting for real time forecasting. However, there performance is completely dependent on the quality and quantity of the data used to develop the model. For data driven models the required data is often obtained

from historical time series observations i.e. wave and tide buoys, weather stations etc. Over the past decades Bayesian Networks (BN) and Artificial Neural Networks (ANN) have become more popular in flood forecasting.

The BN is a probabilistic model that graphically represents the conditional dependence between a set of interdependent variables that characterise a (flood) event. Either discrete or continuous variables can be used in the network and the dependency between the variables are quantified by conditional probability functions.

Boutkhamouine et al. (2020) developed a model for river flood forecasting with a BN approach. The BN is used to estimate the discharge in a river basin. The network is capable of predicting the discharges in upstream rivers of the discharge in the downstream basin is known. The is very use full to estimate river discharge upstream with missing hydrological measurements. BN's are also used for other hydrological purposes like coastal risk analysis, compound flood hazards and flood predictions based on atmospheric variables among others (Couasnon et al., 2018), (Jäger et al., 2018). Finally, we consider the study by Sebastian et al. (2017) in which a BN is developed that considers five TC characteristics at landfall: wind speed, angle of approach, landfall location, radius of maximum winds and forward speed. Observations of over 300 TC's were used to create a synthetic database with 100,000 synthetic storm events for the Gulf of Mexico. These synthetic storms were then used in an empirical wind model to simulate storm surges. With the results, probable values for the peak surge height can be determined. This study shows close resemblance with this thesis research and can be used to determine flood risk.

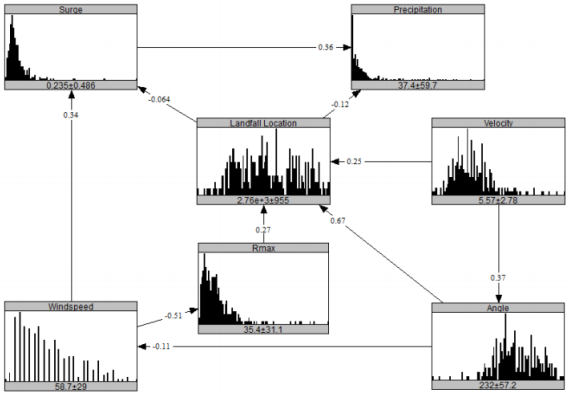


Figure 2.10: Bayesian network schematic for storm surge prediction in the Gulf of Mexico. Adapted from: Sebastian et al. (2017)

The second data driven model that we consider are the artificial neural networks. By feeding input and output data in to the model, the ANN learns the relationship between the in and output, which then can be used for forecasting. The ANN is completely data driven and does not use any physical equations for forecasting.

Over the years, the development of neural networks for the prediction of storm surges has become more popular. Most storm surge prediction neural networks use a set meteorological and/or hydrological input/output parameters for a specified time interval. The output neuron or neurons often represent either the water level or tidal elevation for a specific location. Lee (2006) developed a neural network based on four input parameters at a measurement location in Taiwan: Wind speed, Wind direction, eye pressures and harmonic tidal level in 3 hourly intervals. The network only contains one output parameter, namely the tidal level (water level elevation due to tides and storm). Extensive validations showed, that the network was able to make acceptable water level predictions based on the four input parameters only. Sztobryn (2003) followed a slightly different approach by assuming a tide less sea. This study, investigates two variations: one with continues data (time series) and a set of random 'storm sea level' samples. The results show that the network that used the continuous data set produces better results. Other examples of studies that develop NN capable of making flood prediction are: (Kim et al., 2019) and (French et al., 2017).

Bayesian and neural networks both share similarities and differences. Generally speaking the BN is easier to develop with less decisions about nodes, hidden layer etc. On the other hand, the visual representation of the network structure gives valuable insight on the conditional dependence between the variables. The ANN structure itself does not say anything. ANN generally are able to make accurate predictions if sufficient training data is available. Since we are using synthetic storm simulations during this thesis we are able to simulate many storms and create a substantial amount of data to train the neural network. Combining this with the fact that my predecessor student Guill (2020) focused on Bayesian networks, it is decided to develop an ANN for the storm surge predictions in Hong Kong and Macau. In the next section neural networks are treated in more detail.

2.7. Artificial Neural Networks

As stated before, artificial neural networks are mathematical models that can be modelled and trained for performing particular tasks based on available data. In the literature no universal clear definition of ANNs can be found. The original concept of neural networks was inspired from the conceptualisation of the human brain based on a network of interconnected cells (neurons) (Haykin, 1994). The first neural networks only had one hidden layer and were unable to solve complex problems. Over the years developments have led to the introduction of multiple hidden layers in the model to solve more complex problems. Additionally multiple training and learning algorithm's have been developed for more efficient learning. The development of an ANN can be described by three main phases:

1. Create training database
2. Create networks architecture
3. Train network

2.7.1. Conceptual & mathematical background

A neural network can be characterized by its structure. The structure represents the pattern of connection between nodes, method of determining the connection weights and the activation function. The categorization of ANNs can be based on the number of hidden layers and the direction of information flow and processing.

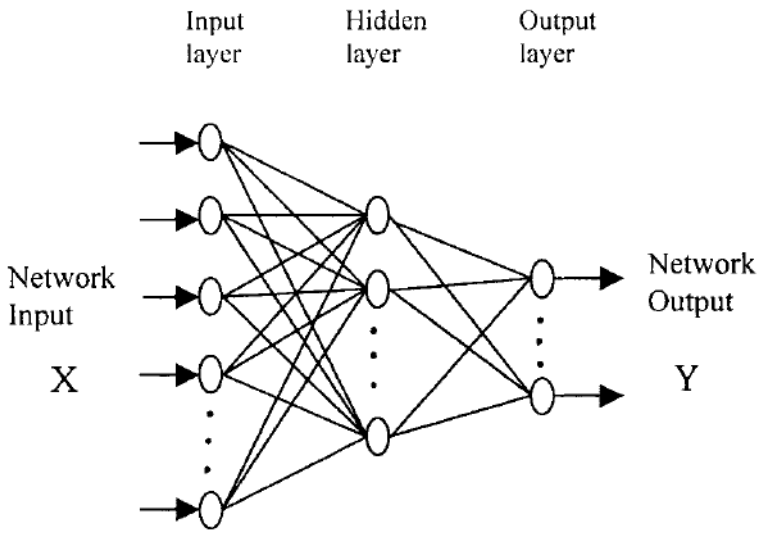


Figure 2.11: Structure of three layer feed forward ANN. Adapted from Govindaraju (2000)

Figure 2.11 shows an example of a simple feed forward neural network with a single hidden layer. In a feed forward network the data passes from the input layer to the output layer. The nodes in the layer

are connected to the nodes in the next layer but are no connected to the nodes in the same layer. Thus, in a feed forward network the output data is only dependent on the input is receives from previous layers and their corresponding weights. However, different structures on ANN exist that allow for information flow in both directions (recurrent networks). Sometimes connection between the nodes in the same layer can also be connected. The input layer is provides the data feed into the network. Each input node represents a different variable or parameter. The final layer (output layer) of the network contains the predicted model output. Although there are some rules of thumb to estimate the required number of nodes in the hidden layer, they are not set in stone. The number of nodes in the hidden layer are often determined by trial and error.

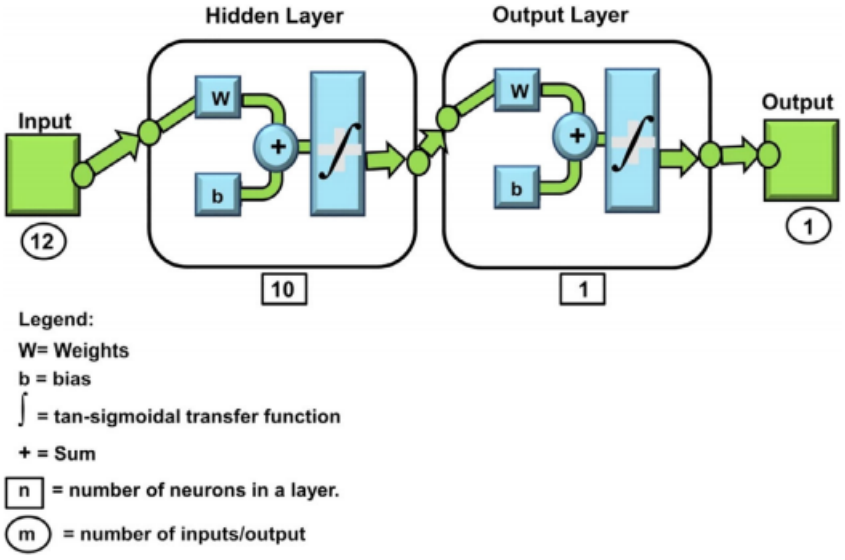


Figure 2.12: Neural Network Schematic. Adapted from Tabbussum and Dar (2020)

Figure 2.12 represents a schematic diagram of a neural network. Suppose we have an input vector $X = (x_1, x_i, x_n)$ that contains the input data that influence the system behaviour. The weight vector $W = (w_{1j}, w_{ij}, w_{nj})$ represents the connection between the nodes in different layers and assigns weights to the neurons. The operation within the network can be defined by the following equation (Govindaraju, 2000):

$$y_j = f\left(\mathbf{X} \cdot \mathbf{W}_j + b_j\right) \tag{2.21}$$

Where b_j is a threshold value (bias) and f an activation or transfer function that determines the response of a node to the input signal. The activation function are equations that determine the output of the network. The function holds for each neuron and determines whether it should be activated. Activation of a neuron should only happen if the input is relevant for the model's prediction. A range of different activation functions exist like the step and sigmoid function (equation 2.22).

$$f(t) = \frac{1}{1 + e^{-t}} \tag{2.22}$$

2.7.2. Training Algorithms

After setting up the structure of the ANN, the learning process can start. This process is used to find the optimal values of the weights W that minimise the error between the in and outputs. During training the network tries the minimise the global error or sum of squared difference error:

$$e = \frac{1}{N} \sum_{n=1}^{N} (y - t)^2 \tag{2.23}$$

Where N is total number of training patterns, e_i is the error for the training pattern n and t is the target. y is the output of the network. There are two types of training: supervised and unsupervised. The supervised training requires a large number of samples with input and output data. During the training process, the weights are adjusted and optimized for each nodes, to minimize the error function. The unsupervised learning process does not need a large number of training samples. Only input samples are provided. The network adapts its connection weight to cluster these input patterns into classes with similar properties. Over the years many different algorithms for training have been developed. For this thesis some widely used algorithms that are used in flood prediction networks are considered. The different training algorithms are discussed in more detail in appendix A.5.

2.8. Synthetic storm catalogue

The synthetic storm catalogue from (Bloemendaal et al., 2020) provides the ability to simulate much more storms than historically have been measured. The catalogue contains the track data of more than 200,000 unique storm events for the Western Pacific Basin. From the database, a number of storms will be selected that influence the water levels near Hong Kong and Macau. The track data of these storms will then be parametrized by means of a wind model and the resulting spiderweb grids will be used as the main forcing in the hydrodynamic model. The catalogue uses 38 years worth of track data from IBTrACS to generate 10000 years worth of TC tracks and intensity data. The sheer size of the database provides excellent opportunities to simulate a large number of storms in Delft3D and use the output data for the development of machine learning models like a neural network. The development of the synthetic catalogue itself can be divided into three phases:

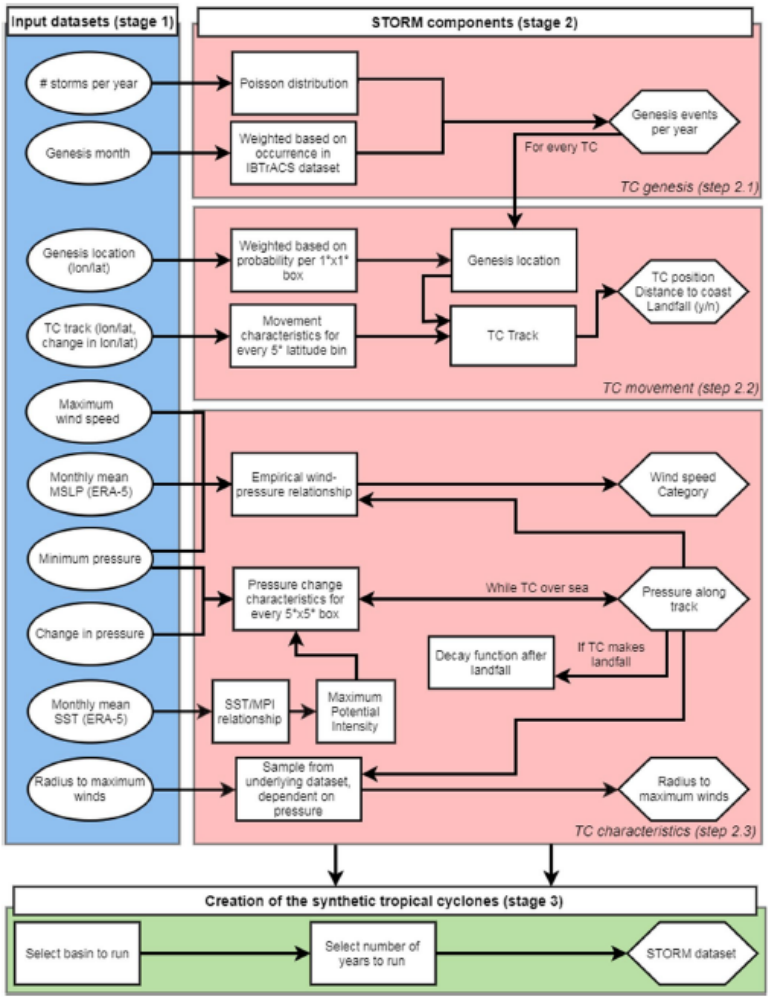


Figure 2.13: Flow chart generation synthetic TC database. Adapted from Bloemendaal et al. (2020)

1. **Data preparation and input**. During this phase TC track data ranging from 1980 to 2018 is extracted from the IBTrACS database for each basin. No data older than 1980 is considered due to comply with modern measurement techniques. Furthermore the minimal threshold is 18 m/s to be selected. Depending on the time intervals of the IBTrACS measurements, the storm data in linearly interpolated to 3 hourly values. The modelling of the synthetic storms also requires additional data like monthly mean sea level pressure (MSLP) and sea-surface temperatures.

2. **Fitting distributions and relationships**. The second stage of the synthetic database development can be divided into three sub-stages namely: TC genesis, TC movement and TC characteristics. During the TC genesis stage, the number of events per year are simulated using a Poisson distribution. Each generated event is randomly assigned to a genesis month.The next step in the development is the generation of the TC genesis location. The genesis locations for the synthetic TC's are generated from the weighted genesis location in the IBTrACS database. For the movement of the synthetic TC, again track data from the IBTrACS database is used and grouped together in 5°latitude sections per basin. For every bin a set of regression formulas are fit to the data. The final step is to generate the characteristics of the TC (eye pressure, maximum wind speed, radius to maximum winds etc.). First, the TC characteristics are assigned along each track. The conversion between the maximum 10-minute averages wind speed and minimum pressure is modelled with the empirical wind-pressure relationship.

$$V_t = a\,(P_{env} - P_t)^b \tag{2.24}$$

In which V_t and P_t are the 10-minute average wind speed and the minimum sea level pressure at time step t. a and b are variables that are different for every month and basin. The data is fitted to the equation using least squares method. The maximum storm intensity is restrained by the so-called maximum potential intensity, to make sure that the synthetic storms don't grow to intense. The pressure drop can then be calculated with the following formula. Where the coefficient A, B, C are estimated using least squares methods.

$$P_{\text{env}} - P = A + Be^{C(STT - T_0)}, \quad T_0 = 30.0°\text{C} \tag{2.25}$$

Finally, the MPI can be determined by subtracting the maximum pressure drop from P_env. The changes in pressure along the TC track are calculated with equation 2. To do so, the pressure changes are extracted from IBTrACS and fitted to the equation. The C coefficients are deduced for every month and basin.

$$\Delta P_t = c_0 + c_1 \Delta P_{t-1} + c_2 e^{-c_3 X}, \quad c_2 > 0, X = \max\{0, P_t - MPI\} \tag{2.26}$$

$$\Delta P_{0.01} \leq \Delta P_t + \varepsilon_p \leq \Delta P_{0.99}, \quad \varepsilon_p \sim N\left(\mu_p, \sigma_{e_p}\right) \tag{2.27}$$

$$P_t = P_{t-1} + \Delta P_t + \varepsilon_p \tag{2.28}$$

Another aspect that is considered during the synthetic storm generation is the fact that the intensity of the storm will decrease after making landfall. To ensure that this effect is captured in the database an addition is made that assumes that the TC intensity decreases as a function of the time and distance it has covered while being over land. Finally, the radius of maximum winds is derived for each time step. Based on historical observations, it seems that intense TC's in general have a smaller RMW. For week TC's a larger RMW is observed. To capture this effect, the derivations of the RMW is divided into three subsets for different storm intensities.

3. **Creating synthetic TC's**. During the third and final stage the 10.000 years of synthetic storms are generated for each basin with the STORM model. This model is consist of series of python programs and follows the components described in the previous two stages.

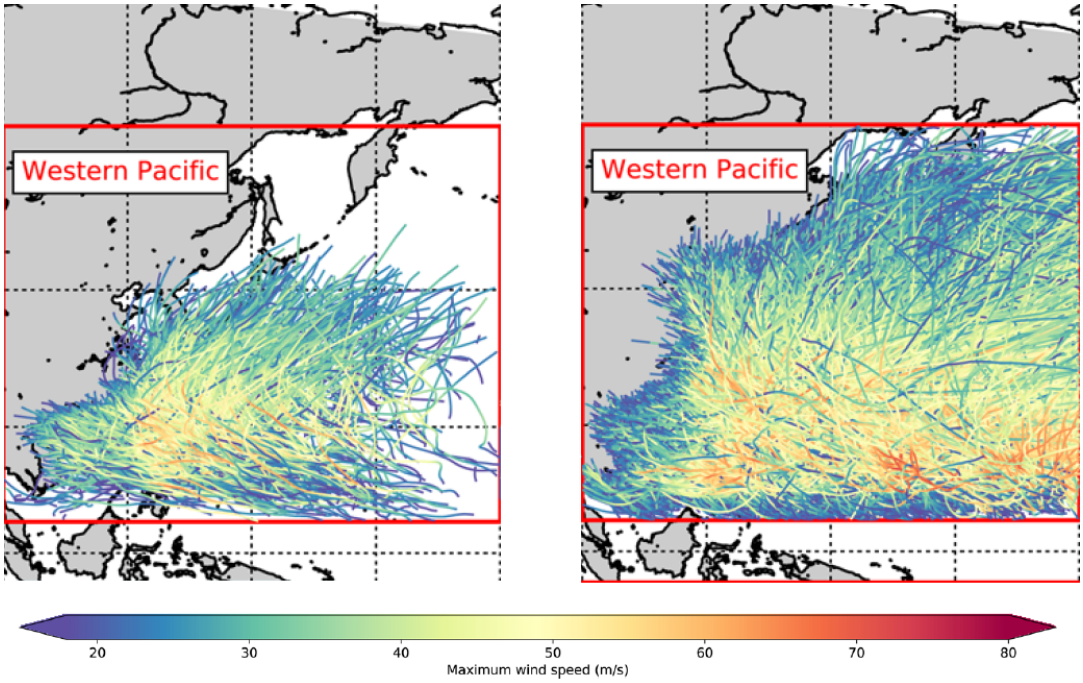


Figure 2.14: Left side: 38 years of TC tracks IBTrACS WP. Right side: 1000 years of synthetic TC tracks WP . Adapted from Bloemendaal et al. (2020)

The final product is a catalogue for each TC basin consisting of 10 text files that all contain 1000 years worth of storms. The track data for each storm is listed in intervals of 3 hours. The columns in the file contain the essential storm track data, like latitude, longitude, wind speed, air pressure, RMW etc.

3

Delft3D Storm Surge Model

This chapter contains all aspects regarding the Delft3D storm surge model for Hong Kong and Macau. In the first section the model setup is discussed. The model setup will form the basis of the synthetic simulations. It is therefore important to reproduce the real world as accurately as possible in the model. In the second section the calibration and validation of the model is discussed. The calibration phase of the model is supported by appendix B in which the sensitivity of different model parameters are discussed. The tidal validation is conducted in two steps. First, only astronomical tide forcing of the model is validated and wind is not yet included in the model. Then the model is validated for simulating storm tide (simulation with astronomical tides and wind forcing). The storm surge validation is done for three historical TC's that have affected Hong Kong and Macau in the past.

3.1. Model setup

The model setup determines for a large part the performance during the simulations. For a correct model construction, many aspects must be considered like data input, grid creation, boundary conditions etc. This section discusses in detail the setup of the model.

3.1.1. Grid

The computational grid forms the basis of the numerical model and can be created in RGFGRID, a subprogram of the Delft3D suite. The key factors during the grid creation that should be considered are the grid extent and the grid size. They both influence the overall performance of the model and must be implemented correctly. Since the focus lies on storm surge simulations, the grid should cover a large area of the South Chinese Sea in order to model the different physical aspects correctly. Tropical Cyclones often have large a radius and travel large distances. To allow for a correct simulation of the TC's physics and gradual entering in the domain, a rectangular grid of approximately 500 by 700 km is used. The grid covers the pearl river delta and extends almost all the way up to Taiwan and the Philippines. The choice of grid cell size is dependent on the desired resolution for the model. In general, shallow coastal waters require smaller grid cells to accurately reproduce the steep changes in bathymetry. The accurate reproduction of the bathymetry in shallow water is needed for accurate storm surge predictions. In the open ocean, the changes in bathymetry are most often much more gradual and since the ocean is deep water the accuracy of the bathymetry is less important. Furthermore, in case of rectangular grid cells, an higher resolution will reduce the number of grid cells that cover both land and water. Since, this focus lies on the water levels in the pearl river delta, this area requires the highest resolution in he model. Delft3D offers a technique called domain decomposition that allows the model to be divided into several domains with different grid sizes. This provides the possibility to have larger grid cells in the open ocean and smaller grid cells with higher resolution in the coastal areas. The grid cell size also influences the computational time of the model i.e. the smaller the grid cells, the higher the total number of cells needed to cover the same area. Therefore, it is beneficial from a computational time perspective to have larger grid cells in the open ocean. The model setup in Delft3D uses three domains with different grid sizes. The large course domain has grid cells of 2 by 2 km, followed by the middle domain with grid cell size of 660 by 660 m. Finally, the fine domain that covers

the water surrounding Hong Kong and Macau has a resolution of 225 by 225 m.

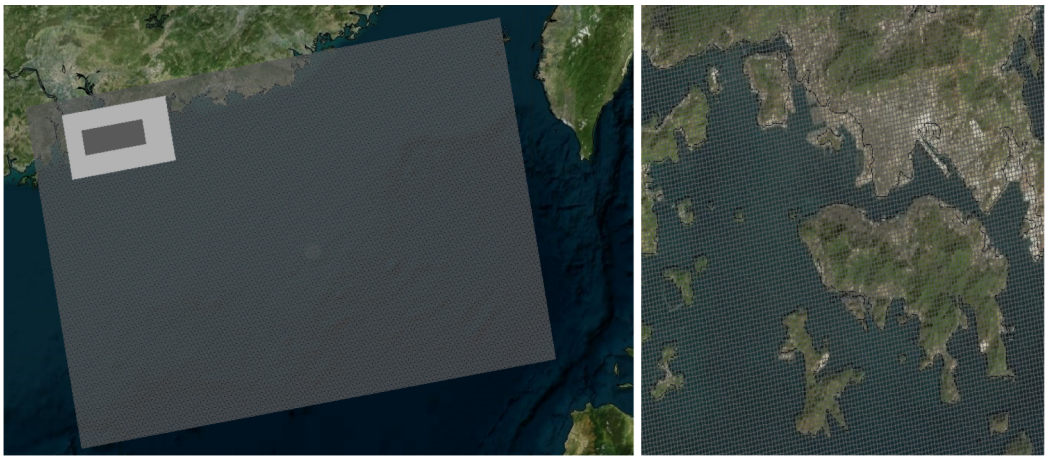


Figure 3.1: Computational grid a) Overview b) Fine grid Hong Kong area

Appendix B.4 aims at investigation the influence of the spatial resolution on the model performance. To do so, two storm surge simulations are conducted with differing resolutions. One model has three domain decomposition steps with a maximum spatial resolution of 225 by 225 m. The second model includes four different domain decomposition steps, with a maximum resolution of 75 by 75 m. Based of the results, it can be concluded that both for both models, almost the same results are obtained in terms of water levels. However, the high resolution model has a much longer computational time than the 3 step domain decomposition model. Based on the minimal difference between the two models and significant time increase of the finer model it is decided to use the 3 step domain decomposition model.

3.1.2. Bathymetry and Topography

Data on the topography and bathymetry for the area of interest are collected from different sources. For bathymetry of the Oceanic waters, the General Bathymetry Card Of Oceans (GEBCO19) is used. This database provides up to 500 m accurate bathymetry and land cards by using different measurements and estimations.

	Coordinate system	**Reference level**
Gebco19	EPSG:4326	Mean Sea Level
Pearl river delta bathymetry	EPSG:4326	Mean Lower Low Water (1.45m below MSL)
Hong Kong topography	EPSG:2326 (orignal HK grid system	Hong Kong Principal Datum (1.3m below MSL)
Macau topography	EPSG:102159 (Macau grid)	Mean Sea Level

Table 3.1: Topography and bathymetry sources

Before the data sets can be used to create the Delft3D model, some source must be transformed and edited. This is due to the fact the different coordinate systems and reference levels are used.The data with different coordinate systems and reference levels are all transformed to EPSG:4326 and MSL. The original topography cards are in 5m resolution. This very high resolution is not needed for the model since the main focus in not on inland flooding. To improve the computational speed, the topography data accuracy is reduced to 100m. Furthermore, some of the data overlap each other and need to be filtered out. The data filtering ensures that the most accurate data is used and the overlapping points are deleted. The local topography and bathymetry sources are significantly more accurate then the global Gebco19 database. The points from the Gebco database that overlap any of the points from the other sources are deleted. All the points from the Gebco database that represent lands with and altitude higher of 30m are also filtered out and deleted. Finally, the transformed and edited sources are combined in one .xyz file that can be used in Delft3D. To make the bathymetry suitable as input for Delft3d, it must be adapted to the computational grid. With QUICKIN the bathymetry points are averaged or interpolated over the grid to obtain bed levels at the grid cell corners.

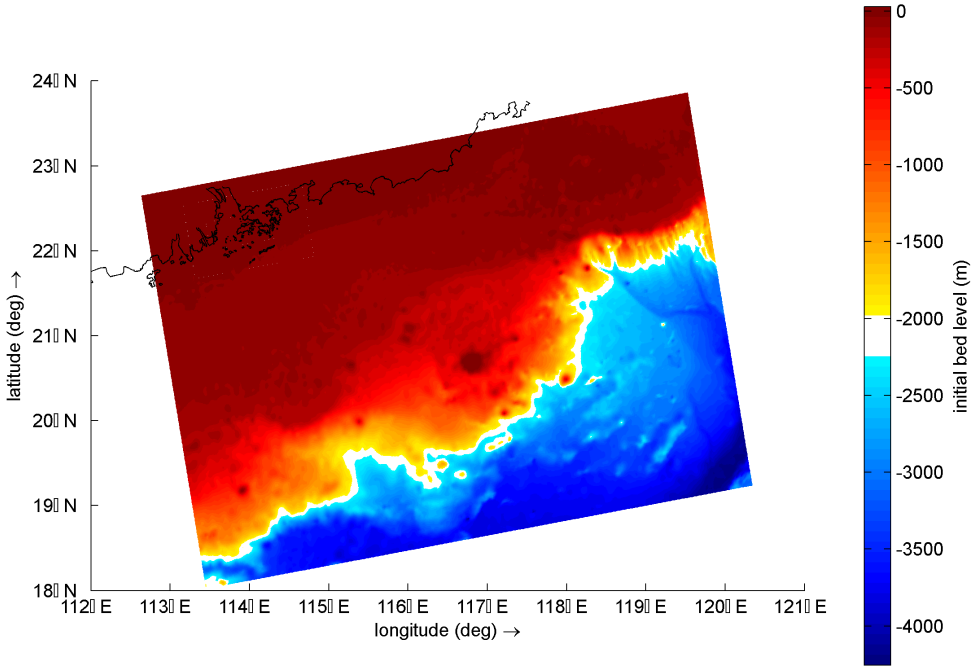


Figure 3.2: Bathymetry computational domain - South Chinese Sea

3.1.3. Time Step

In Delft3D flow the time step can be chosen based on accuracy reasons only, in most cases stability is no issue. A too large Δt will produce inaccurate results. On the other hand, a very small Δt will significantly increase the computational time, without large improvements in accuracy. The Courant-Friedrichs-Lewy number is used to determine the desired time step.

$$CFL = \frac{\Delta t \sqrt{gH}}{\{\Delta x, \Delta y\}} \tag{3.1}$$

For accuracy reasons the value of the Courant number may in general not exceed a value of 10. The Courant number is dependent on the time step, water depth and grid spacing. When using domain decomposition all the grids must have the same time step. The finest grid needs the smallest time step and is therefore leading for the other grids. When considering a grid spacing of 225m, and with Δt=60 seconds, the Courant number becomes 5.9 and is well below 10. An additional simulation was conducted with Δt=30 seconds to check if the accuracy could be improved. The results showed no accuracy improvements compared to the Δt=60 s simulation. The time step used for all simulations will therefore be 60 seconds.

3.1.4. Boundary conditions

To model the influence of the areas outside the model, one needs to impose boundary conditions at the open boundaries of the computational grid. Depending on the type of forcing from outside the domain that must be modelled, different types of boundary conditions can be imposed on the edges of the computational domain. Delft3D allows the user to choose between 5 different types of boundary forcing: water level, current (velocity), Neumann, Discharge or flux and Riemann boundary conditions. The boundary conditions are only imposed on open water boundaries on the outside of the course domain. At the intersection of the course with the middle domain and the middle with fine domain no boundary conditions need to be imposed. The choice of the type of open boundary condition is dependent on the phenomenon to be studied. The discharge or flux boundaries are often used for modelling river

flow problem with an inflow upstream and outflow downstream. The velocity boundary is often used for cross flow problems and the Neumann boundary condition is used to impose alongshore water level gradients. Finally, the Riemann boundary condition or so-called weakly reflective boundary. The main characteristic of this boundary is that it is transparent to a certain level for outgoing waves. This types of boundary allow outgoing waves to pass the boundary without being reflected back into the domain and causing disturbances. The water level boundary is mostly used for larger open ocean models, since this is often the only quantity that is known with some accuracy compared to the other boundary types. A single boundary condition is described at two points: Start point A and end point B. The points that lie between point A and B are calculated by Delft3D by means of linear interpolation. By imposing multiple boundary conditions along the domain, it is possible to simulate the physical behaviour of the water surface along the entire domain.

For the water level, current, discharge boundary types one also need to assign an reflection parameter to each boundary. The reflection parameters specifies the amount by which the open boundary should be less reflective for short wave disturbance. The Delft3D flow manual proposes the following values for the reflection parameter:

$$Water\ level\ boundary:\ \alpha = T_d\sqrt{\frac{H}{g}}, [\mathrm{s}^2] \tag{3.2}$$

$$Velocity\ boundary:\ \alpha = T_d[s] \tag{3.3}$$

As can be seen from the equations, the reflection parameter (α) is dependent on the time it takes a free surface wave to travel from the left boundary to the right boundary (T_d). Being dependent on the travel time, the reflection parameter is thus also dependent on the size of domain. Generally speaking reflection values up to 1000 s2 are considered safe. By increasing the reflection parameter values, to boundary also start to dampen the forcing that is sent by the boundary itself. Due to the size of domain on restrictions on the use of the parameter, the values for α has been determined by trial and error and comparing results with each other. The final input in the boundary section is the type of boundary forcing. Delft3d allows the user to choose between different types of forcing:

- **Astronomic**. Astronomic constituents are used for imposing the phases and amplitudes of the tidal constituents. As is stated in previous chapters, the water level elevation changes cause by tidal forcing can be described by taking the amplitude and phase of a large number of tidal constituents and equation 2.3. The values of each constituent are determined for the start time of the simulation and are updated every time interval which can be set by the user.
- **Harmonic**. Flow conditions are specified using frequencies, amplitudes and phases at the beginning and end of the simulation
- **QH-relation**. water level is calculated from the computed discharge passing through the boundary.
- **Time-series**. flow conditions specified by time series. For example the water level change at an open boundary over a period of time.

Given the purpose of the storm surge simulation model, boundary conditions of the water level type in combination with astronomical forcing is favoured. The water level change due to the tides can be easily determined by taking the astronomical constituents at the outside boundary from one of the many tidal models that are available today. There are multiple open source platforms that provide this data in different spectral resolutions. To ensure the model's quality, three different models are used to determine the performing one. In Appendix B.1, a sensitivity analysis is conducted to determine the best performing tidal model. The model considered are the TPXO8, TPXO9 and FES2014 model. Both the TPXO8 and TPXO9 are regional models and have a resolution of 1/30 degrees (+- 500m. This means that all the grid points are space 500 m apart and that for each grid point a unique set of astronomical constituents are available that can be used for the boundary conditions. The resolution of

the FES2014 global model is 1/16 degrees (+- 250m). From the sensitivity analysis it can be concluded that the regional TPXO8 tidal model has the best performance. The TPXO8 model is therefore used in the boundary conditions for the simulations.

The following constituents are considered in the boundary conditions:

- M2: Principal lunar semidiurnal constituent
- S1: Principal solar semidiurnal constituent
- N2: Larger lunar elliptic semidiurnal constituent
- K2: Lunisolar semidiurnal constituent
- K1: Lunar diurnal constituent
- O1: Lunar diurnal constituent
- P1: Solar diurnal constituent
- Q1: Larger lunar elliptic diurnal constituent
- MM: Lunar monthly constituent
- MF: Lunisolar fortnightly constituent
- M4: Shallow water overtides of principal lunar constituent
- MN4: Shallow water quarter diurnal constituent
- MS4: Shallow water quarter diurnal constituent

3.1.5. Wind and wave forcing

For storm surge simulations, the wind input is the most important forcing factor in the model. As discussed in section 2.5 the observed wind speeds and pressure can be used in a parametric wind model to describe the wind and pressure field of the TC throughout it's life cycle. Data from historical TC's all around the world are stored in the IBTrACS database from the national oceanic and atmospheric administration (NOAA). The database contains information on the location, maximum wind speed, pressures and other relevant parameters for the duration of the TC. The TC track data is used as input for a matlab script that applies the wind model as discussed in section 2.5. The applied wind model creates a file with the wind speed and pressure fields on a moving spiderweb grid. This spiderweb file is used as the TC forcing in the model. Figure 3.3 shows that reproduction of the wind field in Delft3D during an approaching TC.

For the simulation of waves, the Delft3D wave module can be used which runs the SWAN model to simulate the evolution of waves. The Delft3D wave module can be coupled with the flow module to obtain a combined flow/wave simulation. With the SWAN model a wide variety of physics can be simulated, for example: refraction, wave generation by wind, dissipation, shoaling, diffraction etc. For this research the wave module is used to determine the extent of the wave setup during the storm surge. The wave module used the wind forcing form the Delft3D model to compute the wave field and propagation. Just like the flow module, the standalone wave module requires a computational grid to compute to wave field on. The wave module allows for different grid resolutions by means of nesting. For accuracy reasons, three different wave grids are used. The course large grid cover the same space as the flow grid. The two finer wave grids are nested in the course and allow for a more detailed simulation in the coastal region.

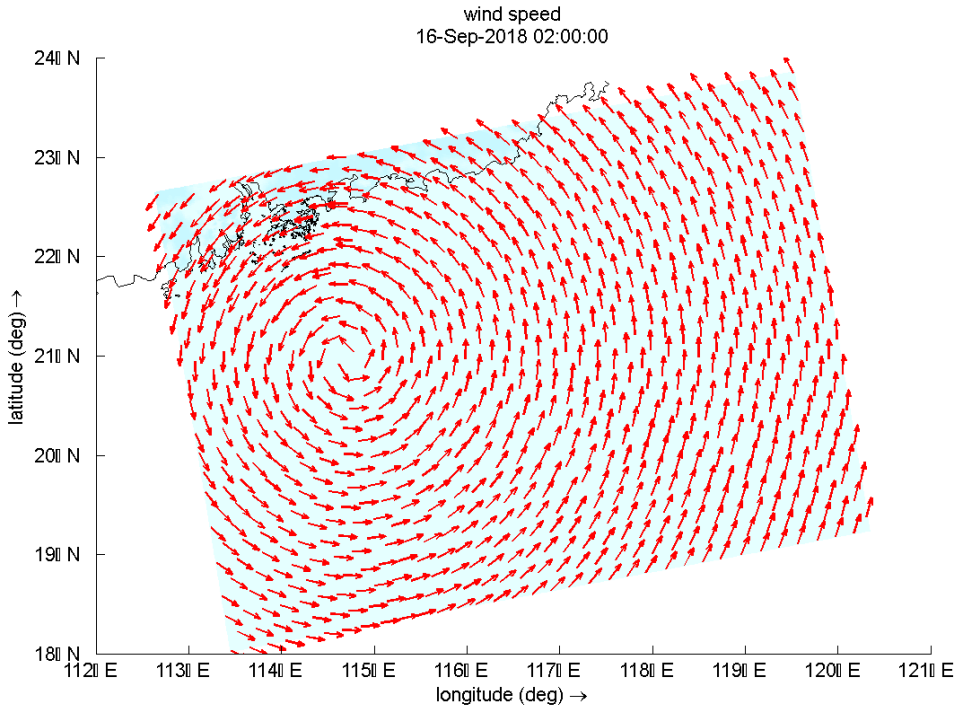


Figure 3.3: Wind field simulated in Delft3D flow for an TC

3.2. Model Calibration and Validation

3.2.1. Astronomical tide validation

As stated in the previous section, the boundary conditions provide the tidal forcing into the computational domain. Multiple tidal models with different resolutions are available. To determine the most accurate model for this study, the different tidal models are tested in Delft3D by means of a tidal validation. The main purpose of the tidal validation is to check if the simulated tides correspond to the measurement station in Hong Kong. For tidal validation purposes, the Hong Kong marine department provided the data sets for three different tidal measurement stations in the Hong Kong waters. These stations measure the water level elevation in 10 minute intervals. The reference level used for tidal measurements in Hong Kong is the principal datum (1.45m below mean sea level). Unfortunately it was impossible to obtain tidal measurement data for the waters surrounding Macau. Therefore, the tidal validation of the hydrodynamic model is based on the tidal movement in Hong Kong only. The CT8 station is location at the entrance of a large contained port. The water depth at stations CT8 and MWC is approximately 18 meter measured from mean sea level. Station CHC is not located near a large port or shipping channel. The water depth at this station is approximately 10 meter

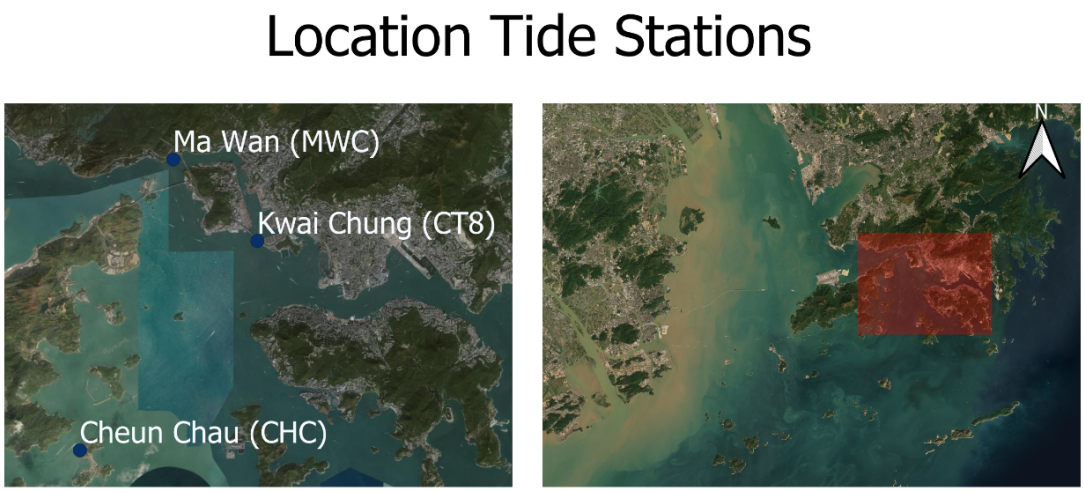


Figure 3.4: Location of tide stations

The tidal measurement stations also measure the water level elevation due to other hydrodynamic processes besides pure tide movement. In the real world tides are not the only forcing of the water level. Wind setup, wave setup, river discharge, precipitation are aspects that can contribute to the measured water level elevation. For the tidal validation, it is therefore important to choose a time frame for the simulation in a period of low wind speeds, wave heights etc. By taking a period of calm weather in the simulation the other processes are minimal and the pure tide simulation will obtain results that are closer to the measured water level elevations. Based on an analysis of historic weather in Hong Kong and Macau Augustus 2015 has been chosen for the tidal validation. During this month the wind speed and waves were historically low and the simulation gave significant better results than for period with more wind and waves.

The analysis of the different models in appendix B shows that all three considered tidal models can reproduce the water levels for Hong Kong with an acceptable error for the time and phase. In general the measured tidal signal is reproduced correctly by the numerical model. However, some under and over estimations during high and low tide can be observed. This can be explained by the fact the only tides were included in the simulation and other physical processes were excluded. Processes like wind and wave setup, river discharges and precipitation can all have an effect on the water level elevation. From the results it can clearly be seen that the simulated tidal signal is in phase for all three tidal measurement stations. This means that the propagation of the tidal wave is captures correctly in the model. On the other hand, some errors can be observed for the simulated water levels during

high and low tide. The highest errors can be observer during the period around neap tide (22-25 Augustus). The results clearly show that around this time period the model performance is reduced for all three models. Since the tidal models all produce similar results it is hard to visually determine the best performing model. Although the root mean square (RMSE) and percentage error are similar for the three models, the TPXO8 model has the best performance overall. For the entire month Augustus 2015 the RMSE is between 10 and 12 cm for all considered tidal stations.

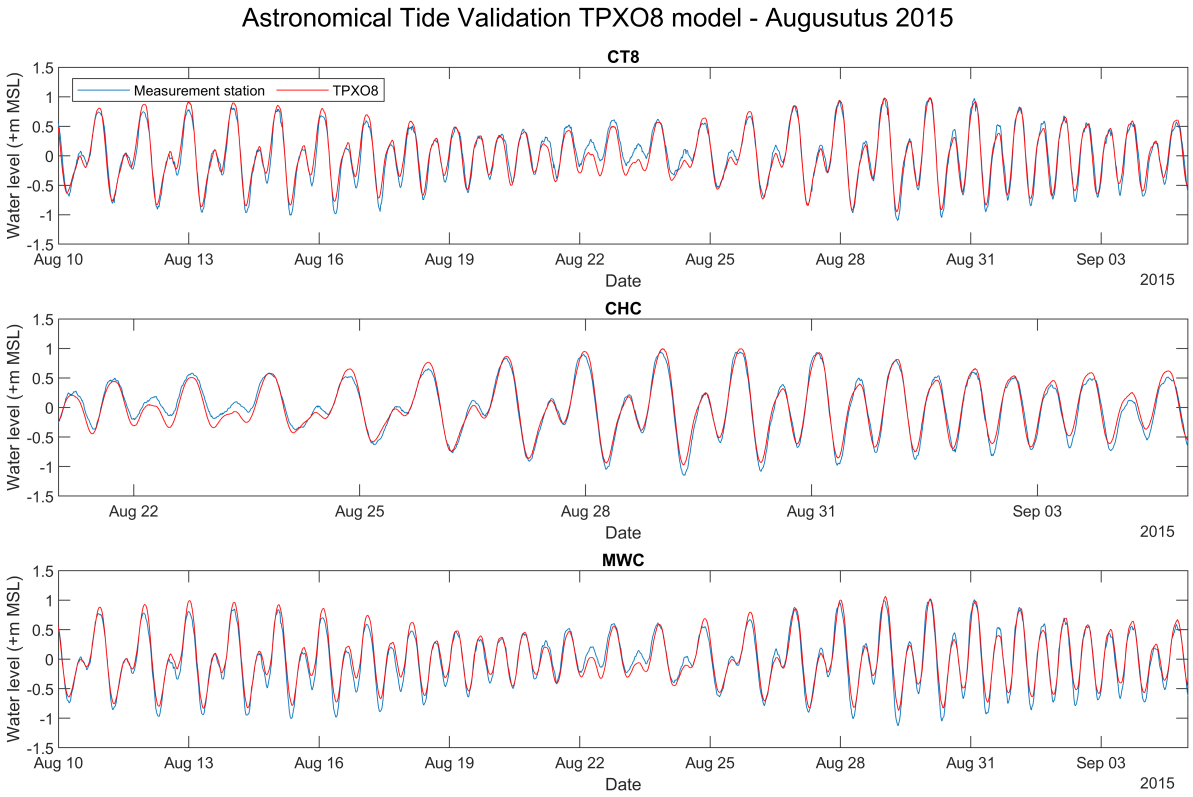


Figure 3.5: Astronomical Tide Validation for three tide station - TPXO8 model

3.2.2. Storm tide validation

Since the tidal validation has shown that the model can reproduce satisfactory tidal results, the next step is to include wind forcing in the model, to simulate storm surges and determine the storm tide (mean storm surge level plus tide). The validation is required to evaluate the performance of the model, can it estimate surge heights within reasonable accuracy? The validation is based on three historical TC's namely: Hato (2017), Mangkhut (2018), and Hagupit (2008). As stated before, the track data from the TC's are parametrized and converted to spiderweb format for the Delft3D input. The different TC agencies all use slightly differing methods or parameters to describe the TC. Appendix B.3 aims at investigating the influence of the data from different agencies and different derived radius of maximum winds. he results show that the simulated surge heights can differ significantly depending the data used. Besides the wind speed and eye pressure, the RMW is plays also an important role in the simulation results.

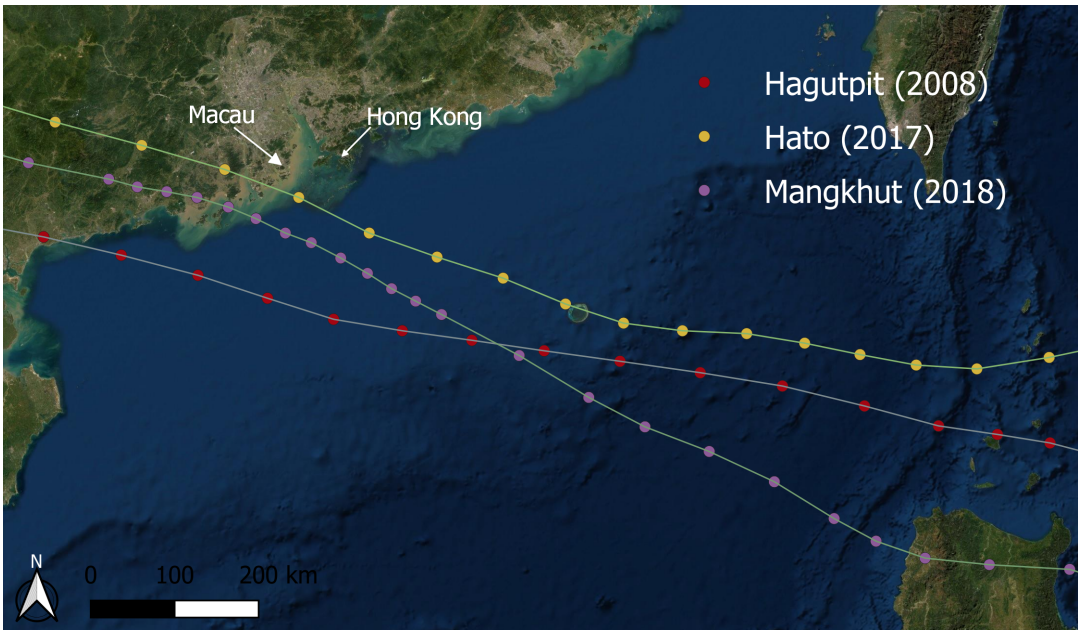


Figure 3.6: Tropical cyclone tracks

Furthermore, in appendix B.2 the influence of the values for wind drag coefficients for varying wind speeds is also investigated. Over the years many investigations have been conducted into the behaviour of the wind drag coefficient under cyclone conditions. It is generally agreed that the drag coefficient increases with the wind speed until a maximum wind speed after which the drag coefficient value will decrease again. However, the exact values for the drag coefficient can differ per research. During the sensitivity analysis a total of 5 different relations for the drag coefficient were used in the simulation for TC Hato. Again the results show that the type of drag coefficient relation used significantly influences the simulated surge height. One limitation of Delft3D becomes clear for the wind drag coefficients. Delft3d only allows for linear interpolation between three values of the drag coefficient for varying wind speeds. However, multiple recent papers suggest a parabolic relationship between the drag coefficient and wind speed. Due to the linear behaviour of the drag coefficients in Delft3D, it is impossible to capture the drag coefficient values for all wind speed points exactly. As a consequence, the linear interpolated drag coefficient values are slightly underestimated compared to the parabolic values. In the end, the sensitivity analysis shows that the drag coefficient by Peng et al. specially derived for the south Chinese sea will generate the most accurate results in Delft3D.

Hato (2017)

Hato was one of the strongest TC to impact Hong Kong and Macau over the last 50 years. Hato started as a tropical depression on August 18 southwest of Taiwan. Over the following days the storm kept increasing in size and intensity and was declared a tropical cyclone. On August 23 Hato made landfall near the pearl river delta, it was classified as an category 2 typhoon by the JWTC. The storm surge

associated with Hato caused flooding of multiple area along the pearl river delta. In Hong Kong city, most damages reported came from wind damage. However some low lying areas were inundated due to the storm surge. Macau was hit even harder by Hato, Many parts of the city were inundated. The eye of the storm passed Macau at a distance of approximately 30 km. According to Takagi et al. (2018), the surge heights reached up to 2m in Hong Kong and up to 2.5 m in Macau. Field surveys after the TC provided evidence of an inundation depth of over 2m for downtown Macau. According to Li et al. (2018) several factors were responsible for the extent of damages in Macau. The first one begin that Hato made landfall during high tide. Furthermore, the wind speed measured was the strongest of all the TC's since 1953.

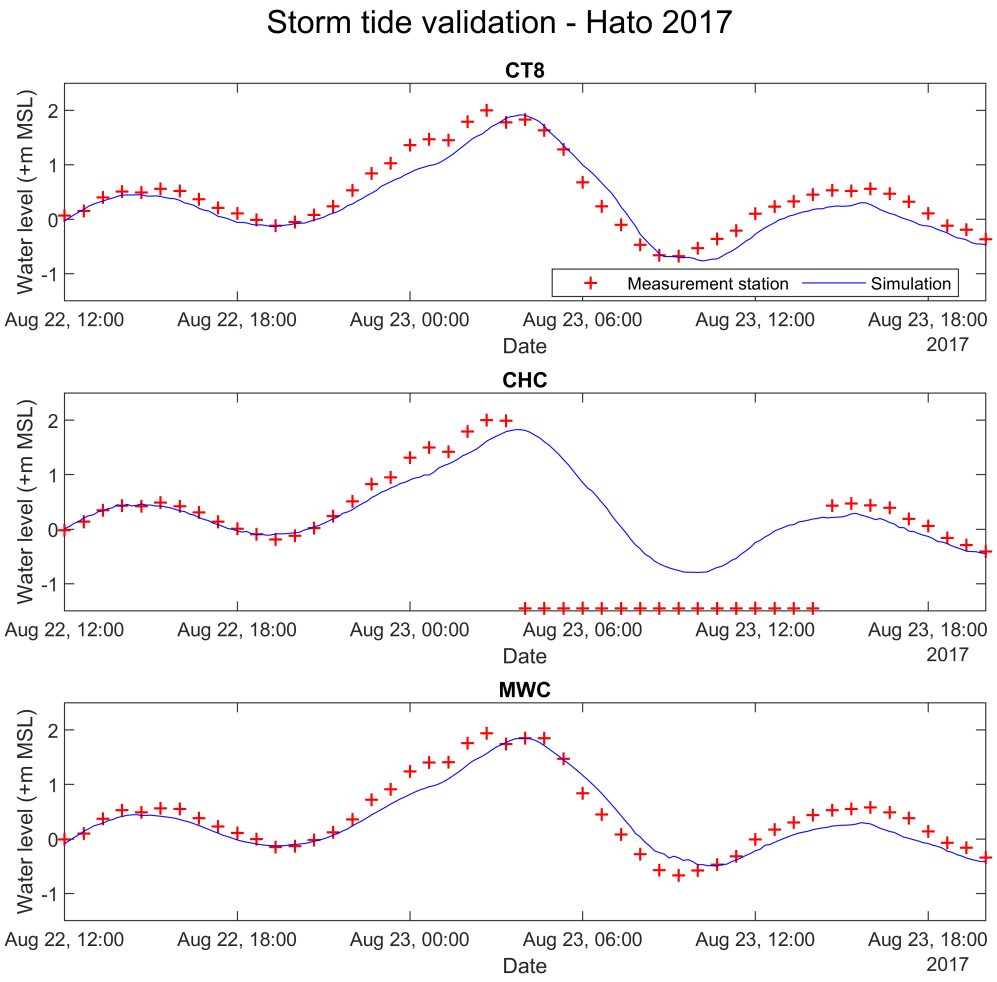


Figure 3.7: Observed vs simulated storm tide for TC Hato

The results of the simulation show that the model is capable of estimating the maximum surge height within reasonable levels. The smallest error is obtained for station ct8 an is 8 cm, the highest error is at CHC with 17 cm difference between the maximum simulated surge height and the measured height. It must be noted that the tidal measurement buoy CHC had a malfunctioning during the peak of the storm. The maximum measured surge heights were all recorded at 23/09 at 2:40 am. However in the simulations, the maximum surge height occurred more than an hour later at approximately 4:00 am.

Mangkhut (2018)

Mangkhut was first noticed as an tropical disturbance on September 5. The following days, the storm increased in intensity and made landfall as a category 5 TC on September 15 at the coast of the Philippines. After traversing the Philippines, Mangkhut weakened to an category 3 TC while passing over the South Chinese Sea. Early September 16 the storm passed in front of the coast of the PRD. Mangkhut is classified as the strongest hitting typhoon in terms of maximum sustained wind since reliable records are kept. In terms of the affected area, Mangkhut was the most severe event since 1979 to hit the PRD Yang et al. (2019). Although Mangkhut had similar meteorological characteristics compared to Hato, some key differences can be defined. First, Mangkhut passed the PRD at neap tide. The radius of the storm was significantly larger for Mangkhut. The eye passed Macau at a distance of approximately 70 km. The storm surge caused flooding in parts of Hong Kong and Macau. Other than during Hato, no fatalities were registered during Mangkhut.

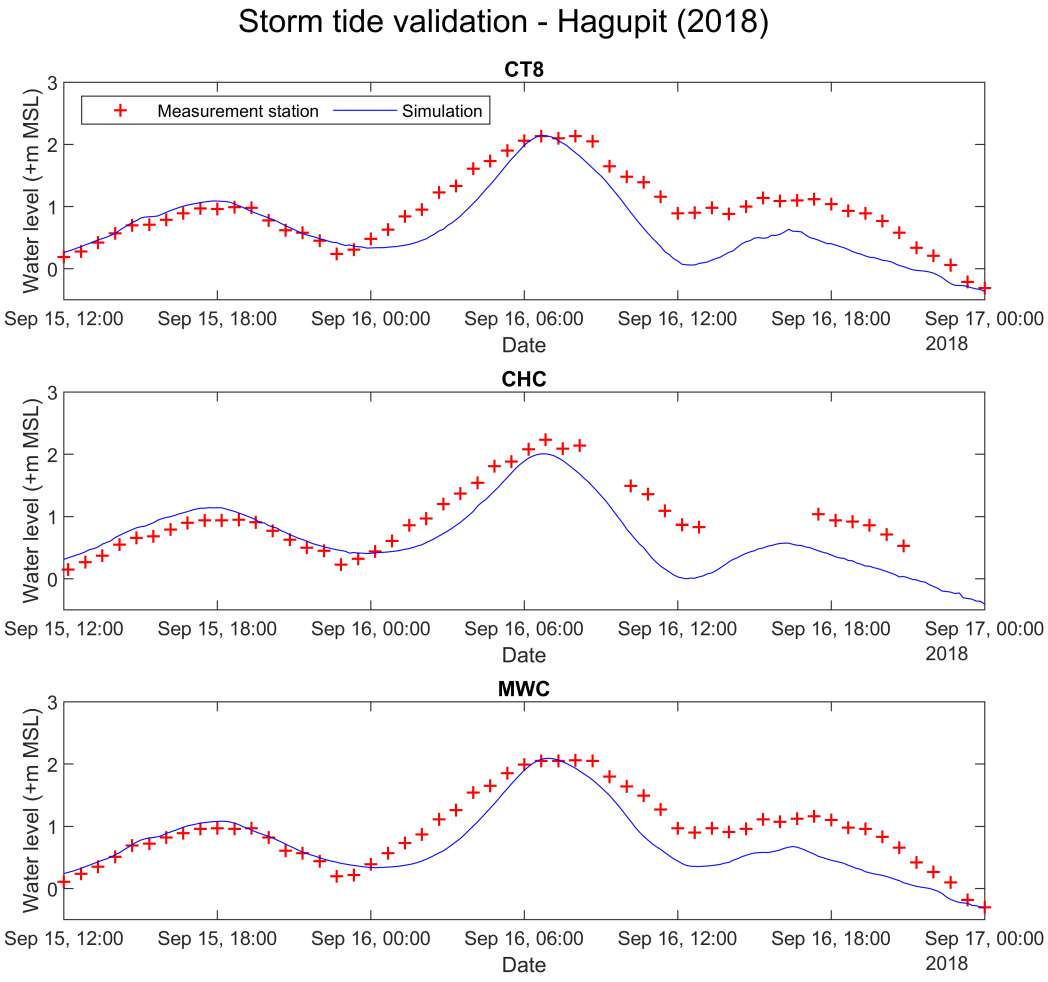


Figure 3.8: Observed vs simulated storm tide for TC Mangkhut

At first sight, the results of the simulation shows that the maximum surge height is estimated in correspondence with the measured signals for station CT8 and MWC. For station CHC a slight underestimation for the maximum surge height is estimated. The arrival of the storm surge peak in the simulation corresponds to the measured arrival time. However, it can also be noted that the water level before and after the passing of Mangkhut are under estimated in the model. Especially the water level after the passing of Mangkhut are underestimated in the simulation for all three measurement stations. The relative poor performance of the model after the passing of the TC can be explained by the fact that Mangkhut occurred during neap tide. As discussed during the tidal validation. The tidal constituents used in the model have the largest errors in the period around neap tide. After the passing of the storm, the tides become the main forcing of the water levels again. Since the tide estimation of

the model during period of neap tide is not optimal, the water levels after the passing of Mangkhut are underestimated. Other factors that can be responsible for part of the discrepancies are possible errors in the bathymetry/ topography. The underestimation of the water levels during the period around neap tide is not necessary a problem for this thesis research. We are mainly interested in the maximum surge heights and the arrival times. Therefore the period before and after the storm surge are of less importance for the research. Furthermore, the synthetic TC simulation will be done without tidal forcing present in the model. The reason behind this will be explained in the next chapter.

3.2.3. Hagupit (2008)

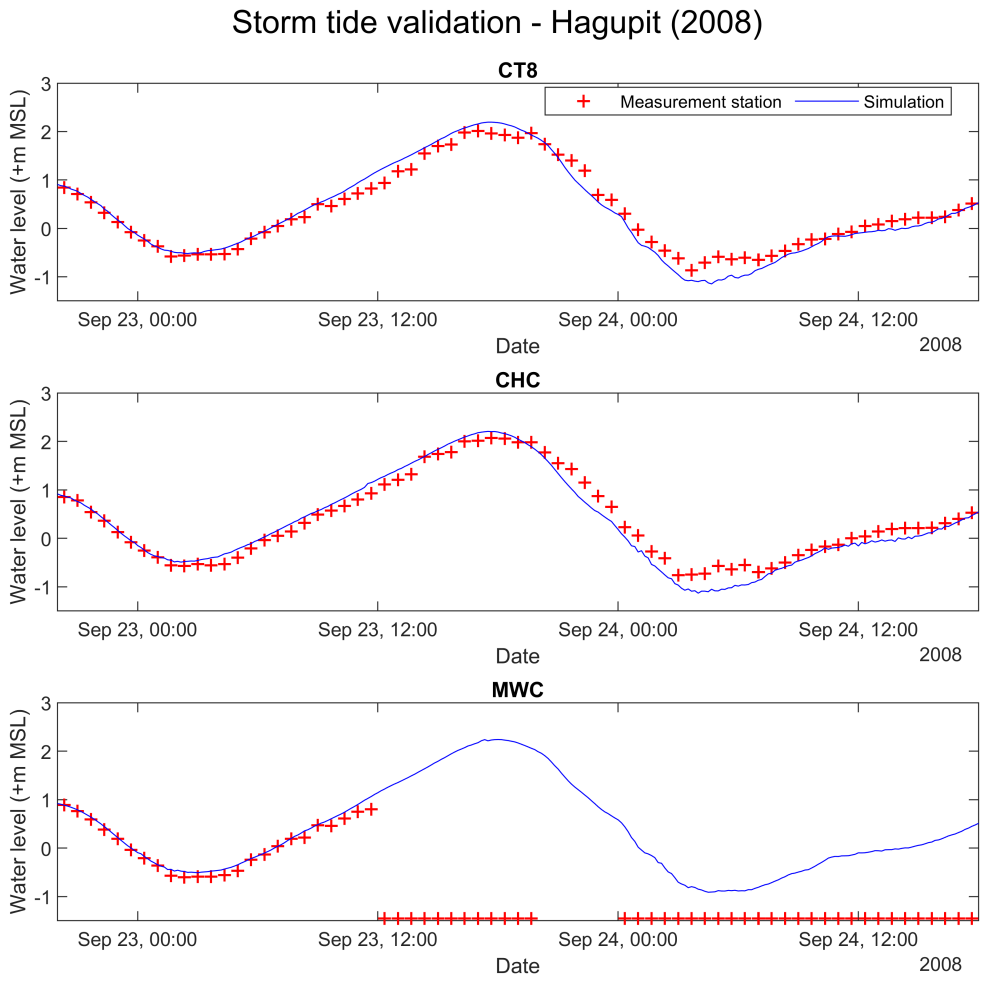


Figure 3.9: Observed vs simulated storm tide for TC Hagupit

The first thing that can be noticed from the figure is the malfunctioning of station MWC during the passing of Hagupit. The results show that for station CT8 and CHC the simulated water levels are in good correspondence with the measured levels. The maximum simulated surge height is slightly overestimated with 10 cm. Other than for Hato and Mangkhut, the rise and fall of the water levels are very similar to the measured levels.

3.2.4. Storm tide simulation with waves

As discussed previously, waves can have significant impacts on coastal areas. Especially in the breaker zone a lot of wave energy is dissipated. The wave simulation shows that high waves are generated during the passing of the storm. Especially, the high wave heights can be observed in front of the PRD during passing of the storm. The many islands in front of the PRD are responsible for refraction and diffraction and create shadow zones with low wave heights on the back side. A storm tide simulation with waves is performed, to gain knowledge into the degree waves are responsible for water level increases during a TC event. The storm tide model for TC Hato is coupled with a separate SWAN (wave) model. The SWAN model is included in the Delft3D package. To so, so a two separate wave grids are created with RGFGRID and QUICKIN. Figures 3.10 and 3.11 show a snapshot of the significant wave height during the peak storm tide for the two grids. The results shows, that the significant wave height can reach up to 5 m in the outskirts of the pearls river delta. Near the cities of Hong Kong and Macau the significant wave heights are up 3 meter.

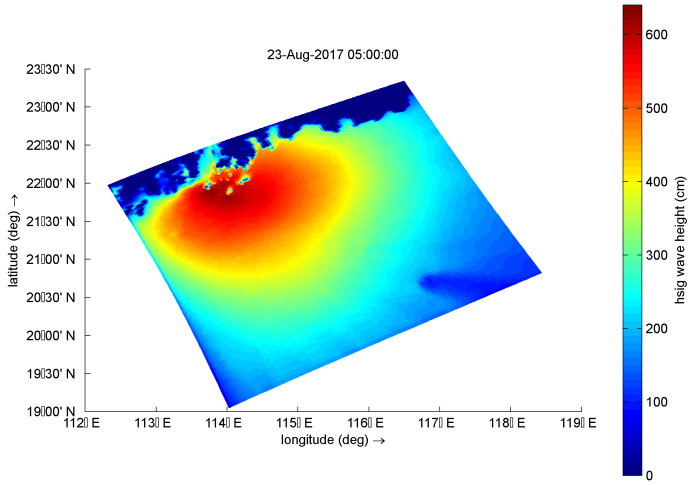


Figure 3.10: Significant wave height at peak storm tide, Course computational grid - Hato (2017)

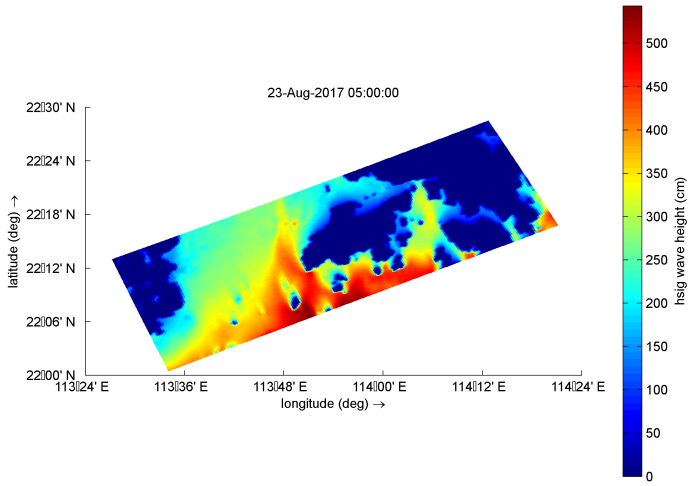


Figure 3.11: Significant wave height at peak storm tide, Fine computational grid - Hato (2017)

For wave validation purposes, the Hong Kong Marine department shared data of two wave measurement buoys. However, relevant wave data was missing for all three validations TC's. Due to the lack of data, the wave model can therefore not be validated accurately. Other data sources mention the maximum wave heights during storms, but without any exact locations. According to Takagi et al. (2018), the off shore significant wave height in front of Hong Kong island was 5.5 m during the passing of Hato. Although it is impossible to to validate the wave model without exact measurement locations, the model is in line with a significant wave height of 5-6 m in front of Hong Kong island. As stated in previous sections, the wave setup depends largely on the radiation stress and water depth. The effect of wave setup will therefore be the largest in the surf zone where radiation stress decreases. The tidal gauges that are used for the validation process are located in deep water and are therefore expected to have minimal increase in water level elevation due to wave setup.

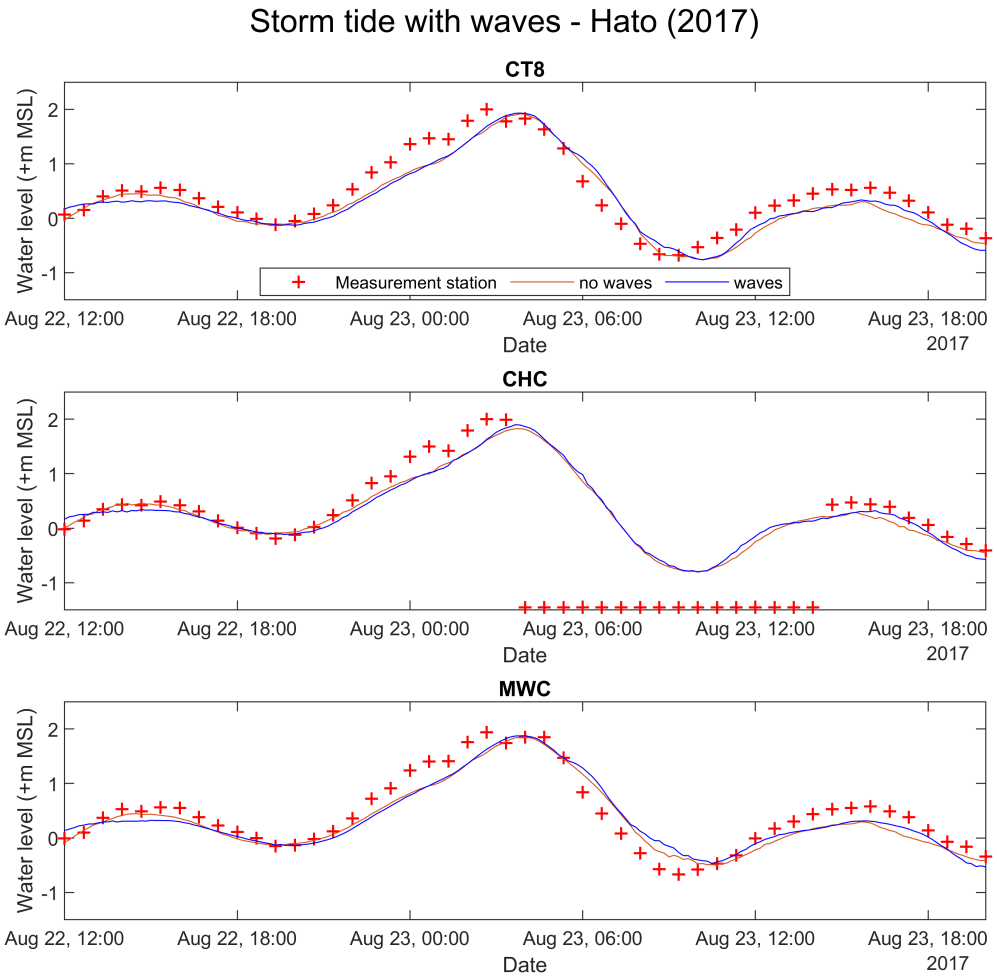


Figure 3.12: Storm tide with waves -TC Hato (2017)

From the results it can clearly be seen that that wave have hardly any influence on the water levels at the measurement locations. The maximum water level increase compared for the simulation with waves compared to the one without waves is 5 cm at the location of the tidal gauges. A consequence of running the coupled flow and wave model, is the sharp increase in computational time. A coupled simulation run takes on average about 300 % more time than a flow model without waves. Since the effect of wave setup is very minimal for the water levels at the tidal gauges, it will be excluded for the synthetic simulation. Due to the large amount of synthetic simulation, the computation time becomes a relevant factor in the process. Although the waves show a very limited effect on the water level of the considered locations, wave setup will be significant in the surf zone. Additionally the enormous amounts of energy dissipated in the surf zone will act on flood defences and coastal structures.

4

Synthetic Storm Surge Simulations

In the previous chapter, the Delft3d model setup and performance has been discussed in detail. The hydrodynamic model forms the basis to gather the data on the water levels variations during storms that will be used in the neural network. The development of the neural network requires a lot of data input for the learning process. Generally speaking, the more data input the more accurate the network will become in predicting the water levels during a storm surge. The meteorological data (TC track data) used by the Delft3D model to estimate the water levels is limited by the amount of historical storms that passed through the PRD. To overcome this limiting data factor, a synthetic storm database is used that contain significantly more track data than the IBTrACS database. The development of the catalogue is discussed in 2.8.

4.1. Synthetic simulations

The contents of the synthetic storm database are used for running a specified number of synthetic storm surge simulations in Delft3D. The synthetic TC track data forms the input for the parametric wind model in order to create the spiderweb file required for each simulation. Other than for historical storms, the synthetic storms don't have a date of occurrence, but only a month and year assigned. This forms a problem for the set-up of the Delft3D model. In order for the model to know when to start and stop the simulation, it requires a start and stop date and time. Furthermore, the tidal elevation and forcing changes over the month and seasons and therefore dependent on the exact date and time. There are two main options to overcome the missing date problem. The first one is to assign a random start date to the synthetic storm in order to capture the astronomical tides. By doing so each storm is assigned a start date, which can be used to model the tidal forcing at that point in time. This option allows for the astronomical constituents to be included in the neural network as input neuron. The other option is to separate the storm and tidal forcing by removing astronomical tides from the model. This option eliminates the need to assign a random date to each storm no neural network input node is required with the tidal forcing, simplifying the structure of the neural network (This will be discussed in the next chapter). However, if the tidal forcing is removed from the simulation, the neural network that is developed will be only be able to predict the water levels without astronomical tides. In the real world the tide plays an important role in the change change of the water levels. To get accurate real life predictions in case of an approaching storms, the tidal elevation must be added to the output of the neural network in case of this option is chosen. By adding the expected tidal elevation at the moment of maximum water levels, one can still get an accurate prediction of the maximum real life water levels during a storm surge.

4.2. TC selection criteria and scripts

Do to the size of the Western Pacific basin and the number of storms in the catalogue a total of 1000 storms are selected from the database based on different criteria. To automate the selection process a Matlab script was created that selects all data from the synthetic storms that are selected and assigns an unique ID to each storm. Creating unique ID's for each storm is needed for correct naming of the corresponding flow files for each storm.

- **Location**. In order for a storm to be selected, it must have a minimal of two track points (Location eye of the storm) inside a specified area surrounding Hong Kong and Macau.
- **Landfall**. The storm must make landfall in the area of previous criteria to be selected (Storms that don't make landfall area not avoided.
- **Category**. The storm must reach a minimal category of 1 on the Saffir-Simpson Hurricane Wind Scale during the duration of the storm. All storms with category 0 are called tropical depressions and will not be considered.
- **Minimal duration**. To ensure that the Delft3D model has sufficient simulation time to allow for model spin up, the minimal duration of a storm must be 48 hours

The Matlab script only selects a storm if the criteria discussed above are met. The track data of the selected storm is then stored in a new array for further use. For each synthetic storm that is simulated in Delft3D a separate spiderweb must be generated based of the track data in the catalogue. To run a Delft3D flow model, the program requires a wide variety of input files. Some files are the same for all synthetic simulation like the grid, bathymetry, domain and boundary conditions files. However, for some files the name and/or their input must be changed for each simulation. For example the spiderweb file is different for each simulation and therefore also needs an unique name. Another example is the master definition file (MDF) and the boundary definitions file (.bnd). The MDF is a file that contains all the settings of the model and also calls the other files required for the simulation. While most of the content stays the same, some settings change for each simulation. The start and stop time (duration) is different for each storm and therefore has to change every time. Additionally a new spiderweb file must be called as well with the corresponding wind forcing input. Due to the sheer size of the simulations to be conducted it would be inefficient to generate all these files manually. To speed up the process, the Matlab script previously discussed is supplemented with a part that automatically generates and names the spiderweb files for each storm. The method of the generation of the spiderweb file from track data is the same as discussed in previous chapter. Additionally the Matlab script also generated uniquely names MDF files for each domain with that contain the correct spiderweb file names and simulation duration.

The synthetic simulations are be computed in the TU Delft cluster, a system that provides high performance computing power for research purposes. The cluster contains 12 nodes, which means that it can compute 12 different jobs at the same time. The cluster completes the jobs significantly faster than a normal laptop and can do multiple at the same time. To make running the simulation on the cluster possible some additional files must be created. The required files to run each simulation in the cluster (.xml, .url. &.sh) are also created. In the end with one press of a button, the Matlab script will select all the storms (or a part) that full fills the set requirements, generate the corresponding Delft3D files with unique ID names and save all files in a unique folder per storm event. Although being setup for the Hong Kong area in the western pacific, the script can be easily adapted for other areas of interest and different selection criteria and therefore very use full for anyone who uses this Synthetic TC catalogue in combination with Delft3D-Flow. The working of the Matlab script is discussed in more detail in appendix D.

4.3. Validation Delft3D model without astronomical tide forcing

The last step before starting with the synthetic simulations is the validation of the storm surge model without any tides. The synthetic models differs from the original model discussed in chapter 3 in the fact that tides are excluded during the synthetic simulations. Since there is no tidal forcing present in the model anymore, the surge height will consequently also decrease depending the tidal elevation at the arrival time of the TC. Since the tide less model will be executed in the TU Delft cluster, this also provides the ability to validate the storm surge model when running in the cluster instead on a PC. This ensures, that Delft3D works and produces the same results as in the PC simulations.

To validate the output of the model with the new Riemann weakly reflective boundary conditions TC Hato is validated. To do so, the water level change at the three tidal stations is selected. To compare the tide less simulation with the actual water level that occurred during TC Hato, the astronomical tide at that time must be manually added to the Delft3D output. Figure 4.1 displays the results for tidal measurement station CHC for the simulation without astronomical tides. The figure shows clearly that the harmonic water level change is not present anymore, the only forcing left in the model is the wind.

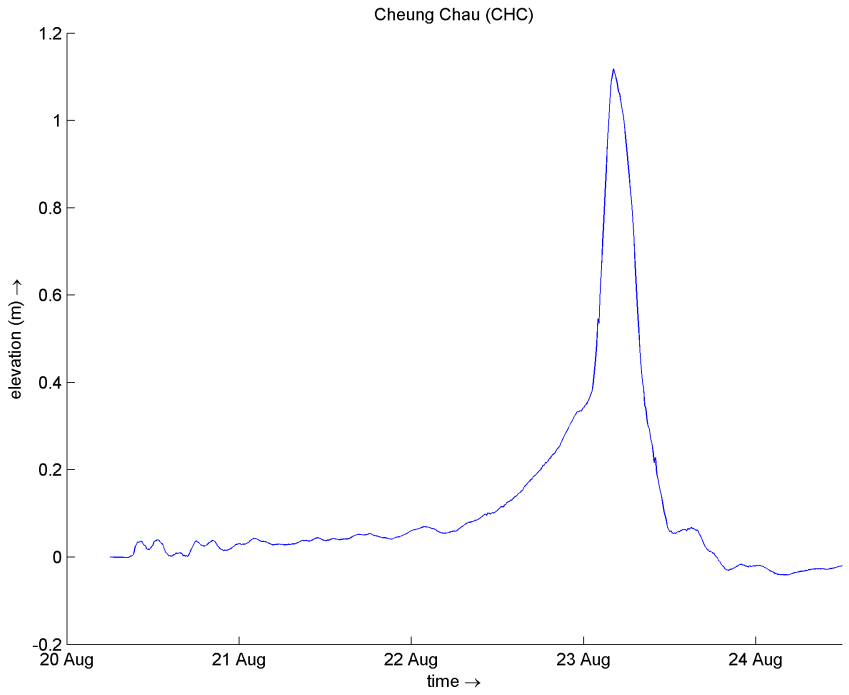


Figure 4.1: Storm surge without astronomical tide - tide station: CHC. Hato (2017)

To compare the tide less simulation with the observed water levels, the astronomical tide during Hato must be added to the water level output. The peak of the water level occurred at approximately 3 am on 23 August 2017. The exact tidal elevation at that time can be determined with the Xtide toolbox. Xtide is an package that provides astronomical tide and current predictions all around the world. The algorithm used by the Xtide package is same as used by the NOAA. Figure 4.2 shows the tidal prediction for Hong Kong on 23 August 2017. The reference level is in Hong Kong Chart Datum or Mean lower low water which is 1.45m below sea level (table 3.1). The tidal elevation at 3 am on 23 August was 2.3 m above HK chart datum, which represent a tidal elevation of 0.85 m above MSL.

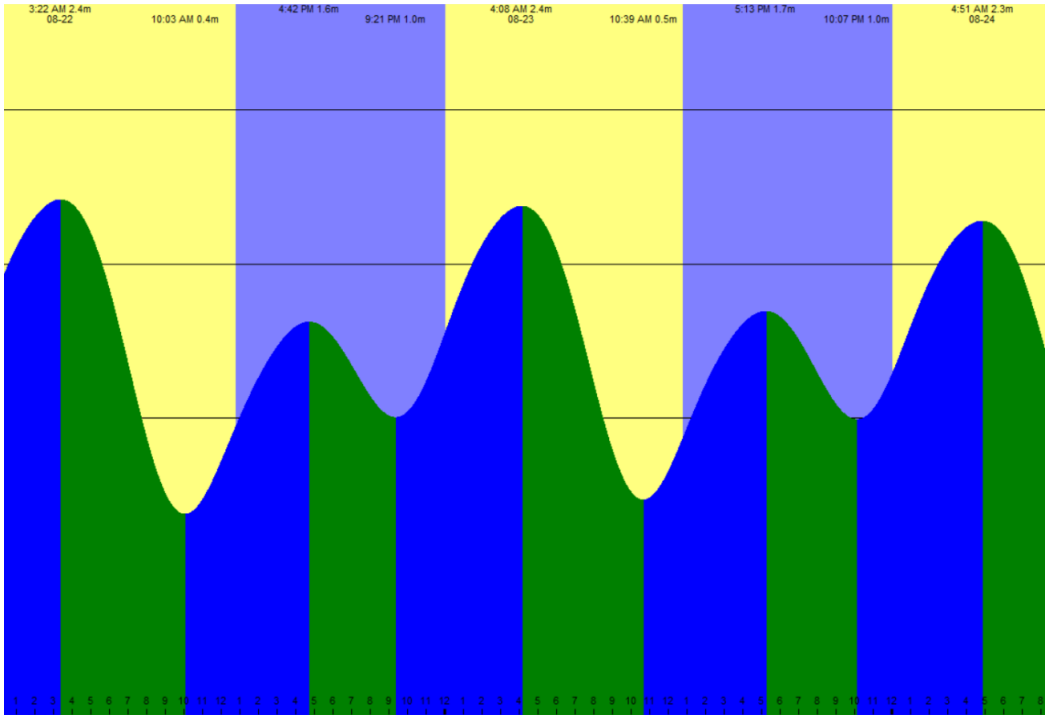


Figure 4.2: Xtide Tidal prediction for Hong Kong 23 August 2017 (reference level HK chart datum)

The comparison between the simulation and observed water levels shows that when the tidal elevation is added to the simulation water level, the results are within 10 cm from the observed values. The low errors show that the performance of the tide less model can be validated to a historical storm. This shows, that removing astronomical tides from the simulation does change the model performance and still produces usable results.

	CT8	CHC	MWC
Synthetic sim wl (m)	1.2+0.75	1.12+0.75	1.17+0.75
Measured wl (m)	2.0	2.0	1.95
Error (m)	-0.05	- 0.13	0.03

Table 4.1: Simulation without tide - Validation TC Hato 2017

5

Artificial Neural Network

The main goal during this thesis study is the development of a neural network capable of predicting maximum water levels in case of an approaching TC. All the work that has been done so-far (setting-up, calibrating, validating the storm surge model and running synthetic simulations) in the cluster was with the purpose of collecting the data needed for the development of the neural network. In this chapter, the development of the neural network is treated in detail. First, an overall approach is presented to obtain maximum predicted water levels from the neural network. This is followed by a detailed explanation of the input and output data needed for training the neural network. The NN input and output data is selected and adapted where needed to make it suitable for the NN input. As discussed in section 2.7, there are many different possible architectures for the development of neural networks. Depending on the problem to solve, complexity and size of the data, one can make changes to the architecture to improve the results of the NN. To find the most optimal architecture, different configurations and training algorithms are investigated in section 5.3. Finally, the different NN configurations will be validated and their performance checked (section 5.4, 5.5 and 5.6)

5.1. General approach

Most recent applications on storm surge predictions have been conducted for historical storms. For these storms, the meteorological data is actually obtained from measurements stations. The same holds for the water levels that are recorded by measurement buoys at specific locations. During this study the approach towards NN modelling will be different compared to the previous discussed approaches. The reason for the different approach is the fact that during this study, the NN data is obtained from synthetic track data and simulations. Other than real life water level measurements, the synthetic simulations give insight on the water levels in the entire computational domain and not only a specific measurement location. Thus the synthetic simulations provide the data to increase the number of output neurons in the network, to give insight in the water levels of the entire domain instead at only specific locations. Same as the output of Delft3D, it is possible to develop a neural network that instead of predicting water levels at one location, is capable to output a map with predicted maximum water levels. An advantage of working with synthetic data compared to historical data is the ability to obtain data from a large number of storms. The synthetic track data contains 10000 years worth of storm while historical data sets like IBTrACS only contain up to 40 years of historical data. Generally speaking, neural networks will become more accurate when increasing the number of training samples. On the other side, by increasing the number of output neurons the networks architecture will become more complicated and will require more hidden neurons and more computational time and memory. Furthermore, this approach requires extensive testing and validating of the storm surge model to calculate water levels. When checking the literature, no publications can be found of NN's capable of predicting maximum water levels in map format. Therefore, the development of the NN will investigate the suitability and accuracy of such an network compared to the more 'traditional' networks that only give the output for a few locations. The flowchart in figure 5.1 shows the steps for developing the neural network. The prediction of the maximum water levels is based of seven relevant track parameters: latitude, longitude, maximum wind speed, eye pressure, radius of maximum winds, forward speed TC

and forward direction.These parameters provide valuable information on the location, movement and intensity of the TC and will be linked in the NN to the maximum water levels.

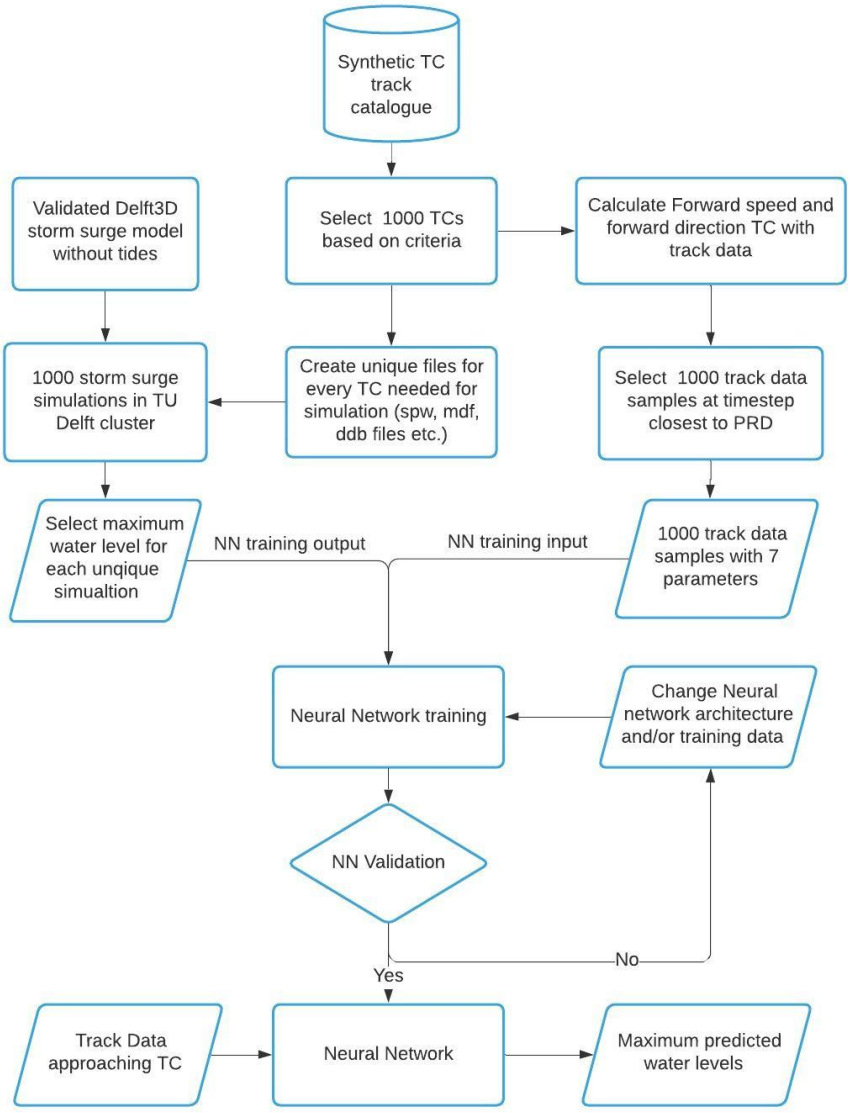


Figure 5.1: Neural Network Flowchart

5.2. Neural Network data collection and processing

NN input data

Since the aim lies at predicting the maximum water levels (peak storm tide), the input track parameters should correspond closely to the conditions at that peak time. Generally this means that the maximum water levels will occur when the TC makes landfall and/or the TC eye is close to or above the Pearl River Delta. One obvious way it to simply select the track parameters of the first time step when the TC makes landfall. This method would be sufficient if all the TC's make landfall in the PRD, however this is not the case. Although all selected storms make landfall, there landfall location can vary along the Chinese coast. The fact that not all storms make landfall in the PRD, would mean that for some storms the selected track data does not completely correspond to the time when the maximum water levels occur. Suppose we have a storm making landfall west of the Pearl River Delta. Due to the general storm direction in the western Pacific (from east to west), the storm will have passed Hong

Kong longitude wise prior to landfall. Depending on the exact location this means that the maximum water levels in the PRD occur before the storm making landfall. Thus by selected the first time step after landfall as input data, some samples will have less accurate data. One way to overcome this problem is to selected the track data not based on first landfall but by selecting the time step that the TC eye has the closest distance to the PRD. Of the considered input parameters all but two are already displayed in the synthetic TC catalogue. Only the forward speed and direction of the TC must be calculated. FD and FS are already calculated in the Matlab script that creates the spiderweb files and other Delft3d files. By making a quick modification to the script the FD and FS can be added to the track parameters for all time steps.

NN output data
The results of the synthetic simulations are stored in special history (trih) and map (trim) files produced by Delft3D. For each computational domain one trih and trim file is produced per simulation. These files contain the results for all parameters (Water level, water depth, bed level, wind speed, velocity etc.). The trih (history) files only contains the results for specified observation locations. These observation locations must be identified by the user prior to the simulation. The observation points are generally used to observe the results of the simulation in points specific points of interest. The trim files does not work with observation points but instead displays the results for the entire computational domain. For both the trih and trim file, the time interval for which the simulation results are stored can be set by the user. Since the trih and trim files are so-called data files and contain the results for different parameters, one needs additional tools to process the results. Provided with the Delft3D suite is the Quickplot toolbox, which allows for easy reading, exporting and displaying of the simulation results. Quickplot also allows to select the parameter of interest (for example water level) and export it as a .mat file. However, this action must be done manually for each simulation and it would be very time consuming to manually create a .mat file for each simulation. To overcome the problem of big data data processing, Deltares hosts a platform called Open Earth Tools with open source tools, data and models for Delft3D applications. The open earth tools also includes a section with tool for data processing in Matlab. These functions can be downloaded and added to Matlab, providing efficient and direct access to the Delft3D results. For efficient processing of the simulation results, a Matlab script is developed that uses the open earth tools and can read and save the results of up to 100 simulations at the same time. The exact workings of the Matlab script is discussed in appendix B, but in short the scripts reads the results of the water levels for all time intervals for the fine domain. From all the time intervals, the maximum water levels are selected and stored in a array. Finally, the result is an array with the maximum water levels for each simulation. These files can be combined into one data file containing the maximum water levels for all synthetic simulations. The file containing the maximum water levels for all simulations will be adapted further and is finally used in the neural network training. More information on the Matlab scripts can be found in Appendix D.

5.3. Architecture

The neural network is created with the Deep learning toolbox from Matlab. This toolbox provides the framework for designing and implementing neural networks and deep learning networks. After the input and output data has been obtained, one can start with the actual implementation of the NN. To do so, decisions must be made concerning the training algorithm, number of hidden layers, training ratios and type of activation function.

- **Training algorithm.** The NN is trained by means of a training algorithm or optimization algorithm. The goal is to minimize the error between the NN output and target output by changing the weights in the hidden neurons. Presently, a wide variety of training algorithms exist with different learning times and memory requirements. The deep learning toolbox includes about 10 different learning algorithms. For a NN with the size of figure 5.2, many of these algorithms are unusable on this problem due to the memory requirements of the training algorithms. The Levenberg-Marquardt (LM) and Bayesian regularization (BR) algorithms require too much memory for training this specific network. On the other side the scaled gradient conjugate (SCG) and the Conjugate Gradient with Powell/Bealle (SGB) will take longer time to train but can be used for training without the memory problems.

- **Number of hidden neurons.** In the field of deep learning there is no general consensus on how many hidden neurons a network should have. In many applications, the optimal number of neurons are determined by trial and error. Multiple NN's with different number of hidden nodes are trained and cross validated with the test sample set. Based on the validation performance one can select the best performing network. It must be noted that an increase in the number of hidden neurons, will also increase the complexity of the NN function in Matlab. In case the Matlab function becomes too large, it cannot be executed

- **Number of hidden layers.** A neural network with more than one hidden layer is called a deep learning network. The number of hidden layers in a NN can be related to the model capacity. The model capacity or complexity represents the number of connections or number of output parameters. The higher the model capacity the more complex function the NN can learn. This by increasing the number of hidden layers, the number of connection increase and this generally provides the ability to learn more complex functions with potentially better performance. A drawback of increasing the complexity of the NN is that it requires more data samples to learn from.

- **Training ratio.** The data samples are divided based on user defined ratios into three classes: training samples, validation samples and test samples. The training samples are only used for training the NN. The validation samples are used to validate that the network generalizing and makes sure that the training stops before over fitting. Over fitting can occur when the network keeps learning while the weights already have reached an optimum values. As a consequence, the accuracy of the network goes down. To prevent over fitting to validation samples are used to determine the MSE, if the MSE is not decreasing over a number of iterations, the training will stop. Finally, the test samples are used as a completely independent test. For learning the following ratio is used: 80% training, 10% validation and 10% test samples.

- **Machine mode or regression mode.** When the NN runs in machine mode, it will return a class label i.e. The inputs are classified toward a set of target categories. In regression mode the NN will return a specific value. Since the focus lies on water levels, the NN is ran in regression mode.

- **Training stop criterion.** The stop criterion is important to prevent over fitting of the network since it reduces the accuracy of the network. This means that the training script determines when the optimal accuracy is reached and when to stop training based on the validation checks. During the training process, the weights of the neurons are constantly adjusted in an effort to reduce the error. The generalization errors will reduce fast at the start of training gradually slowing down. At some time during training, the minimal error (optimal values for weights found) will be achieved, after which the error will start increasing again. The key to prevent over fitting in the network is to make sure to stop training if the error starts increasing again. To do so, the Matlab training script defines the number of validation checks that must fail in order to stop training. For the development of the network, the number of validation checks is set to 10. In other words, If the generalization error increases for 10 validation samples in a row, the training will be stopped. Another measure that checks for over fitting is the variance, which is defined as the difference between the training error and validation error. By considering the variance, one can gain insight in the degree of under or over fitting of the already trained network.

- **Other settings** Network settings that also are considered are the activation or transfer function and maximum number of epochs. The maximum number of epoch means the number of iterations allowed before the training is stopped. During training the maximum number of epochs is set 2000. The deep learning toolbox allows for three different transfer functions to be implemented. The transfer function is responsible for determining the network's output by taking the sum of the weighted inputs and bias as the input in the function. The types if transfer functions are: log-sigmoid, tan-sigmoid and a linear transfer function. For pattern recognition problems the sigmoid functions are often used. For function fitting the linear function is most widely used. Feed forward networks often have a sigmoid function for the hidden layers. The non linear function for the neurons in the hidden layer allows the network to learn the non linear relationships between the input and output vectors. A linear output layer is used for function fitting.

5.4. Neural Network - Water level map output

This section aims at the development and validation of a neural network that is capable of predicting the maximum water level for the entire domain. Each output neuron of the network will represent a pixel with a water level from the map. The water level map consists of a total of 60952 pixels. The enormous amount of output neurons needed for this network will lead to higher computational demand. It can be noted that the water level map also contains land areas. By removing the land areas from the NN, the total number of output neurons can be reduced to 38189 representing only the water pixels. By lowering the number of output neurons, the NN is simplified and requires less memory.

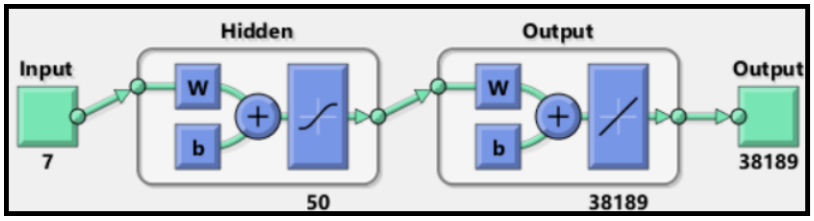


Figure 5.2: 1 layer 50 hidden nodes feed forward neural network

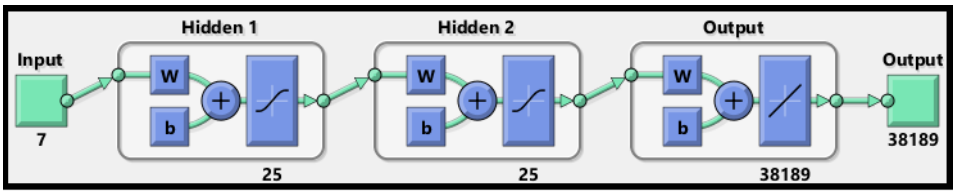


Figure 5.3: 2 layer 25x25 hidden nodes feed forward neural network

To gain insight in the most suitable network architecture to predict the maximum water levels, different configurations are learned and tested. Figure 5.2 and 5.3 display the network structure for two of the considered network configurations. The NN validation consists of different stages. First, the global errors of the networks are compared to each other. These error give an initial insight in the performance of the network. During the second phase certain track data samples are used to obtain the NN output map. This output map is then compared to the water level map of the corresponding synthetic TC. The final phase in the validation process is testing the NN prediction performance in case of an historical storm. The tidal elevation at that point in time is added to the NN results and can then be compared to the observed water level during that specified historical storm.

5.4.1. Performance indicators

Table 5.1 displays the results of the six different architectures investigated. It must be noted that single hidden layer NN's with more than 100 hidden neurons, requires too much memory to run on a normal PC. The network performance can be expressed in the mean square errors as in equation 2.7.2. This error represents the average squared difference between the target vector and the output vector. As previously discussed the samples are divided based on ratio into different classes for training purposes. The errors are calculated for each sample division by taking the average errors of the corresponding sample set.

		Performance error (m)	Training error (m)	Validation error (m)	Test error (m)
1 layer	25 Nodes	0.218	0.2177	0.2189	0.2196
1 layer	50 Nodes	0.1314	0.128	0.1322	0.1334
1 layer	100 Nodes	0.2169	0.2154	0.2240	0.2219
2 layers	25*25 Nodes	0.3924	0.3936	0.3866	0.3674
2 layers	50*50 Nodes	0.3569	0.3537	0.4176	0.3325
2 layers	25*100 Nodes	0.3887	0.3882	0.4645	0.3173

Table 5.1: Neural network performance

The results show that the 1 layer feed forward networks outperforms all double layer networks. Thus by increasing the complexity of the network the errors increase for this specific case. The NN is essentially a black-box model. Input parameters are inserted in the network and results come out with very few insight what happens in between. Due to the mathematical complexity it is hard to understand what causes the errors. To gain more knowledge on the network behaviour, other indicators that are often used in error analyses. The bias and variance give further insight in the network performance. High bias means that the output of the network is considered bad for all data i.e. the model is not fitting on the training data. This means the training error will be large. Low bias means the model is fitting well, and training error will be low. The Variance of a model is the difference between validation error and training error. It gives good insight in whether the network is under or over fitting. If the variance is high, the network is not performing well on the validation set. For network performance the variance should be low as possible. To lower the bias, one can increase the number of epochs (iterations) or increase the size of the network. The variance can be possibly be lowered by increasing the number of data samples or changing the network architecture. The quality of the samples data is also very important for network performance. When working with large data sets, there will be errors present. For example, while checking the results of random synthetic simulations, a few simulations were found that had an error in it. As a consequence these simulations stopped prior to the TC arriving, and therefore only produced maximum water levels of about 20 cm. If these samples are used in the NN in combination with the correct track data, it will lead to significant under estimations of the water level for the more intense TC's.

The highest variances are obtained for the double layer networks. Adding additional data to the samples is not an options since no more than 1000 synthetic storm surge simulations have been conducted. If there not enough data to train the so-call randomness of the initial values in the weights will remain. This will lead to incorrect results. By reducing the model size or changing the architecture, better results might be obtained. Based on this first analysis step it becomes clear that it is likely that the network has insufficient data samples for good double layer network performance. Do to the high number of connections between hidden neurons in the double layer network the non-linear function becomes more complicated and requires more training data, for higher model performance.

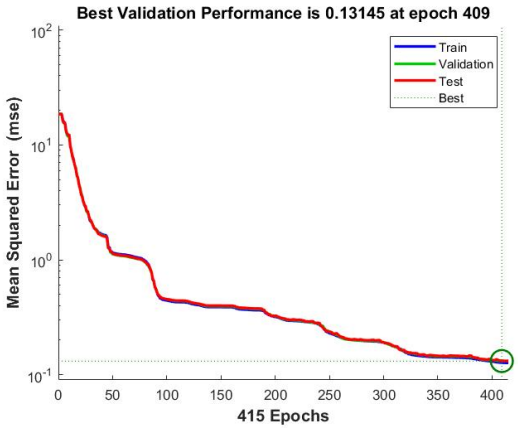


Figure 5.4: Training performance

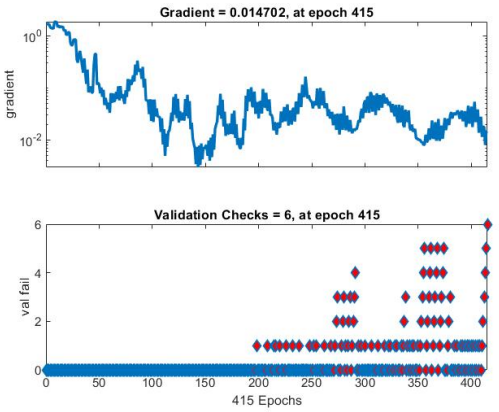


Figure 5.5: Training state

Figure 5.4 shows the NN training phase of the single layer 50 hidden nodes network. The network is trained with 892 samples i.e. the samples remaining after removing the samples with low water level elevation. At the start of the training process, the MSE was over 10m because the weights are initially assigned a random values. The error improvement occurs mostly on the first 100 epochs after the error is already reduced to approximately 0.4 meters. After the first 100 epochs, the learning rate slows down considerably. Finally after 415 epochs the network reaches the optimum values for the weights to predict the maximum water level. The gradients displayed in figure 5.5 are used during training the improve the accuracy of the network. Mathematically, the gradient is an vector that gives the direction in which the loss function increases the fastest. A high gradient value means the the weights are changing significantly. The steps in weight change will become smaller as the network approaches a

global minimum. Finally the number of validation checks is set to 10. The network must pass these check is order for the training the stop.

5.4.2. Neural Network - Model performance and validation

During this validation phase the output of the neural network is compared with the output of the synthetic simulations. Since the NN produces a water level map, it can easily be compared to the water level map output of the simulation with the maximum water levels. The first map comparison is made for a training sample. The track data of an training sample if selected. The data is inserted in the NN function that has been created by training the single layer 50 hidden nodes network. The NN function then calculates the output which only contains the results for water points. To create the water level map, the land masses must be added again to the output data. To do so a Matlab script is written that uses the locations of the original landmass points to recreate the water level map.

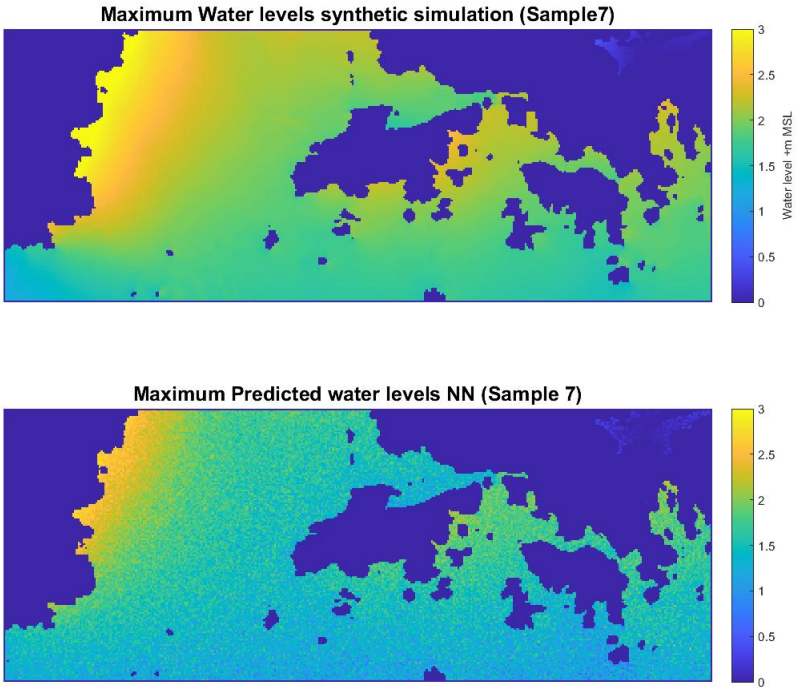


Figure 5.6: Comparison between maximum water levels of Delft3D and NN output. Training sample 7

As can be seen from figure 5.6, the predicted NN output shows a relatively 'grainy' map. This means that neighbouring grid cells have different values for the water level. This also clearly illustrates the difference between the synthetic simulation and the NN. Where the synthetic simulation recreates the physical processes that causes the water levels the change. The NN is a purely mathematical function that does not include any physical relations. The values of the output neurons that represent the water levels are independent from their neighboring neurons. When the results are transformed back to the map format, that independent behaviour becomes visible. To improve the output results an effort is made to smooth the map. For image processing often 2-dimensional convolution is used. With convolution the 2d matrix that represents the image is convoluted with a smaller matrix called the kernel, the resulting in a smoother map with less values differences between neighboring cells (figure 5.7). equation 5.4.2 shows the formulation that defines the convolution of A and B. Where A is the original matrix, B a matrix in the form $1/9 * \text{ones}(3)$ and the results are stored in array C.

$$C(j,k) = \sum_{p} \sum_{q} A(p,q)B(j-p+1, k-q+1) \tag{5.1}$$

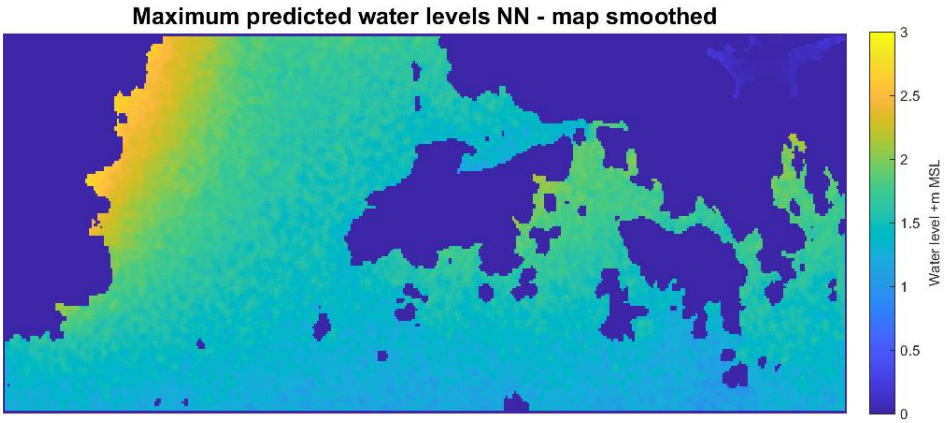


Figure 5.7: Smoothed NN output - Training sample 7

Appendix C contains more figures with different samples comparisons. To get insight in the overall performance, samples with different storm intensities are selected. The validation and test samples are also compared besides the training samples. Although the visual comparison gives good insight in the overall model performance it does not show the water level errors at a specific locations. Therefore, a closer look is taken at the locations highlighted in figure 5.8.

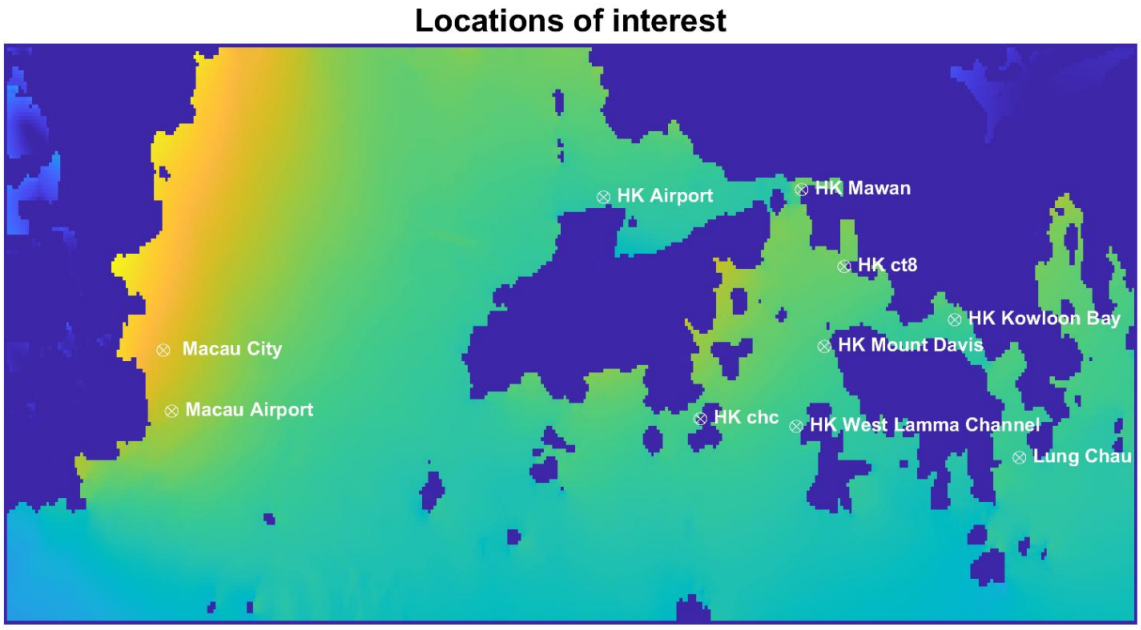


Figure 5.8: Locations of interest that are used during the neural network validation

For each of the considered locations, the error is calculated by subtracting the results of the synthetic simulation from the NN output. Besides the sample error, the relative error is also calculated. The relative error compares the measurement (NN output) with the real values (Delft3D output). This puts the error in perspective and gives insight to how far or close the NN output comes to the Delft3D output. In other words, this type of error is relative to the size of the item being measured. The relative error can be easily adapted to display the error in percentages.

$$\text{Relative Error} = \frac{|\ \text{measured - real}\ |}{\text{real}} \tag{5.2}$$

$$\text{Percent Error} = \frac{|\ \text{measured} - \text{real}\ |}{\text{real}} \times 100\% \tag{5.3}$$

	HK Kowloon Bay		HK Airport		HK Mawan		HK ct8		Macau Airport		Macau City	
Sample	error (m)	relative error (m)	error (m)	relative error	error (m)	relative error	error (m)	relative error	error (m)	relative error	error (m)	relative error
Train Sample 20	0.02	0.011	-0.36	0.163	-0.34	0.170	0.06	0.028	-0.58	0.197	-0.13	0.043
Train sample 45	-0.4	0.297	0.08	0.133	-0.10	0.119	-0.31	0.259	0.52	1.916	0.604	2.007
Train sample 154	-0.4	0.249	-0.29	0.227	-0.27	0.212	-0.489	0.280	-0.099	0.077	-0.176	0.107
Train sample 425	0.4	1.789	0.26	2.120	0.35	2.099	0.34	1.761	0.1	0.391	0.0345	0.127
Val sample 772	-0.22	0.195	0.09	0.164	0.05	0.070	-0.18	0.174	0.43	1.420	0.5	1.559
Val sample 779	-0.5	0.253	-0.46	0.269	-0.45	0.267	-0.51	0.246	-0.59	0.274	-0.39	0.167
Test sample 818	-0.0428	0.451	-0.285	0.226	-0.22	0.196	-0.2	0.353	-0.71	0.018	-0.93	0.030
Test sample 828	-0.5	0.251	-0.1	0.284	-0.51	0.291	-0.54	0.255	-0.43	0.211	-0.35	0.150
Test sample 848	-0.33	0.174	-0.49	0.265	-0.50	0.275	-0.36	0.177	-0.59	0.257	-0.304	0.121

Table 5.2: Global errors and relative errors for different samples and locations of interest - Map NN

Table 5.2 shows the results of the error analysis for the locations of interest for different samples. Negative errors imply an underestimation of the surge heights and positive values indicate an overestimation of the surge heights by the NN. It becomes clear that the range of errors is significant. Depending the sample considered, the water level location in Hong Kong show errors up to 0.5m, sometimes under estimating the surge height while for other samples over estimations are made. The model performance is worse for the two locations in Macau, where errors up 1 m are found. When considering the relative errors it becomes clear that some predictions also have large errors. The highest relative errors is obtained for sample 425 for station HK Airport with a value of 2.1 or 210%. When one compares the relative errors with the real errors it becomes clear that a small real errors does not automatically means that the relative error is also small. Overall, the results in table 5.2 shows that there is a wide range of small and large relative errors. The following possible reasons are identified as a cause for the model performance:

- The size of the output layer (38189 output neurons) is very large and increases the NN complexity significantly. as a consequence many training algorithms cannot be used and the maximum number of hidden nodes is 100 to prevent memory issues.
- Neural network training data. As discussed earlier, the NN output data consists of the maximum water levels during the entire simulation. This can be considered a simplification compared to the real world. In case of a real TC approaching, the water levels in HK and Macau will not reach the maximum valued at the same time. For most TC's, HK will experience the maximum water level earlier than Macau. The maximum water level map used in the NN, does not distinguish between the arrival time difference of the maximum water levels. Especially, when linking the NN output data with the TC track parameters this spatial difference becomes more important. The track data that is selected for the NN is the time interval with the TC eye locations closest to HK. However, this often means that the selected track parameters do no completely comply with the model for Macau. The maximum water levels at Macau occur when the TC eye is closer to Macau. This problem can be avoided when using time series in the NN. In that case, the track parameter for each time step are combined with the water level at that same time interval. This also would give insight in the propagation of the water level increase due to the storm surge. However, introducing time series would have increased the model complexity and memory required even further.
- Number of samples. It is also possible that there are simply to few samples for this network to learn the non-linear function with reasonable accuracy. Because, the sample data consists very intense and less intense storms, the maximum water level range is also large. The NN therefore requires a lot a sample data to accurately predict the water levels with large range.

The current state of the neural network is not very accurate in the prediction of the maximum water levels. In an effort to increase the NN performance, two new NN architectures are set-up. Both new architectures aim at increasing the network performance, by reducing the overall model complexity. The easiest way to decrease the complexity of the network is by removing a large part of the neurons in the output layer. For storm surge forecasting, the water levels near the coast are of most value. The

first new NN architecture only consider the water levels at the coastline. Finally, an even less complex network is developed that only considers the water levels at ten specified locations.

5.5. Neural Network - Coastline water levels

The coastline neural network uses the same track data as the previous network. From the maximum water level data, only the cells are selected that are neighboured by the coastline. The cells that represent land masses have a value 0 in the water level data set. To select the coastline points from the original data set, a Matlab script is developed that selects or deselects a cell based on their neighbouring cells. If a grid cell is surrounded by one or more cells that represents land, it will be selected by the script to form the coastline. The coastline that covers the entire grid has a total of 2795 grid cells. The resolution of the coastline is the same as the resolution of the fine computational domain used in the Delft3D simulations i.e. one grid cells represent an area of 225 by 225 meter. Figure 5.9 shows the grid cells that are selected to form the coastline. Due to the number of small and large islands in the PRD, the entire coastline still consists of almost 2800 grid cells. In an effort to reduce the number of output neurons even more, the current resolution of the coastline is reduced by a factor 5, generating a new resolution of approximately 1100m. By reducing the coastline resolution, the number of output neurons are reduces to 590 output neurons in total. The consequence of reducing the resolution of the coastline is the fact that the coastline grid cells are no longer connected to each other. Therefore, the coastline is dotted in figure 5.10. By reducing the resolution of the coastline, it is possible to significantly reduce the complexity of the neural network without losing the water level prediction capability for the domain.

Figure 5.9: Extracted Water level grid cells that form the coastline of Hong Kong and Macau (+m MSL)

5.5.1. Model Performance and Validation

The validation process for the coastline model is largely the same as for the map network from the previous section. Table C.1 in appendix C shows the network training results for different configurations. The 100 and 200 nodes networks have the best overall performance. Increasing the number of hidden nodes more, results in a decrease of the model performance. The two best performing networks are further investigated by determining the errors between the network output and training samples. Table 5.3 and 5.4 are displayed below with the results of the validation of the 200 nodes coastline network. The table validation tables for the 100 nodes network can be found in Appendix C.2.

error (m) Sample Number	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mt Davis	West lamma Channel	HK Airport	HK Lung Chau	Macau Airport	Macau City
7	-0.183	-0.235	-0.160	-0.146	-0.163	-0.127	-0.156	-0.158	-0.133	-0.136
20	-0.060	-0.083	-0.159	-0.072	-0.072	-0.131	-0.167	-0.310	-0.404	-0.442
45	-0.034	-0.004	-0.027	-0.077	-0.031	-0.067	-0.066	-0.027	0.091	0.128
154	-0.048	-0.075	-0.131	-0.035	-0.088	-0.128	-0.054	-0.057	0.280	0.126
425	0.073	0.066	0.066	0.079	0.058	0.089	0.081	0.094	0.134	0.161
772	0.048	0.123	0.093	0.047	0.111	0.081	0.000	0.046	0.179	0.297
779	0.142	0.133	0.009	0.138	0.128	0.041	0.057	0.009	-0.280	-0.121
796	0.010	-0.065	-0.073	-0.071	-0.011	-0.057	0.004	-0.191	-0.024	-0.080
818	-0.237	-0.267	-0.224	-0.187	-0.208	-0.189	-0.274	-0.023	-0.017	-0.096
828	-0.080	-0.115	-0.055	-0.073	-0.052	-0.084	0.002	-0.032	0.030	0.033
848	-0.093	-0.081	-0.167	-0.054	-0.060	-0.125	-0.101	-0.059	-0.337	-0.170

Table 5.3: Sample errors (m)- Coastline Network, 200 Nodes

The results show that in general the network is capable of predicting the maximum water levels with reasonable accuracy. However, it must be noted that the accuracy can vary significantly over the samples. For some samples, very low errors are obtained while for other samples the errors can be significantly. The results also show that the maximum water level predictions are more accurate for the coastline in Hong Kong than for Macau. When comparing the results to the previous network that predicts the water level for the entire domain, some aspects stand out. Firstly, the reduction of the number of neurons in the output layer directly influence the number of hidden neurons that can be used. For the coastline network, a hidden layer with more than 2000 nodes is possible, while for the map network the maximum number of nodes in the hidden layer was 100. Secondly, it is observed that by reducing the complexity of the network, the training performance is improved significantly. The training, validation and test errors are reduced substantially compared to the map network. Despite, the significant reduction in the complexity, the coastline network is still too complex to be trained with other training algorithms than the scaled conjugate gradient (SCG) algorithm.

When the mean errors over the selected samples are calculated, it becomes clear that the 200 nodes network has a slightly better performance than the 100 nodes network. For the 200 nodes network, the mean error is -0.0587 m while the 100 nodes network has a mean error of -0.08018 m. The negative values mean that the neural network is underestimating the maximum water levels. It must be noted that for most samples the network is underestimating the maximum surge height. On the other hand, a few samples are identified that are overestimated by the network.

Relative Error Sample	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mt Davis	HK West Lamma Channel	HK Airport	HK Lung Chau	Macau Airport	Macau City
7	0.084	0.108	0.079	0.073	0.079	0.065	0.082	0.082	0.056	0.044
20	0.027	0.038	0.073	0.036	0.034	0.064	0.078	0.156	0.138	0.130
45	0.028	0.004	0.029	0.056	0.027	0.063	0.109	0.025	0.353	0.423
154	0.029	0.045	0.084	0.021	0.053	0.083	0.042	0.036	0.273	0.073
425	0.412	0.371	0.290	0.351	0.266	0.387	0.625	0.301	0.576	0.673
772	0.045	0.115	0.121	0.040	0.113	0.092	0.000	0.048	0.601	0.584
779	0.070	0.065	0.005	0.069	0.064	0.021	0.034	0.004	0.128	0.047
796	0.008	0.055	0.064	0.069	0.010	0.054	0.003	0.191	0.014	0.038
818	0.196	0.220	0.205	0.179	0.193	0.189	0.259	0.029	0.012	0.046
828	0.038	0.054	0.028	0.036	0.026	0.044	0.001	0.017	0.015	0.013
848	0.045	0.039	0.085	0.028	0.030	0.066	0.055	0.032	0.145	0.064

Table 5.4: Relative errors - Coastline network, 200 Nodes

The relative errors are also lower for the coastline network. A substantial part of the samples have relative errors below 10 % which can be considered accurate. However, as discussed before for some samples a significant error is observed of up to 60 %.

Figure 5.10: Maximum predicted water levels Neural Network output (+m MSL)

5.5.2. Validation for Historical TCs

The final step in the NN validation is to check whether, the NN is capable of predicting the water levels for an historical TC. By taking the track data from TC Hato, and adding the tidal elevation to the NN output, the networks performance on historical storms can be checked. Like for the storm surge validation, TC Hato, Mangkhut and Hagupit are considered again. For each TC, the track data is selected from IBTrACS. The closest eye location to Hong Kong is then used in the NN function. The results of the NN are stored and the corresponding tidal elevation at that time is exported from Xtide. The tidal elevation is added to the NN results and are then compared to the observed maximum water levels during these TC's. It must be noted that this comparison can only be made for the three tidal stations (ct8,chc and mwc), due to lack of data for other locations.

Tide station and storm	NN results (+ m MSL)	\Tidal elevation (+m MSL)	NN combined with tides (+m MSL)	Observed water levels (+m MSL)	Error (m)
chc (Hato)	1.41	0.75	2.16	2	0.16
ct8 (Hato)	1.61	0.75	2.37	2	0.37
mwc (Hato)	1.6	0.75	2.35	1.95	0.4
chc (Mangkhut)	1.97	-0.05	1.92	2.23	-0.33
ct8 (Mangkhut)	2.15	-0.05	2.10	2.13	-0.03
mwc (Mangkhut)	2.14	-0.05	2.09	2.05	-0.04
chc (Hagupit)	1.96	0	1.96	2	0.04
ct8 (Hagupit)	1.97	0	1.97	2	0.03
mwc (Hagupit)	1.95	0	1.95	no measurement	

Table 5.5: 200 Nodes network - Performance for Historical TC

The network has the poorest performance for the prediction of the maximum water level for TC Hato. The largest error for Hato is obtained for station ct8 and mwc and are up to 40 cm. On the other hand, the predictions for Mangkhut and Hagupit are much better. It is likely that the training data set did not include input parameters that closely correspond to the track data for TC Hato. As a consequence, the network is not trained well, to solve for these parameters. This illustrates the importance of having a large variety of input parameter values during training. If there is only a small variety of the values in the training track data set, the network will only be capable of predicting accurate values for input parameters that are close to the values used in the training data set.

Figure 5.11: Neural Network prediction for TC Mangkhut track parameters (+m MSL)

5.6. Neural Network - water levels for 10 locations

The final neural network that is developed also has the lowest complexity. This NN will only predict the maximum water levels at 10 locations instead of for the entire map. The selected locations are near the city centres, airports and other relevant locations. By not generating a map or coastline output, the NN is significantly less complex, allowing for different training algorithms and new configurations. For the NN with 10 output locations, multiple configurations are tested again.

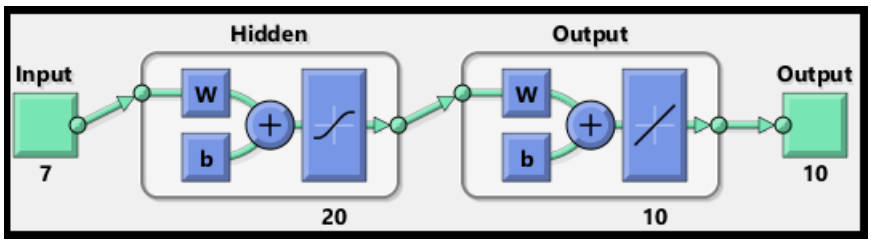


Figure 5.12: Neural Network architecture

In total 15 different configurations are trained. The results are displayed in table C.4 in appendix C. Reducing the complexity of the NN by only training for the locations of interest causes a significant reduction is the global errors. For all considered configurations the errors are reduced compared to the NN water level map version. The reduction in errors is most likely due to the reduction of complexity. Furthermore, instead of only using the SCG algorithm for training, others can be used as well. In total 18 different configuration are trained. There are multiple configurations that produce low errors. To selected the best NN, the most promising configurations are analysed further by calculating the errors for the locations of interest. For validation purposes, two of the best performing networks are analysed: the 30 nodes SCG network (Table 5.6, 5.7) and 20 nodes LM network for which the results can be found in appendix C.3.

Error (m) Sample	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mount Davis	West Lamma channel	HK Airport	Lung Chau	Macau Airport	Macau City
7	-0.12	-0.07	-0.07	-0.02	-0.09	-0.05	0.02	0.01	0.03	0.04
20	-0.22	-0.11	-0.29	-0.13	-0.28	-0.29	-0.16	-0.35	-0.28	-0.32
45	-0.13	-0.12	-0.10	-0.15	-0.12	-0.12	-0.09	-0.07	0.17	0.13
154	-0.19	-0.21	-0.24	-0.16	-0.21	-0.21	-0.10	-0.17	0.11	0.06
425	0.16	0.17	0.07	0.19	0.12	0.09	0.06	0.12	0.03	0.04
772	-0.02	0.05	0.09	-0.01	0.05	0.04	0.01	0.00	0.20	0.16
779	0.06	0.02	-0.07	0.01	0.02	0.00	-0.06	-0.05	-0.33	-0.24
796	-0.05	-0.02	-0.11	-0.03	-0.11	-0.13	-0.08	-0.19	-0.17	-0.24
818	-0.29	-0.29	-0.27	-0.18	-0.21	-0.22	-0.13	-0.09	-0.14	-0.20
828	-0.13	-0.21	-0.13	-0.21	-0.16	-0.14	-0.05	-0.18	0.04	0.09
848	-0.04	-0.05	-0.11	-0.04	-0.01	0.01	-0.16	0.16	-0.30	-0.26

Table 5.6: Water level errors for 10 locations - Single layer, 30 hidden nodes network, SCG training algorithm

Relative error Sample	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mount Davis	West Lamma channel	HK Airport	Lung Chau	Macau Airport	Macau City
7	0.056	0.032	0.034	0.010	0.044	0.026	0.010	0.005	0.012	0.015
20	0.100	0.051	0.133	0.064	0.132	0.140	0.073	0.176	0.088	0.102
45	0.112	0.094	0.108	0.111	0.104	0.119	0.149	0.068	0.624	0.447
154	0.116	0.119	0.153	0.097	0.127	0.136	0.078	0.109	0.084	0.039
425	0.848	0.852	0.306	0.826	0.548	0.373	0.465	0.377	0.117	0.157
772	0.019	0.046	0.117	0.009	0.052	0.048	0.018	0.000	0.652	0.489
779	0.030	0.010	0.036	0.005	0.010	0.000	0.035	0.025	0.151	0.103
796	0.043	0.018	0.095	0.029	0.100	0.121	0.061	0.189	0.097	0.127
818	0.245	0.242	0.246	0.174	0.195	0.222	0.125	0.113	0.090	0.111
828	0.063	0.097	0.067	0.104	0.079	0.074	0.028	0.100	0.020	0.040
848	0.020	0.024	0.056	0.021	0.005	0.005	0.086	0.087	0.129	0.105

Table 5.7: Relative errors - Single layer, 30 hidden nodes network, SCG training algorithm

The Single layer 30 hidden nodes network for which the error results are displayed in table 5.6, shows that for the 10 considered sample the prediction errors for the locations of interest stay within a

maximum underestimation of 0.3 and a maximum overestimation of 0.2 m. For some samples the NN shows a slight underestimation of the maximum surge height for the locations. The prediction for the locations in Macau still have poorer performance than the locations in Hong Kong. As stated before, this can be related back to the way the track data and maximum water level were selected. When considering the relative error some aspects stand out. Firstly, for the majority of the samples, the relative error is reasonable. However, some samples show differences between the errors for the locations. This means that the errors change over the locations within the same sample. For examples, the errors for Macau airport and City have significant errors for some samples. By introducing time series in the NN, the error in Macau would have probably been reduced even more. Another trend that became visible during the analysis is the fact that the degree of surge underestimation generally increases for more intense TC's. For simulated events with high surge heights the NN will often underestimate the surge height slightly. When comparing the performance of the map NN with the 10 locations NN it becomes clear that by reducing the complexity of the network, significantly increases in accuracy have been obtained for this case. While for the map NN, errors up to 0.9 m were observed, this new network outperforms the old for every tested sample. For actual maximum water level prediction, the locations NN, is capable of producing more accurate results. Additionally, it is relatively easy to add more locations to the network to include more areas of interest.

5.6.1. Network validation for historical TC

The same as for the coastline network, the current network performance is checked for the historical TCs Hato, Mangkhut and Hagupit.

Tide station and Storm ID	**NN result (+m MSL)**	**Tidal elevation (+m MSL)**	**NN combined Water level (+m MSL)**	**Observed water level (+m MSL)**
ct8 (Hato)	1.35	0.75	2.2	2
chc (Hato)	1.26	0.75	2.11	2
mwc (Hato)	1.36	0.75	2.21	1.95
ct8 (Mangkhut)	2.03	-0.05	1.97	2.13
chc (Mangkhut)	1.81	-0.05	1.76	2.23
mwc (Mangkhut)	1.90	-0.05	1.85	2.05
ct8 (Hagupit)	2.1	0	2.1	2
chc (Hagupit)	2.016	0	2.016	2
mwc (Hagupit)	2.0	0	2.0	No measurement

Table 5.8: NN network validation for three historical TC's

The results show that if the tidal elevation is added to the NN output, the maximum predicted water levels errors are below 0.3 m for all considered TC's. This current method gives good insight in the maximum water levels that can be expected in case of an approaching TC. However, The NN output is completely trained on 7 input parameters. Therefore, the correct selection of the track data is of the upmost importance. For applicability, the user should always select the track data point that comes closest to Hong Kong. If other track data points are used, the accuracy of the NN will go down considerably. Since in the NN, the model is purely mathematical, a slight change in the input parameters, can have significant consequences for the NN output. The accuracy of the maximum water level prediction is therefore for a large part dependent on the quality of the track data.

6

Discussion

This chapter focuses on discussing the methodology followed and results obtained during this Thesis research. The discussion is divided into two main parts: the hydrodynamic model and neural network.

6.1. Hydrodynamic storm surge model

The storm surge model discussed in chapter 3 has been developed for the purpose of generating data, that can be used in the neural network. Since the results of the storm surge model are used as input in the second model, it is important that the results of the storm surge model are sufficiently accurate to be used in the neural network. An inaccurate storm surge model will automatically lead to an inaccurate neural network. To make sure that the results of the storm surge model are accurate, an emphasis towards understanding the physical processes and a detailed validation process has been made during development. The performance of the storm surge model itself is dependent on a large number of aspects which are discussed in the next part.

6.1.1. Data availability and quality

The data on that is used by the model like bathymetry, topography, storm track data, tidal data etc. is very important for the overall performance of the model. The GEBCO database provides the information of the bathymetry in the South Chinese sea, with a resolution of 500m. For open oceans, this resolution is considered sufficient. However, for the coastal region in the PRD, a higher resolution is preferred to account for all the local bathymetry changes. Therefore, instead of using GEBCO for the PRD region, a local bathymetry data set is used that provides more detailed (higher resolution) information on the bathymetry. The bathymetry directly relates to the water depth, which has a influence on the physical processes. Although to a lesser extent, other aspects like river discharge and precipitation runoff also influence the water levels. Collecting the required data to model these processes proved to be impossible. Therefore, the decision was made to exclude river discharge and precipitation from the model. Including these physical processes into the hydrodynamic model might in increase the model's accuracy.

For validation purposes it is necessary to collect actual observed water level and wave data from measurement stations. In the PRD there are many tidal gauges and wave buoys operated by different local authorities. For this research it was possible to collect the water level data for three stations near Hong Kong and two wave stations. The water level data is the main source for validation. The quality of the water level data was sufficient but sometimes a measurement buoys malfunctioned during a storm preventing a complete validation of the model. The wave data obtained from the buoys displayed no data during the period of all three validation storms (Hato, Mangkhut and Hagupit). Since critical data on the waves was missing from the data set, the storm surge model with waves could not be validated. Based on the data aspects discussed, it becomes clear that data availability and quality of very important for physical modelling. Missing data sources lead to the need for making assumptions and simplifications to the model or prevent a complete validation process.

6.1.2. Physical parameters

There of many physical parameters that must be calibrated in order to simulate the processes accurately. These parameters are identified in the sensitivity analysis (Appendix B). The main purpose of this analysis is to investigate the influence these parameters have on the result of the model. The parameters investigated in this appendix are: tidal constituents, wind drag coefficients and TC parameters.

- **Tidal constituents.** The tidal constituents are the forcing on the boundary of the model. In order the reproduce the actual tidal changes, these constituents must be chosen correctly. In the analysis, the output of three tidal simulations from different tidal models is compared to the observed tidal changes at the measurement buoys. The considered tidal models produce results that are in phase with the observed tidal signal. In terms of the tidal amplitude, the models show an increase in the error around neap and spring tide but overall the amplitude are in line with the observe values. The average error for a tidal simulation of 30 days is 12 cm, which is considered sufficient for storm surge modelling. It is not completely clear why the error increase during neap and spring tide. A strong possibility is the fact that some of the tidal constituents used may not be entirely correct. An harmonic analysis can be conducted to find a possible error in the tidal constituents and improve the model results.
- **Wind drag coefficients.** A very important physical parameter that can influence the results significantly is the wind drag coefficient. This drag coefficient is used in the equation to calculate the stress (wind shear stress) exerted by the wind on water. The value of this coefficient changes for changing wind speeds. Over the years, multiple relationships have been developed to determine the correct value for the coefficient. Some of the drag coefficient relations are linear while other follow a hyperbolic path. Additionally Delft3D only allows the user to enter drag coefficients for three different wind speeds. The values in between are linearly interpolated. The sensitivity analysis shows that the best model results are obtained for the wind drag relationship specially developed for the South Chinese sea by (Peng and Li, 2015).
- **Tropical cyclone parameters.** The TC track parameters form the input of the spiderweb grid calculation and consequently have a direct influence on the wind forcing within the model. There are multiple TC agencies that register the track data of historical TC's. These agencies all have slightly different methods to measure/determine these parameters. During the sensitivity analysis, track data from the different agencies are tested. IBTrACS lists the data from all the different TC agencies. The ones that are considered in the sensitivity analysis in appendix B.3 are the JWTC (USA), Tokyo agency (Japan) and CMA (China). Some data like the latitude,longitude, eye pressure en max wind speed can be determined from satellite imagery. However data on the RMW is not registered by all agencies and must therefore be calculated in some cases. The sensitivity analysis showed that the different methods for determining the RMW produce significant difference in the water levels for the model output. The influence of the TC parameters illustrates the need to investigate the influence these parameters have on the output.

6.1.3. Model validation

The model validation process consists of simulating three historical TC's and comparing the results with the observed data. Based on the validation it can be concluded that the model is capable of predicting the water levels correctly. Since the focus lies on the maximum water levels, the errors are also calculated for the maximum water level that occur during a storm. For the three considered storms, the error for the maximum water level is below 20 cm for the three locations. This translates to a percentage error below 10% and is considered sufficient for storm surge purposes. It must be noted that the three tidal measurement stations are all located in Hong Kong. For Macau, which is located approximately 70 km from Hong Kong no tidal data is available. Therefore, the results of the simulation for Macau cannot be completely validated, but will be used.

The last step in the validation process is a simulation of a storm surge with waves. Since the focus is not on coastal flooding, there is no interest in the impact of the on coastal areas. However, waves can cause an increase in water level in shallow water due to wave setup. Since a large part of the PRD water depth is considered deep, it is expected that wave setup will have a minimal effect on the water

level in the areas of interest. The simulation shows indeed, that waves have a minimal contribution to the water levels. It must be noted, that wave-setup plays a much more significant role in shallow water areas close to the coast. But since we are not looking at overland flooding, this is neglected. A additional reason to neglect waves in the model is the fact that wave data collected from the two buoys, all have data removed during the occurrence of Hato, Mangkhut and Hagupit. It is therefore not possible to validate the wave parameters like significant wave height, length and period.

6.1.4. Limitations hydrodynamic model

Like any other representation of a physical process in a numerical model, it is not possible to reproduce the real world with 100 percent accuracy. There are always some trade-offs and neglections to be made in an effort to reproduce the real world accurately.

- Model does not include precipitation and run-off from land.
- Model does not consider discharges from the rivers connection to the PRD
- Delft3D only allows linear relationships for the wind drag coefficients. Recent studies have suggested the use of hyperbolic relationship for the drag coefficients. This hyperbolic behaviour cannot be reproduced in Delft3D.
- No complete water level validation for Macau due to missing water level observations for tidal buoys
- No wave validation due to missing obseverd data.
- Spring and neap tide not captures completely correct in model. Opportunity to improve tidal signal by means of a tidal analysis of harmonic constituents

6.2. Neural Network

6.2.1. Data selection

The performance of the network is for a large part determined by the training data. As can be expected for large data set, errors were found in some of the 1000 synthetic simulations. Some of the simulations were forced to stop before the stop time. Since it is very time consuming the manually check all these simulations for errors, a threshold criteria was set-up to filter the simulations for which the maximum water level results hardly changes compared to the simulation. All simulations that produces maximum water levels below 20 cm compared to the initial level were deleted from the sample set. This ensures that the samples that did not experience any significant surge heights are not included in the network. A a consequence only 892 samples remain in the training data set. Of these removed samples not all simulations had errors, some simulation simply induces a low surge height. To investigate the effect of the sample change, a network with all samples and 892 samples are trained exactly the same and the result compared. Analysis shows indeed, that although the number of samples are reduced in the 892 sample network, the performance was significantly better than the original network.

At first instance the development of the NN focused on developing a network capable of producing a map with maximum predicted water levels for TC track parameters. For this version 38189 output neurons are needed to represent the water level in each pixel. The development of a network of such complexity comes with problems. The size of the output layer, means there are many connections between the nodes in the hidden layer. This increases the memory demand during significantly. As a consequence some of the most used training algorithms cannot be used. Furthermore, due to the already complex network, the size of the hidden layer is also limited. In Matlab, networks up to 100 hidden neurons can be trained before running in to memory problems. However, a network with many output neurons generally also requires many hidden neurons to accurately solve the non-linear problem. Due to the size of the output layer it is not possible to test preferable configuration which is limiting for networks accuracy.

6.2.2. Network Validation

In chapter 2.7, three different neural network architectures have been developed. The purpose of developing different networks, is to gain insight in the training behaviour en determine to best suited configuration for maximum water level predictions caused by storm surges. First, a 'complex' network was developed capable of predicting the water levels for the entire domain. The second network only considers the water level at the coastline, reducing the complexity of the network significantly. The final network developed is a very simple NN, only capable of prediction the maximum water levels at 10 locations surrounding Hong Kong and Macau. The three networks, are analysed in detail.

NN - water level map output

Although there are some significant limitations to this network configuration, the network is capable of producing predicted water level maps. If the output maps are compared to the output map of the synthetic simulation, the limitations come to light. Visual inspection of the maps shows that for sample case the network can predict the water levels with a wide variety of error sizes. For some samples the predicted values are relatively good. While for other samples the errors a large and up to 1 meter. Especially the predictions for Macau are very poor.

The most likely reason for the inaccurate performance can be linked back the to number of nodes in the hidden layer. The number of hidden nodes are limited and as a consequence, the network is not capable of accurately predicting the water levels. It is shown, that by reducing the models complexity significant improvements can be made in the predictions for the same number of training samples. Additionally, it is likely that the performance of the network will significantly increase if the number of hidden nodes are increased. However, this requires way more computing power than a normal PC can deliver. In the future it might be worth it to try to improve the network for example by running the training process in a computing cluster. Another possibility for the relative poor performance of the

map NN is the way the in and output data is selected. The network does not use time series. As discussed in chapter 5.2, this means that for the input data only the track parameters of the time step with shortest distance to HK is considered. The output data selected represent the maximum water levels that occurred during the entire simulation. The fact that we are not working with time series means that the maximum water levels selected did not all occur at the same time. For example for a TC passing from East to West, Hong Kong would have experienced the TC before it arrives at Macau. Consequently the maximum water levels in Hong Kong would have occurred earlier. The track data selected corresponds more to the situation in Hong Kong. This often means that the selected track parameters do no completely comply with the model for Macau. The maximum water levels at Macau occur when the TC eye is closer to Macau.

NN- water levels coastline output

This network has a significant better performance than the previous network. By reducing the number of output neurons, it is possible to increase the number of nodes in the hidden layer. Although it was expected that the networks performance would increase with an increase in the number of hidden nodes, this was not completely true. The analysis in section 5.5 shows that the best performance is obtained for a 200 Nodes single layer network. Increasing the number of hidden nodes to over 200, leads to a reduction in the performance. Apparently, if more than 200 nodes in the hidden layer are used for this specific problem, the number of connections between the nodes in different layers becomes too large and ultimately will decrease the overall model performance. Although the networks complexity has been reduced, the LM and BR traing algorithm still require too much memory to be used. Therefore, this network is also trained with the SCG algorithm. When comparing the network output with the training data, there is no real trend visible in whether the network over or under estimates the maximum water levels. Generally speaking, the ratio of under and over estimations is almost equal for the 200 nodes network.

The output of the network is also checked for three historical TC's. We know for sure that the network input parameters for these storms are not used in the training process. The performance of the network for historical storms will show whether the network is capable of making predictions for input parameters that the network has never seen before. The results in table 5.5 shows that the 200

nodes network can accurately predict the maximum water levels for TC Mangkhut and Hagupit for most locations. On the other hand the predictions for TC Hato are of less quality. For this storm, over estimations up to 40 cm are observed.

NN - water level locations output
In an effort to improve the network even further, a change towards a more traditional network is made. Instead of generating a map or coastline network, this new configuration only has the maximum water level output for 10 locations of interest. The reduction in the complexity of the network allows it to be trained with more types of algorithms. Since there are less output neurons, the number of connections between the nodes is reduced enormously. This implies that the non linear function the network tries to create can also be less complex.

The results of the trained network shows an maximum error of 30 cm between the output data and actual maximum water levels. Again the largest errors are obtained for the locations in Macau, but they have improved considerably since the previous configuration. By testing multiple samples it becomes clear that generally speaking the errors are smaller for less intense TC's. For very intense TC's the network tends to underestimate the surge heights. When comparing the performance indicators of the coastline network with the 10 locations network, it becomes clear that a lower MSE is obtained for the coastline network. The performance indicators only give information on the overall performance of the network by calculating the mean error over all the output nodes. This does not give any information about the variability on the performance over the different samples. The coastline may have a smaller MSE than the 10 locations network, but this does not directly means that it can make more accurate predictions. For the coastline network, the MSE is calculated over 2795 nodes while for the locations network the error is only calculated over the 10 nodes. Due to the large number of nodes, there can be nodes with large errors, but due to averaging of the error over all the nodes, this will have a limited effect on the MSE. For the 10 locations network however, one or two nodes with large errors will already have a significant influence on the MSE. The validation process shows that the coastline network in general has larger errors than the 10 locations network.

Based on the validation process, it can be concluded that the 10 locations network is able to predict the maximum water levels with accuracy for most samples. A downside of this network, is the fact that the maximum water levels are only predicted for 10 locations, thus it gives a limited amount of information. The coastline network, is capable of making sufficient predictions in most cases. However, for some combinations of input parameters significant errors are observed, which makes is less usable for for real world predictions. Finally the map network, provides the most information on the water levels. However, this network is also the most inaccurate and is not suitable for predictions in it's current state. Due to time constraints it was not possible to further increase the performance of the considered networks.

6.2.3. Neural Network limitations

- Network configuration limited by computational memory. As stated multiple times, the maximum complexity of the network is dependent on the computational capacity of the PC.
- Black box model. Hard to identify the reasons the network behaves like it does. Purely mathematical model, hard to distinguish the effect of configuration changes.
- Results not dependent on physical processes. The calculation of the network is purely mathematical en neglects any physical processes completely. The performance of the network is therefore completely dependent on the training data and training process. It is possible that the network produces results that are physically impossible. To prevent this from happening, extensive validation is required for both the training data of the network en the output performance.
- Large amounts of data needed for training the network. The performance of the network is completely dependent on the training data. Quantity and quality of the data used is very important. More samples used for training leads to more variety in the training data set. The network is therefore, better capable to train for a wide variety of input parameters leading the more accurate

results. If a limited number of samples are used during training the network will only perform well on the input parameters that have been used for training.

- No times series data are for training neural network. This means that the maximum water level used in the training output data, don't all occur at the same in real life. This can have an effect on the accuracy of the predictions and prevents the network to be able to predict arrival times. Introduction of time series in the NN, will also lead to the possibility of predictions of the arrival time of the maximum surge. it must be noted that introducing time series will also have severe consequences for the required memory. It is likely, that normal PC's are unable to handle the training of a map or coastline network in combination with time series.

7

Conclusion and Recommendations

This chapter summarises the conclusion of this study by relating them back to the research questions as displayed in chapter 1.

Main research question:

- Is it possible to develop a neural network using synthetic storm data that is capable of accurately prediction the maximum water levels in case of an approaching tropical cyclone?

Sub research questions:

- What is the best way to set-up and calibrate a hydrodynamic model accurately for a large number of synthetic storm simulations?
- How should a synthetic storm database be implemented to efficiently process large amounts of data for use in synthetic storm surge simulations and neural networks?
- What are the advantages and or disadvantages of using neural networks to predict maximum storm surge heights, compared to traditional hydrodynamic modelling?

7.1. Key findings

7.1.1. Set-up and calibration hydrodynamic storm surge model

During this research, the hydrodynamic model forms the basis for the production of reliable data in the neural network. Since the NN uses the results from the hydrodynamic model it is of the upmost importance to calibrate and validate the surge model in a way it represents the real world situation as close as possible. The first aspect to be considered is the size of the computational grid. The grid cell resolution is a trade-off between the level of detail and the computational demand. For the open ocean a grid cell resolution of 2 km is used while at the location of interest the resolution increase to 220 m. The Bathymetry and topography are combined from different sources with various resolution to obtain one bathymetry and topography map.

The calibrations of the model is conducted based on many different simulations with slightly different settings. An effort towards understanding the sensitivity of important parameters have been made. The analysis shows that it is important to calibrate these parameters for their specific purpose. Especially the wind drag coefficients and RMW parameter has a significant influence on the storm surge height and must be used with care. There are many methods in the literature to derive these values but an sensitivity analysis is always preferred to ensure sufficient knowledge in the behaviour of the parameter of interest.

For a correct validation of the different physical processes in the model one should divide the model into different steps: tidal validation, storm surge validation and storm surge with waves. For the tidal validation, the observed water level are compared to the tidal elevations in the model. A tidal simulation of 30 days was conducted to ensure that the model is capable of capturing the daily and monthly changes in the tidal elevation. To validate the performance of the model for different storm situations, three historical storms are tested in the surge model: Hato (2017), Mangkhut (2018) and Hagupit (2008). For these historical storms, observed water level data is available ensuring that a comparison between the model and real world can be made. Furthermore, by validating the model for multiple storms, one can ensure that the model performs well for different storm situations. Overall, the validation process showed that the model is capable of predicting the surge heights with good accuracy in case of a TC passing the PRD. During the final validation step the storm surge model is coupled with a SWAN wave model and the effect of wave setup was investigated. The simulations show that waves are responsible for a negligible increase in the water levels for the locations of interest. Depending on the purpose of the model, waves can be included. For the storm surge model it can be neglected due to the very low influence on the water levels.

When considering the entire validation process, one can conclude that the storm surge model is capable of predicting surge height with a maximum error of 20 cm for the considered cases. In summary, for the efficient setup of a storm surge model used for many simulations, sufficient data must be available. Additionally, it is important to gain knowledge on the effect certain parameters have on results of the model. Validation should be conducted by testing the physical processes separately and combined in order to distinguish between the physical processes. Furthermore, validation for more cases, ensures that the model is capable of producing usable results for different storm situations.

7.1.2. Implementation of synthetic storm data

When referring back to the handling of the synthetic storm database, efficient handling data handling and automatization of processes becomes important. Due to the size of data, it would take way too many time to manually prepare all the synthetic simulations. During this research multiple Matlab script have been produced with the goal of automatizing as much as possible processes. Although, these scripts are set-up for this specific case, they can easily be adapted for future research purposes on other subject and locations. The most important tools developed during this research for synthetic data handling and processing are listed below.

- **TC selection tool.** This tool can be applied to the synthetic catalogue for all storm basins. With this script, the track data of storms can be selected based on criteria imposed by the user. This script provides the possibility to find and select storm of interest from a large database and eliminates the need to select storms manually. This script is especially use full in the case that many synthetic storms must be simulated.
- **Synthetic storm Delft3d file creator tool.** This tool is the main tool that creates all the required Delft3D files for the synthetic simulations automatically. It uses the track data of the storms selected in previous tool to create the spiderweb, mdf, ddb and shell files which need different configurations for each unique simulation. This tool enables researchers to quickly prepare multiple simulations in Delft3D. Although the script is currently set-up for this specific research, it can be easily adapted for different Delft3D model setups and for locations all around the world.
- **Delft3D output processing tool.** This tool is developed to increase the efficiency of processing Delft3D output files. Rather than using Quickplot and manually export the results for each locations, the Matlab script used Open Earth Tools functions to load and store the results up to 100 simulation at the same time.

Another aspect that is solved in order to work with synthetic storm data in Delft3D is the fact that the synthetic storms are not assigned a specific date. Delft3D requires a start date and stop date to simulate to correct tidal elevation at that time. To overcome this problem, it is decided to remove tidal forcing in the synthetic storm surge model. Now, the date of occurrence of the storm does not matter anymore. As a consequence for accurate water level predictions, the tidal elevation should be added manually to represent the actual water level. The Matlab tools that have been developed of the course

of this thesis research can be easily used and adapted for other locations and purposes. Efficient data handling is guaranteed with these tools.

7.1.3. Hydrodynamic model versus neural network

Unlike the hydrodynamic model the neural network is purely based on mathematics and does not include any physical processes. The behaviour of the output cells is independent, meaning that the value of a output cell is not dependent on the neighbouring cell. This is a disadvantage for the NN capable of producing the water level map. The results show that there can be significant water level difference between neighbouring grid cells. This would have been impossible in storm surge model, where the physical processes are included. This also illustrated the need for high quality training data. If the simulation data has large errors, these errors will automatically also be included in the NN.

An advantage of the ease of used and prediction speed of the network. The NN trained for predicting the maximum water levels for Hong Kong and Macau can be used with minimal knowledge and gives the prediction results immediately. A hydrodynamic storm surge model must be set-up, and simulated to obtain the results. Even if the surge model is already setup, in case of an approaching TC, the parametric wind field still must be calculated and the storm simulated. Depending on the state of the model, this can take from a few hours up to days or weeks if no surge model is set-up yet. In the case of an approaching TC, the NN is capable of predicting the maximum water level based on only 7 track parameters, which published by the TC agencies. The results of NN will give a very fast indication of the predicted surge height at the locations considered.

A possible disadvantage for the use of NN in storm surge predictions is the fact that many data samples are required for training. Actual observed data is often only available for shorter periods of time and for a few specific locations. So when using historical storms, it can be hard to find sufficient data. This problem is completely overcome by introducing the synthetic storms. With synthetic simulation the training database can be expanded to large number of samples. Furthermore, instead of water level data of a few locations (tidal measurement stations), the synthetic simulation provide the results for the entire computational domain, giving more possibilities for predictions.

7.1.4. Overall Neural Network storm surge prediction performance

During this thesis research three different neural networks were developed, maximum water level prediction is a map version, coastline version and 10 locations version. These network have been tested and calibrated extensively in an effort to create an accurate prediction network.

At first, an effort was made to develop an accurate network with a map output. During the development of this network some limiting factors were discovered. Firstly, the large number of output neurons needed to produce a water level increase the complexity of the network significantly. As discussed previously, complex networks are limited by the training algorithm, number of hidden neurons and memory capacity. With the data from the 1000 synthetic simulations it was impossible to train the network with reasonable accuracy. The variation in performance for different storms is too high. Although the visual result of maximum water levels by means of a map output can be very insight full, significant improvements must be made to the current network, to be used in forecasting. However, the current network requires improvements for reliable estimates.

In an effort to create a network more capable of predicting the water levels, the map network was simplified. At first, a network that only predicts the water level at the coastline was developed. Although the performance indicators show a small mean error for the training, validation and test data sets, the accuracy varies with the samples. For some samples, the network predictions are very accurate while for other samples the errors can be significant. By reducing the number of water levels cells by taking the coastline in a 1 km resolution, predictions still can be made without losing valuable water level information.

The final network that is developed, is the NN capable of making maximum water level predictions for 10 locations. This network is capable of making the most accurate predictions. However, this network

still shows a variety in the errors for different samples. The validation of the network shows that the water levels can be predicted with a maximum error of 30 cm. Although the network is trained with synthetic data, it can be also be used for historical storms. In order to make 'real world' predictions with the developed network, one needs the track parameters of the TC when it is close to Hong Kong and the tidal elevation at the moment of the TC passing. Executing the network function for these parameters will result in reasonable accurate results for the maximum water levels if the tidal elevation is added to the result (see section 5.6.1). Furthermore, the network gives insight on where the maximum water levels will occur. Due to missing validation data for Macau and simplification made for the data used in the network training, the errors for Macau are larger than for Hong Kong. This is likely due to the spatial separation between the two cities. For accurate water level predictions in Macau, it would be preferred to develop a separate network that selects the input parameters based on the track data being close to Macau.

Although three different networks have been developed, there is still room for improvements for these networks. The main problem is the variety in the quality of the predictions. Especially the coastline and locations network are capable of making accurate predictions for many different input data combinations. However, as we have seen some input combinations still lead to significant errors. Due to time constraints it was not possible to further improve these networks. However, there are still many options and aspects that can be considered in an effort to increase the accuracy. These aspects could be further investigated in future studies. Due to the current variability in the performance of the networks, the developed networks are not yet ready to be used for operational forecasting. The obtained results however show that artificial neural networks provide a promising technique for the prediction of the maximum water levels. The recent focus towards data science will only increase the use of these data driven models further in the future.

7.2. Recommendations

- **Change hydrodynamic model set-up to include overland flooding.** This thesis research has neglected overland flooding and only focused on predicted water level induces to storm surges. By adapting the Delft3D model to include over land flooding, one can gain insight in the most vulnerable locations in Hong Kong and Macau in case of an approaching TC. Additionally, the flood defences can be represented in the model. This allows for identification of possible weak spots in the flood defences and gives valuable insights for local authorities and policy makers for disaster and evacuation management.

- **Neural network development with time-series instead of one time-step**. By taking the time series of the track parameters and water levels, an NN can be trained that is also capable of predicting the arrival time of the maximum surge heights instead of only the level. Introducing time series can also lead to more accurate results, since the predictions are no longer based on the track parameters of one time step.

- **Additional focus on improving the current Neural Network Architecture.** Although the current networks are currently not suited to be used for real life forecasting due to the variety in the prediction accuracy. There is still room for improvement for the current networks. An in depth analysis of the effects of different architectures (number of nodes and layers etc.) should be performed, to find the best performing setup. Also it would be interesting to analyse the relationship between the number of hidden nodes and output nodes on the performance of the network. Additionally, more types of neural networks should be investigated. During this study only feed forward networks have been developed. There are more types of ANN available like recurrent ANN's and feedback networks.

- **Increase number of synthetic simulations** The synthetic TC catalogue, contains more than 200000 storms for the Western Pacific Basin only. During this thesis only a very small part of these storms have been simulated. The implementation of the synthetic storm data and the Matlab tools developed provides the opportunity to simulate a large number of synthetic storms efficiently and with a high degree of automatization. By increasing the number of simulations,

more data can be obtained and also the effect of storms that pass the PRD at some distance can be analysed.

- **Influence of sea level rise.** All over the world sea levels are rising. What are the implications for the PRD on the frequency of flooding, mortality rate etc. if the sea level keep rising like predicted. By linking storm surge research to possible sea level rise scenarios, one can gain additional insight in the possible week spots in the coastal flood defence system,
- **Storm surge simulation with waves included.** By including wave in the simulations, one can gain additional knowledge on the wave impact on the coast and flood defences during a storm. Wave interactions will become more important if overland flooding is also considered.

Bibliography

Arcadis. Sustainable Cities Index 2015. 2015.

Nathalie Asselman. Flood inundation modelling model choice and proper application. Technical report, 2009. URL http://www.floodsite.net/html/partner{_}area/project{_}docs/T08{_}09{_}03{_}Flood{_}inundation{_}modelling{_}D8{_}1{_}V3{_}3{_}P01.pdf.

Shaowu Bao, Lian Xie, and Leonard J. Pietrafesa. An asymmetric hurricane wind model for storm surge and wave forecasting. *27th Conference on Hurricanes and Tropical Meteorology*, (2000):4–8, 2006.

Ray Bell. Literature Review on the structure and dynamics of Tropical Cyclones. pages 1–16, 2010.

Xueyan Bi, Zhiqiu Gao, Yangang Liu, Feng Liu, Qingtao Song, Jian Huang, Huijun Huang, Weikang Mao, and Chunxia Liu. Observed drag coefficients in high winds in the near offshore of the South China Sea. J. Geophys. Res. Atmos.pdf. pages 6444–6459, 2015. doi: 10.1002/2015JD023172. Received.

Nadia Bloemendaal, Ivan D. Haigh, Hans de Moel, Sanne Muis, Reindert J. Haarsma, and Jeroen C.J.H. Aerts. Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data*, 7(1):1–12, 2020. ISSN 20524463. doi: 10.1038/s41597-020-0381-2. URL http://dx.doi.org/10.1038/s41597-020-0381-2.

Judith Bosboom and Marcel J.F. Stive. *Coastal Dynamics 1*. Delft Acedemic Press, 0.5 edition, 2015. ISBN 978-90-6562-3720.

Brahim Boutkhamouine, Hélène Roux, and François Pérés. Data-driven model for river flood forecasting based on a Bayesian network approach. *Journal of Contingencies and Crisis Management*, 28 (3):215–227, 2020. ISSN 14685973. doi: 10.1111/1468-5973.12316.

Anaïs Couasnon, Antonia Sebastian, and Oswaldo Morales-Nápoles. A Copula-based bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the houston ship channel, Texas. *Water (Switzerland)*, 10(9), 2018. ISSN 20734441. doi: 10.3390/w10091190.

Alex de Sherbinin, Andrew Schiller, and Alex Pulsipher. The vulnerability of global cities to climate hazards. *Environment and Urbanization*, 19(1):39–64, 2007. ISSN 09562478. doi: 10.1177/0956247807076725.

Bob Dean, Ian Collins, David Divoky Darryl Hatheway, and CFM Norm Scheffner. WAVE SETUP 1 FEMA COASTAL FLOOD HAZARD ANALYSIS AND MAPPING GUIDELINES Wave Setup FEMA Coastal Flood Hazard Analysis and Mapping Guidelines Focused Study Report Focused Study Leader Team Members W AVE SETUP i FEMA COASTAL FLOOD HAZARD ANALYSIS AND MAPPIN. 2005.

Deltares. Delft3D-FLOW user manual. Technical report, Deltares, 2018.

Ioana Dima and Melicie Desflots. Wind Profiles in Parametric Hurricane Models. page 6, 2010.

Robert DiPietro and Gregory D. Hager. *Deep learning: RNNs and LSTM*. Elsevier Inc., 2019. ISBN 9780128161760. doi: 10.1016/B978-0-12-816176-0.00026-0. URL https://doi.org/10.1016/B978-0-12-816176-0.00026-0.

Wenqi Du, Tso-Chien Pan, and Edmond Y M Lo. Evaluation of Natural Catastrophe Impact on the Pearl River Delta (PRD) Region – Earthquake Risk. *ICRM tropical report*, (October):1–6, 2018.

Gary D. Egbert and Svetlana Y. Erofeeva. Efficient Inverse Modeling of Barotropic Ocean Tides. *Journal of Atmospheric and Oceanic Technology*, 19(2):183–204, feb 2002. ISSN 0739-0572. doi: 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2. URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0426{%}282002{%}29019{%}3C0183{%}3AEIMOBO{%}3E2.0.CO{%}3B2.

Jean T. Ellis and Douglas J. Sherman. Perspectives on Coastal and Marine Hazards and Disasters. *Coastal and Marine Hazards, Risks, and Disasters*, pages 1–13, 2015. doi: 10.1016/B978-0-12-396483-0.00001-7.

EM-DAT. EM-DAT | The international disasters database, 2019. URL https://www.emdat.be/.

Kerry Emanuel. Tropical Cyclones. *Annual Review of Earth and Planetary Sciences*, 31(1):75–104, 2003. ISSN 0084-6597. doi: 10.1146/annurev.earth.31.100901.141259.

Jon French, Robert Mawdsley, Taku Fujiyama, and Kamal Achuthan. Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports. *Procedia IUTAM*, 25:28–35, 2017. ISSN 22109838. doi: 10.1016/j.piutam.2017.09.005.

Rao S. Govindaraju. ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY . By the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 1. *Journal of Hydrologic Engineering*, 5(April):124–137, 2000. ISSN 1084-0699. doi: 10.5121/ijsc.2012.3203.

Lvaro Prida Guill. Modelling of storm surge due to hurricanes in mississippi (usa). (February), 2020.

Stephane Hallegatte, Colin Green, Robert J. Nicholls, and Jan Corfee-Morlot. Future flood losses in major coastal cities. *Nature Climate Change*, 3(9):802–806, 2013. ISSN 1758678X. doi: 10.1038/nclimate1979. URL http://dx.doi.org/10.1038/nclimate1979.

Lee D. Harris. Characteristics of the hurricane storm surge. Technical Report 5, 1964.

S Haykin. *Neural Networks: A Comprehensive Foundation",*, volume 05. 1994. doi: 10.1142/s0129065794000372.

G. J. Holland. An analytic model of the wind and pressure profiles in hurricanes. *Monthly Weather Review*, 108(8):1212–1218, 1980. ISSN 00270644. doi: 10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2.

Greg J. Holland, James I. Belanger, and Angela Fritz. A revised model for radial profiles of hurricane winds. *Monthly Weather Review*, 138(12):4393–4401, 2010. ISSN 00270644. doi: 10.1175/2010MWR3317.1.

Leo H. Holthuijsen. *Waves in Oceanic and Coastal Waters*. Cambridge University Press, 1 edition, 2007. ISBN 78-0-521-86028-4.

Leo H. Holthuijsen, Mark D. Powell, and Julie D. Pietrzak. Wind and waves in extreme hurricanes. *Journal of Geophysical Research: Oceans*, 117(9):1–15, 2012. ISSN 21699291. doi: 10.1029/2012JC007983.

James R. Holton and Gregory J. Hakim. *An introduction to dynamic meteorology: Fifth edition*, volume 9780123848666. 2012. ISBN 9780123848666. doi: 10.1016/C2009-0-63394-8.

Jennifer L. Irish, Donald T. Resio, and Jay J. Ratcliff. The influence of storm size on hurricane surge. *Journal of Physical Oceanography*, 38(9):2003–2013, 2008. ISSN 00223670. doi: 10.1175/2008JPO3727.1.

W. S. Jäger, E. K. Christie, A. M. Hanea, C. den Heijer, and T. Spencer. A Bayesian network approach for coastal risk analysis and decision making. *Coastal Engineering*, 134:48–61, 2018. ISSN 03783839. doi: 10.1016/j.coastaleng.2017.05.004.

Jeffrey D. Kepert. Chapter 1: Tropical Cyclone Structure and Dynamics. In *Global Perspective on Tropical Cyclones: From Sceince to Mitigation*, chapter 1. World Sceintific, 2010. ISBN 978-981-4293-47-1.

Sooyoul Kim, Shunqi Pan, and Hajime Mase. Artificial neural network-based storm surge forecast model: Practical application to Sakai Minato, Japan. *Applied Ocean Research*, 91(July):101871, 2019. ISSN 01411187. doi: 10.1016/j.apor.2019.101871. URL https://doi.org/10.1016/j.apor.2019.101871.

Ying Ho Kwong and Mathew Y.H. Wong. State size and democratization in hybrid regimes: The chinese island cities of Macau and Hong Kong. *Island Studies Journal*, 12(2):113–126, 2017. ISSN 17152593. doi: 10.24043/isj.36.

France Lajoie and Kevin Walsh. A technique to determine the radius of maximum wind of a tropical cyclone. *Weather and Forecasting*, 23(5):1007–1015, 2008. ISSN 08828156. doi: 10.1175/2008WAF2007077.1.

H Lam and C C Lam. Reprint 603 Risk Management of Tropical Cyclone-induced Flooding in Hong Kong Workshop on Risk Management towards Millennium Development Goals and Socio-economic Impact Assessment of Typhoon-related Disaster , Kuala Lumpur , Malaysia , 5-9 September 2005. Number September, pages 5–9, 2005.

D S Lau, S T Chan, and Processes Iwtclp. Reprint 1302 STORM SURGE RISK IN HONG KONG ASSOCIATED WITH SUPER TYPHOON HATO (1713) 4th International Workshop on Tropical Cyclone Landfall. 2017.

T. L. Lee. Neural network prediction of a storm surge. *Ocean Engineering*, 33(3-4):483–494, 2006. ISSN 00298018. doi: 10.1016/j.oceaneng.2005.04.012.

Linlin Li, Adam D. Switzer, Yu Wang, Chung Han Chan, Qiang Qiu, and Robert Weiss. A modest 0.5-m rise in sea level will double the tsunami hazard in Macau. *Science Advances*, 4(8):1–11, 2018. ISSN 23752548. doi: 10.1126/sciadv.aat1180.

Xiaofeng Li, Jun A. Zhang, Xiaofeng Yang, William G. Pichel, Mark Demaria, David Long, and Ziwei Li. Tropical cyclone morphology from spaceborne synthetic aperture radar. *Bulletin of the American Meteorological Society*, 94(2):215–230, 2013. ISSN 00030007. doi: 10.1175/BAMS-D-11-00211.1.

David. j. Livingstone. *Artificial Neural Networks: Methods and Applications*. Humana Press, 2008. ISBN 9781588297181. doi: 10.1007/978-3-030-35743-6_3.

M. T. Montgomery and B. F. Farrell. Tropical cyclone formation. *Journal of the Atmospheric Sciences*, 50 (2):285–310, 1993. ISSN 00224928. doi: 10.1175/1520-0469(1993)050<0285:TCF>2.0.CO;2.

A. Mosavi, P. Oztruk, and K. Chau. Flood Prediction Using Machine Learning Models: Literature Review. *water*, 2016.

Jonathan F. Nott. *Palaeostorm Surges and Inundations*. Elsevier Inc., 2015. ISBN 9780123965387. doi: 10.1016/B978-0-12-396483-0.00005-4. URL http://dx.doi.org/10.1016/B978-0-12-396483-0.00005-4.

Shiqiu Peng and Yineng Li. A parabolic model of drag coefficient for storm surge simulation in the South China Sea. *Scientific Reports*, 5(1):1–6, 2015. ISSN 2045-2322. doi: 10.1038/srep15496.

Roger A. Jr. Pielke and Roger A. Sr. Pielke. *Hurricanes: their Nature and Impacts on Society*. John Wiley & Sons Ltd, 1997. ISBN 0471973548.

Preventionweb. Hong Kong (China) - Disaster & Risk Profile | PreventionWeb.net, 2019. URL https://www.preventionweb.net/countries/hkg/data/.

Pablo Ruiz-Salcines, Paulo Salles, Lucia Robles-Díaz, Gabriel Díaz-Hernández, Alec Torres-Freyermuth, and Christian M. Appendini. On the use of parametric wind models for wind wave modeling under tropical cyclones. *Water (Switzerland)*, 11(10), 2019. ISSN 20734441. doi: 10.3390/w11102044.

Robert W Schloemer. Analysis and Synthesis of Hurricane Wind Patterns over Lake Okeechobee, Florida. *Hydrometeorological Report*, 31:49, 1954.

A. Sebastian, E. J.C. Dupuits, and O. Morales-Nápoles. Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed. *Coastal Engineering*, 125:42–50, 2017. ISSN 03783839. doi: 10.1016/j.coastaleng.2017.03.008.

Ignacio Sepúlveda, Philip L.F. Liu, and Mircea Grigoriu. Probabilistic Tsunami Hazard Assessment in South China Sea With Consideration of Uncertain Earthquake Characteristics. *Journal of Geophysical Research: Solid Earth*, 124(1):658–688, 2019. ISSN 21699356. doi: 10.1029/2018JB016620.

Timothy Sim, Dongming Wang, and Ziqiang Han. Assessing the disaster resilience of megacities: The case of Hong Kong. *Sustainability (Switzerland)*, 10(4):1–16, 2018. ISSN 20711050. doi: 10.3390/su10041137.

A Sterl. Drag at high wind velocities - a review, Technical report, TR-361. Technical report, KNMI, 2017.

Marzenna Sztobryn. Forecast of storm surge by means of artificial neural network. *Journal of Sea Research*, 49(4):317–322, 2003. ISSN 13851101. doi: 10.1016/S1385-1101(03)00024-8.

Ruhhee Tabbussum and Abdul Qayoom Dar. Comparative analysis of neural network training algorithms for the flood forecast modelling of an alluvial Himalayan river. *Journal of Flood Risk Management*, (June 2019):1–18, 2020. ISSN 1753318X. doi: 10.1111/jfr3.12656.

Hiroshi Takagi and Wenjie Wu. Maximum wind radius estimated by the 50 kt radius: Improvement of storm surge forecasting over the western North Pacific. *Natural Hazards and Earth System Sciences*, 16(3):705–717, 2016. ISSN 16849981. doi: 10.5194/nhess-16-705-2016.

Hiroshi Takagi, Yi Xiong, and Fumitaka Furukawa. Track analysis and storm surge investigation of 2017 Typhoon Hato: were the warning signals issued in Macau and Hong Kong timed appropriately? *Georisk*, 12(4):297–307, 2018. ISSN 17499526. doi: 10.1080/17499518.2018.1465573. URL https://doi.org/10.1080/17499518.2018.1465573.

J. Teng, A. J. Jakeman, J. Vaze, B. F.W. Croke, D. Dutta, and S. Kim. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling and Software*, 90:201–216, 2017. ISSN 13648152. doi: 10.1016/j.envsoft.2017.01.006. URL http://dx.doi.org/10.1016/j.envsoft.2017.01.006.

Kevin J. Tory and William M. Frank. Chapter 2: Tropical Cyclone Structure. In *Global Perspectives on Tropical Cyclones: From Science to Mitigation*. World Sceintific, 2010. ISBN 10 981-4293-47-4.

United Nations. World Urbanization Prospects The 2011 Revision. Technical report, United Nations, 2011.

Albena Veltcheva and Hiroyasu Kawai. Investigation of the Typhoon Pressure and Wind Field with Application for Storm Surge Estimation. *Port and Airport Research institute Japan*, 2002.

Peter J. Vickery and Diuraj Wadhera. Statistical models of Holland pressure profile parameter and radius to maximum winds of hurricanes from flight-level pressure and H* wind data. *Journal of Applied Meteorology and Climatology*, 47(10):2497–2517, 2008. ISSN 15588424. doi: 10.1175/2008JAMC1837.1.

Todd L. Walton and Robert G. Dean. Landward limit of wind setup on beaches. *Ocean Engineering*, 36(9-10):763–766, 2009. ISSN 00298018. doi: 10.1016/j.oceaneng.2009.03.004. URL http://dx.doi.org/10.1016/j.oceaneng.2009.03.004.

Yuqing Wang and C. C. Wu. Current understanding of tropical cyclone structure and intensity changes - A review. *Meteorology and Atmospheric Physics*, 87(4):257–278, 2004. ISSN 01777971. doi: 10.1007/s00703-003-0055-6.

Jie Yang, Linlin Li, Kuifeng Zhao, Peitao Wang, Dong Wang, In Mei Sou, Zhengtong Yang, Jie Hu, Xiaochun Tang, Kai Meng Mok, and Philip Li Fan Liu. A Comparative Study of Typhoon Hato (2017) and Typhoon Mangkhut (2018)—Their Impacts on Coastal Inundation in Macau. *Journal of Geophysical Research: Oceans*, 124(12):9590–9619, 2019. ISSN 21699291. doi: 10.1029/2019JC015249.

Hijioka. Yasuaki, Erda Lin, Joy J. Pereira, R.T. Corlett, X. Cui, Insarov R.D. Lasco, E. Lindgren, and A. Surjan. Chapter 24: Asia. In *Climate change 2014: Impacts, Adaption, and Vulnerability. Part B-Regional aspects*, chapter 24. Cambridge University Press, 2014. ISBN 9781107683860.

M. Zijlema, G. Ph Van Vledder, and L. H. Holthuijsen. Bottom friction and wind drag for wave models. *Coastal Engineering*, 65:19–26, 2012. ISSN 03783839. doi: 10.1016/j.coastaleng.2012.03.002. URL http://dx.doi.org/10.1016/j.coastaleng.2012.03.002.

A

Literature Background and Theory

A.1. Wind setup

Moving air exerts a shear stress (τ) on the water surface that can be described by:

$$\tau_{wind} = C_d \rho_a W^2 \tag{A.1}$$

Where C_d is the wind drag coefficient, ρ_a the density of air and W the wind speed at the water surface. Due to the shear stress the water surface starts moving in the same direction as the wind. When the wind is directed towards the coast, a current towards the coast is generated in the upper layers due to the wind. The coast forms a barrier for the landwards current. To compensate for the landward currents, an opposite directed water mass transport occurs in the lower layers. As a consequence setup or set-down of water level occurs at the coast to balance the wind induced shear stresses. The wind induces water level setup can be described by the following equation (Bosboom and Stive, 2015):

$$\rho g h \frac{d\eta}{dx} = \tau_{wind,x} \tag{A.2}$$

$$\eta(x) = \int_0^L \frac{\tau_{wind}}{\rho g h} dx \tag{A.3}$$

From the equation it can be seen that the water level setup is inversely proportional to the water depth. This implies that for shallow coastal zones, the water level can pile up to high heights (storm surge). Walton and Dean (2009) investigated the effect of wind setup caused by storm surges on varying beach profiles. Te study showed that wind setup is typically the main driver for the increase of the still water level. Furthermore, it is shown that mild slopes can lead to very high water levels at the land/water interface.

A.2. Linear wave theory

The linear wave theory is often applied in coastal and oceanic engineering to describe the characteristics of wind waves. As discussed in the previous section, it is obtained from the Navier- Stokes and continuity equations. In this section, the equations of the linear wave theory are briefly discussed without the derivation. The picture below display a single wave with relevant terms of the linear wave theory.

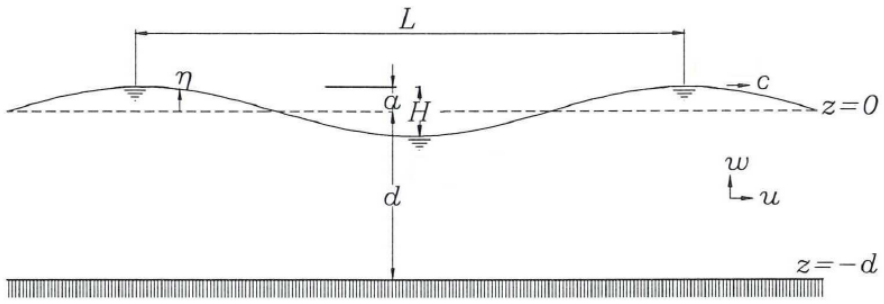


Figure A.1: Schematization of a single wave with relevant parameters displayed

The displacement of the water surface can be described by means of the angular frequency ω and the wave number (k).

$$\eta(x,t) = a\sin(\omega t - kx + \alpha) \tag{A.4}$$

$$\omega = \frac{2\pi}{T} \tag{A.5}$$

$$k = \frac{2\pi}{L} \tag{A.6}$$

The displacement of the wave is dependent on the water depth. In the linear wave theory waves can either be in shallow, deep or transitional water. In shallow water the wave motion extends all the way down to the bed. While for deep water the wave motion does not reach the bed. The criteria for deep or shallow water is besides the water depth also dependent on the wave length. In other words if the water depth is larger than about half the wave length the water is considered deep.

- Shallow water $h/L < 1/20$
- Deep water: $h/L > 0.5$
- Transitional water: $1/20 < h/L < 0.5$

The dispersion relationship is often used in the linear wave theory. It relates the wave number and length of a wave to its frequency and can be used to determine the phase and group velocity

$$\omega^2 = (\frac{2\pi}{T})^2 = gk * tanh(kh) \tag{A.7}$$

- For deep water ($h/L > 0.5$), the dispersion relation reduces to: $\omega^2 = gk$. The wave length becomes: $L = \frac{gT^2}{2\pi}$
- For shallow water, the wave length becomes; $L = T * \sqrt{gh}$

The energy plays an important part in the description of waves in a wave field. The total energy can be divided in potential and kinetic energy. The displacement of the water surface is related to the potential energy and the orbital movement to the kinetic energy. The total energy per unit of surface area can be calculated with the following equation:

$$E = \frac{1}{8}\rho g H^2 L \tag{A.8}$$

A.3. Wave setup

During a TC, waves generated by storm winds are propagating towards the coast. These waves not only carry energy towards the coast but also momentum. When the waves reach coastal areas they can induce wave setup. Wave setup is defined as the increase in water level above the stillwater level due to momentum transfer by waves that are in the surf zone. The radiation stress is defined as the depth integrated and wave averages flow of momentum due to waves. Spatial changes in the radiation stress causes wave forces to act on the fluid impacting the water motion and levels. The radiation stress includes the transfer of momentum and the wave induced pressure force. The cross shore radiation stress of x-momentum in the x-direction is defined as (Dean et al., 2005):

$$S_{xx} = \int_{-h}^{\eta} \left(p + \rho u^2\right) dz \tag{A.9}$$

By using the linear wave theory, one can express the radiation stress in terms of wave energy. If the wave propagating expression is perpendicular to the coast ($\theta = 0$) the equation reduces to equation A.11. The magnitude of the radiation stress is dependent on the wave height, length and water depth.

$$S_{xx} = (n - \frac{1}{2} + ncos^2\theta)E \tag{A.10}$$

$$S_{xx} = (2n - \frac{1}{2})E \tag{A.11}$$

$$\frac{d\bar{\eta}}{dx} = -\frac{1}{\rho g(\bar{\eta} + h)} \frac{dS_{xx}}{dx} \tag{A.12}$$

The surf zone is defined as the region in front of the coast where wave breaking occurs. In the surf zone the value of S_{xx} decreases, which results in a negative gradient.

$$F_x = -\frac{dS_{xx}}{dx} = \rho g h \frac{d\bar{\eta}}{dx} = \rho g(h_0 + \bar{\eta}) \frac{d\bar{\eta}}{dx} \tag{A.13}$$

$$\Delta\bar{\eta} = -\frac{3}{8}\gamma^2 \Delta h \tag{A.14}$$

If the dissipation of energy due to wave breaking is included, one can solve equation A.13 and estimate the maximum water level increase due to wave setup. The discussed theory on wave setup shows that this effect only occurs in shallow waters in the surf zone close to shore. However, wave setup can also occur when there is a flooded barrier island or reef present in front of the coast.

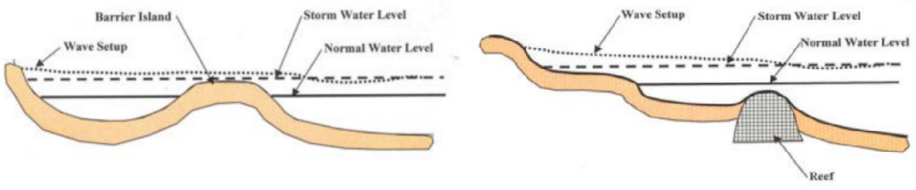


Figure A.2: Wave setup in front of the coast. Adapted from Dean et al. (2005)

Today, different models exist that can be used for the modelling of wind wave to describe the sea state. For example, the SWAN wave model developed by Delft University of Technology, can computed random, short crested wind generated waves in coastal regions and inland waters.

A.4. Conceptual description Delft3D

With the Delft3D-FLOW software either 2D (depth averaged) or 3D (unsteady flow and transport) simulations can be conducted. A wide range of natural phenomena like tidal and astronomical forcing can be simulated. The 2D option uses the depth averaged equations for simulation where the variation in vertical direction is of less importance. This approach is often used for tidal waves, surges and tsunamis. The 3D approach is of interest for transport problems where there is significant variations in the vertical direction and are more used for specific flow problems like salt intrusion, waste water, sediment transport and river water discharges among others. Both the 2 and 3D model uses a version of the shallow water equations derived from the three dimensional Navier-Stokes equations Deltares (2018). Instead of flat xyz coordinates, Delft3D uses orthogonal curvilinear coordinates in which the coordinate line may be curved. The two types of coordinates systems that can be used are Cartesian or spherical. The vertical grid used is called a σ, with two planes that are not strictly horizontal but follow the bathymetry and free surface.

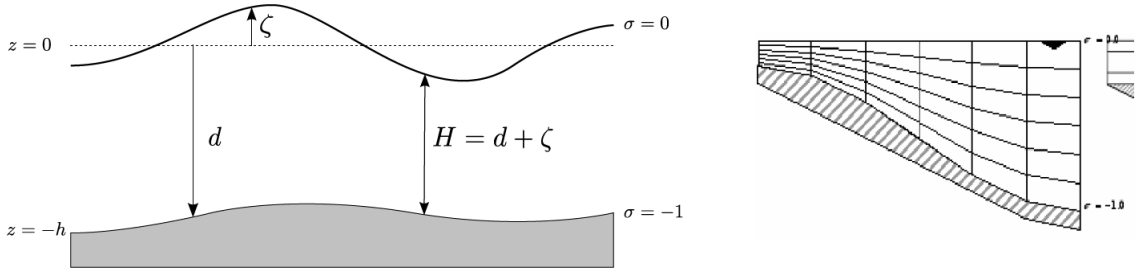


Figure A.3: (a) Definition of water level, depth and total depth (b) σ grid. Adapted from Deltares (2018)

The shallow water equations are given by the momentum equations in horizontal direction (equations A.15 and A.16) and the continuity equation (equation A.17). The equations are displayed in the form of the curvinlinear grid coordinates (ξ and η).

$$\frac{\partial u}{\partial t} + \frac{u}{\sqrt{G_{\xi\xi}}}\frac{\partial u}{\partial \xi} + \frac{v}{\sqrt{G_{\eta\eta}}}\frac{\partial u}{\partial \eta} + \frac{\omega}{d+\zeta}\frac{\partial u}{\partial \sigma} - \frac{v^2}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\eta\eta}}}{\partial \xi} + \\ + \frac{uv}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\xi\xi}}}{\partial \eta} - fv = -\frac{1}{\rho_0\sqrt{G_{\xi\xi}}}P_\xi + F_\xi + \frac{1}{(d+\zeta)^2}\frac{\partial}{\partial \sigma}\left(\nu_V \frac{\partial u}{\partial \sigma}\right) + M_\xi \tag{A.15}$$

$$\frac{\partial v}{\partial t} + \frac{u}{\sqrt{G_{\xi\xi}}}\frac{\partial v}{\partial \xi} + \frac{v}{\sqrt{G_{\eta\eta}}}\frac{\partial v}{\partial \eta} + \frac{\omega}{d+\zeta}\frac{\partial v}{\partial \sigma} + \frac{uv}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\eta\eta}}}{\partial \xi} + \\ \frac{u^2}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\xi\xi}}}{\partial \eta} + fu = -\frac{1}{\rho_0\sqrt{G_{\eta\eta}}}P_\eta + F_\eta + + \frac{1}{(d+\zeta)^2}\frac{\partial}{\partial \sigma}\left(\nu_V \frac{\partial v}{\partial \sigma}\right) + M_\eta \tag{A.16}$$

$$\frac{\partial \zeta}{\partial t} + \frac{1}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial((d+\zeta)u\sqrt{G_{\eta\eta}})}{\partial \xi} + \frac{1}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}} + \frac{\partial \omega}{\partial \sigma} = (d+\zeta)\,(q_{in} - q_{\text{out}}) \tag{A.17}$$

Where u and v are the depth average flow velocities in horizontal direction, w the velocity in vertical direction, v_v is the vertical viscosity coefficient, P is pressure gradient, F are horizontal Reynolds stresses and M represent contributions due to external sources. One of the assumptions of the shallow water equations is the hydro-static pressures assumption, the vertical momentum equation is reduced to a hydro-static pressure distribution.

$$\frac{1}{\rho_0\sqrt{G_{\xi\xi}}}P_\xi = \frac{g}{\sqrt{G_{\xi\xi}}}\frac{\partial \zeta}{\partial \xi} + \frac{1}{\rho_0\sqrt{G_{\xi\xi}}}\frac{\partial P_{atm}}{\partial \xi} \tag{A.18}$$

$$F_\xi = \frac{1}{\sqrt{G_{\xi\xi}}}\frac{\partial \tau_{\xi\xi}}{\partial \xi} + \frac{1}{\sqrt{G_{\eta\eta}}}\frac{\partial \tau_{\xi\eta}}{\partial \eta} \tag{A.19}$$

$$M_\xi = q_{in}(\hat{U} - u) \tag{A.20}$$

The equation that describe the flow in Delft3D are a set of partial differential equations. To solve this mathematical problem, a set of initial and boundary conditions must be imposed. The boundary conditions are either closed or open. The open boundaries are imposed on the 'water' boundaries and are used to restrict the computational area. Boundary conditions can also be imposed in the vertical direction, like free surface and seabed boundary conditions. For the depth averaged 2D flow the bed shear stress induced by turbulent flow is given by the quadratic friction law which used the Chézy coefficient. The Chézy coefficient is widely used coefficient to include friction and be derived from the Chézy, Manning or White-Colebrook formulations

$$\vec{\tau}_b = \frac{\rho_0 g \vec{U}|\vec{U}|}{C_{2D}^2} \tag{A.21}$$

The 'open' boundaries are imposed on the edges of the domain and provide forcing into the model for example tidal forcing. Multiple types of open boundary conditions can be used in Delft3D for different purposes:

- Water level: $\zeta = F_\zeta(t) + \delta_{atm}$
- Velocity: $U = F_U(t)$
- Discharge: $Q = F_Q(t)$
- Neumann: $\frac{\partial \zeta}{\partial \vec{n}} = f(t)$
- Riemann invariant: $U \pm \zeta\sqrt{\frac{g}{d}} = F_R(t)$

A.5. Neural Network - Training Algorithms

Back Propagation algorithm

Back propagation is the most widely used training algorithm. The algorithm minimises the the error function. The input samples are passed through the network to the output layer. Then the output is compared to the target output and the error is calculated. At the beginning of the training process, all weights are awarded a small numeric value. These weights are optimized iterative using the steepest gradient principle (gradient descent). The back propagation algorithm consists of two phases: forward pass and backward pass. During the forward pass, processing of information occurs from the input to the output layer. In the backward pass, the error obtained in the output layer is sent back to the input layer and the weights are modified. The weights are adjusted according the following equation.

$$\Delta w_{ij}(n) = -\varepsilon^* \frac{\partial E}{\partial w_{ij}} + \alpha^* \Delta w_{ij}(n-1) \tag{A.22}$$

Where $\Delta w_{ij}(n)\ and\ \Delta w_{ij}(n-1)$ are the weight increments between the nodes during the nth and $(n-1)$ pass, also called epoch (Govindaraju, 2000). In ANN an epoch refers to one cycle through the full training data set. In the above equation ϵ and α are the learning rate and momentum. The momentum factor allows for speeding up the training in areas with small errors and prevent oscillation in the weights. Since back propagation is an iterative procedure, there must be a criterion imposed to stop the training. The is usually done, by assigning part of training samples as validation samples. The iterations are stopped when the squared error in the prediction of the validation set is minimum (loss function minimum). More iterations after reaching this minimum will lead to larger errors and is often referred to as over training of the network.

Conjugate Gradient algorithm
This method does not proceed along the error gradient but in a direction orthogonal from the previous step. In general, solutions are faster obtained than in the gradient method. The method uses equation A.5 to determine the weights of the neurons.

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \varepsilon \mathbf{P}(n) \tag{A.23}$$

$$\mathbf{P}(n+1) = -\mathbf{g}(n+1) + \beta(n)\mathbf{P}(n) \tag{A.24}$$

Where $P(n)$ is the direction vector and $\beta(n)$ a time dependent parameter and $g(n)$ the gradient vector. The initial direction is set equal to the negative gradient vector $P(0) = -g(0)$.

Bayesian Regularisation algorithm
Bayesian regularisation is a mathematical process that converts a non linear regression into a statistical problem (Livingstone, 2008). The algorithm is considered to be more robust than back propagation and no over training can occur.

Levenberg-Marquardt algorithm
This algorithm is also known as the damped least squares method and is used to solve non linear problems. This iterative technique locates the minimum of a function that is expressed as the sum of squares and is widely used in fitting problems. It has a fast learning rate but has a large memory capacity. It can therefore, not be used for very complex networks.

A.6. Recurrent ANNs

Recurrent or feedback neural networks, is another kind of network structure in which the outputs from the neurons are use as feedback to the neuron in the previous layer (Current output is considered as an input for the next output). These types of NNs are often used for time series prediction and processing control. In flood forecasting these types of networks are often used for short term predictions.

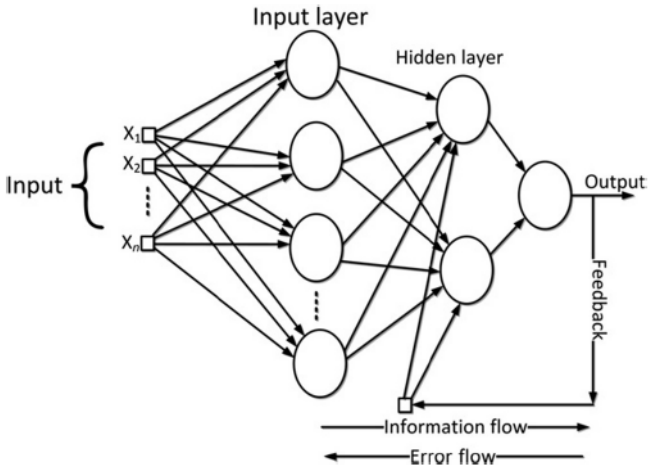


Figure A.4: Recurrent Neural Network. Adapted from DiPietro and Hager (2019)

B

Sensitivity Analysis

In this chapter a sensitivity analysis for the calibration of the numerical model is treated in detail. The main goal of the sensitivity analysis is to investigate the influence of certain parameters on the results of the simulation. First different tidal models are tested that will represent the boundary conditions in the model. Secondly an analysis is made concerning the best values for the wind drag coefficient. The final part of the sensitivity analysis is focused on the best determination of the tropical cyclone parameters.

B.1. Boundary Conditions - Astronomical Constituents

In this section, the influence of the use of different boundary models is investigated. For this purpose, three different tidal models are used in the simulation: the TPXO8, TPXO9 and FES2014 models. Figure B.1 shows the results of the simulations with the different models used as tide forcing input. At first sight the tidal models produce results that are in phase with the measured signals from the tidal buoys CT8 and CHC. Station MWC clearly shows an phase error between the signals.

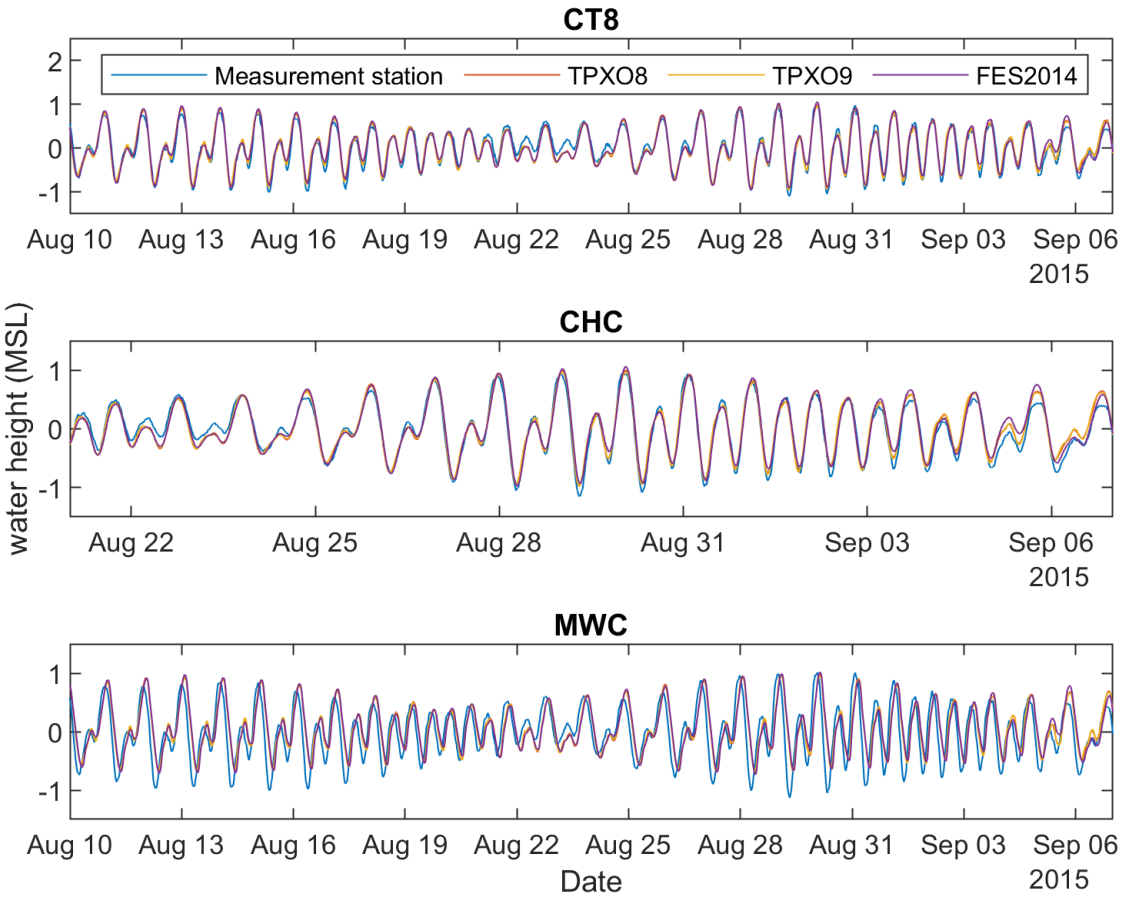


Figure B.1: Tidal simulation results for different tidal models

A closer inspection shows that simulated tide are as expected of mixed type with a predominantly semi-diurnal (two high and low waters per day) signal. For the most part, the simulated signals for CT8 and CHC follows the measured signal pretty close. However, from the figure it can be seen that between 22 and 25 Augustus the difference becomes more significantly. During this period around neap tide, the simulated signals consistently underestimate the water level. There are multiple possible explanations for these underestimations during the period around neap tide. During the tidal simulations there is no wind forcing included in the model, therefore wind and wave setup are not included. Another possible reason is the absence of river discharges in the model. As discussed previously, there are many rivers with significant discharge to flow out into the pearl river delta. It also likely that the neap tide under estimations are due to the used astronomical constituents in the boundary conditions. Since it is hard to visually determine the best suited tidal model, the root mean square and percent error are calculated.These statistics give good insight on the average error between the measured signal and simulated signal.

	CT8		CHC		MWC	
	RMSE [cm]	% error	RMSE [cm]	% error	RMSE [cm]	% error
TPXO8	11.07	8.89	10.22	8.8	30.48	24.88
TPXO9	11.22	9.01	10.48	8.9	30.49	24.81
FES2014	10.06	8.52	11.04	9.01	33.63	27.95

Table B.1: Errors between real and simulated tide

The results in table B.1 show that the different tide models produce similar results for the considered stations. The average errors over the simulation period are close to each other. For the three considered locations, the poorest results are obtained for station MWC for all tidal model used. As can be seen from the figure, the model consistently shows an phase error and an error in the estimation of the low waters. In the real world the water level during low water drops lower than the model calculates. An explanation for the poor performance of the MWC simulation is the location of the station. Tidal buoy MWC is located in a relative small channel surrounded by multiple islands. These islands are connected with bridges to each other. When taking a closer look at the bathymetry used in the model, it can be seen that at the locations of the bridges the bathymetry file uses the elevation of the bridges instead of the bed level. As a consequence the bridges form a barrier for the flow in the model. By changing the bathymetry and rerunning the model, the results for station mwc improve significantly as can been seen in figure B.2. The new RMSE error reduces to 12.1 cm, which is a significant reduction compared to the simulation with the original bathymetry.

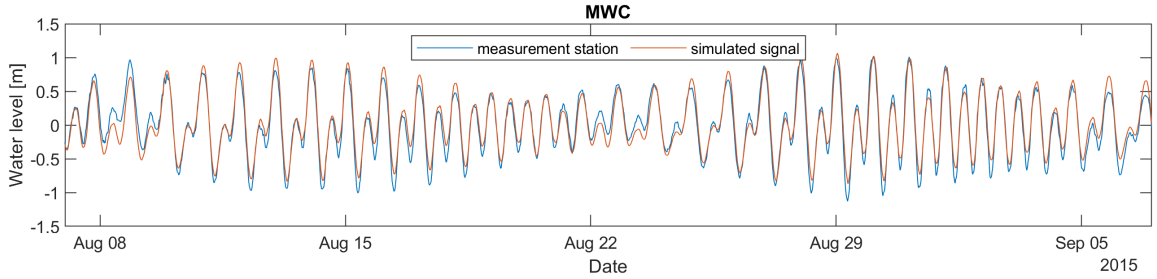


Figure B.2: Tidal simulation station MWC

Since there is no distinct difference between the simulated signals the best working tidal model is picked based on the RMS and percentage error. Overall, the TPXO8 atlas seems to be performing the best for the three tidal buoys. The average error of the tidal signal in the model are below 10%, which is acceptable for storm surge modelling. Since there is no distinct difference between the cons

B.2. Wind drag coefficients

The dimensionless wind drag coefficients are usually used to determine the stress (equation B.1) exerted by the wind on water (air-sea interface). In fluid dynamics, the wind shear stress is used to determine the water level setup due to wind (wind setup) among other things. Over the years a lot of research has been conducted into the values and behaviour of the drag coefficient.

$$\tau = \rho_a * C_D * U_z^2 \tag{B.1}$$

The total wind stress can be described by the taking the sum of three components: viscous, turbulent and form stresses. For low wind speeds the water surface is flat and turbulence is low. The stress is dominated by viscous stresses for these low wind speeds. For wind speeds exceeding 5 m/s the turbulent and stresses become dominant. The turbulent stresses are generated from momentum being transferred to the water body by turbulent eddies that propagate with the wind field. The form stresses are caused by interactions between the air surface and waves that are formed by the wind.

$$U_z = \frac{u_*}{\kappa} ln \frac{z}{z_0} \tag{B.2}$$

$$C_d = \frac{u_*}{U_{10}}^2 = \frac{\kappa^2}{ln^2(10/Z_0)} \tag{B.3}$$

Where u_* is the friction velocity, κ the von Kármánn constant, z_0 the roughness length, ν the kinematic viscosity and α the Charnock parameter.In general it is agreed that the drag coefficient increases almost linearly with the wind speed for low to moderate speeds. The behaviour of the drag coefficient above 30 m/s is less established. Figure B.3 shows the results of different wind drag coefficient researches. It can bee seen that for most researches, the maximum value for the drag coefficient is obtained for wind speeds between 25 to 35 m/s with values between 0.18 to 0.28 10^-3.

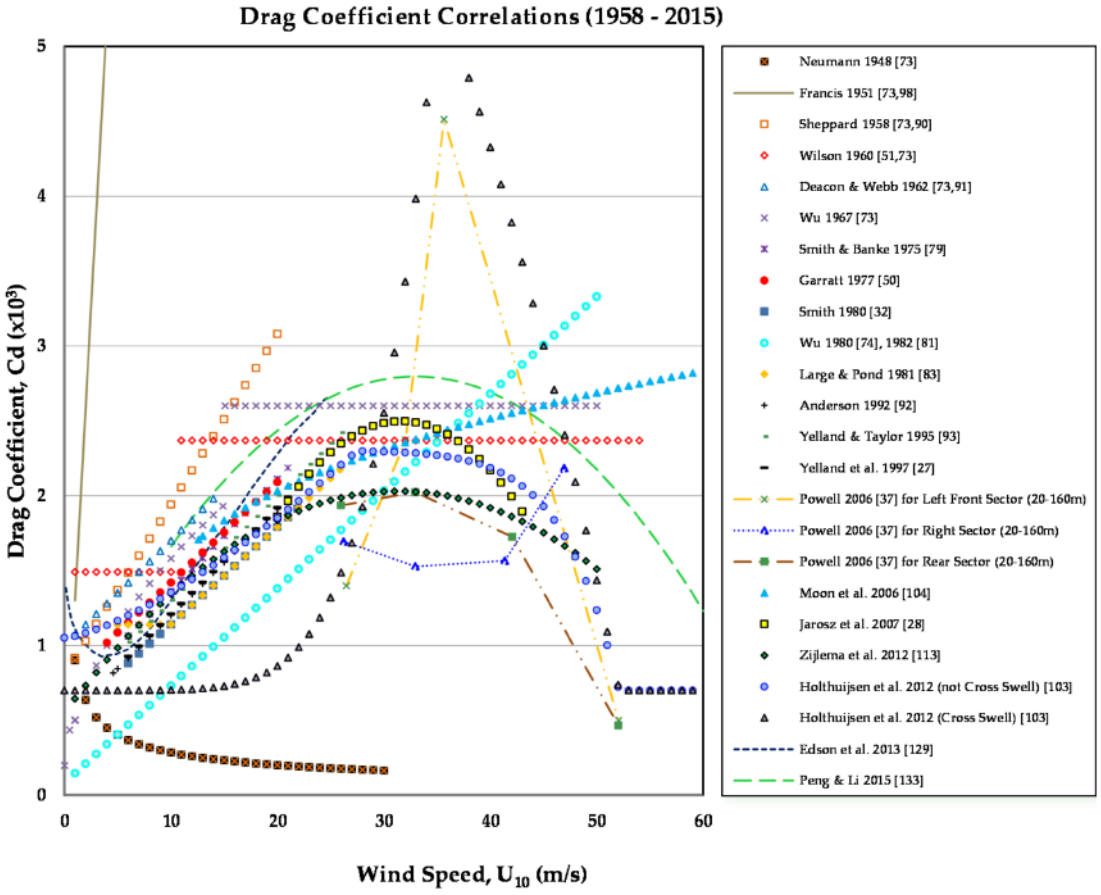


Figure B.3: Different Wind Drag Correlation. Adapted from Sterl (2017)

Some older studies suggested that the drag coefficients increase even further, while most recent studies show that the value of the drag coefficients reaches a maximum around 30 m/s after which it decreases again. The wind drag coefficient is dependent on the surface roughness, sea state and stratification (Sterl, 2017). For increasing wind speeds the sea surface gets rougher and the drag coefficient increases. Since the drag coefficient is of importance for the Delft3D results and different studies suggest different drag coefficient values, a sensitivity analysis is conducted to investigate the effect on the surge height for different drag coefficients. To do so, different storm surge simulation (tides and wind) are executed in Delft3D. The results of the different simulations are then compared to each other and the best suited drag relationship is chosen and used for the synthetic TC simulations.

Bi et al. (2015) investigated the observed drag coefficients in high winds in the near offshore of the South Chinese Sea. For this research data of seven typhoon event from two offshore observation towers were used. The study suggested that the drag coefficients decreases for increasing speeds between 5-10 m/s. Then the coefficient increases and reaches a maximum for a wind speed around 18 m/s. For highest wind speeds the coefficient decreases again. An other study for the South Chinese Sea suggested a parabolic model to determine the drag coefficients for different wind speeds based on storm surge observations and simulations (Peng and Li, 2015). Equation B.4 is suggested for the relationship between the drag coefficient and wind speed. α and c are constants determined from different TC's. This relationship implies that the maximum value for the drag coefficient is reached for wind speeds of 33 m/s after which the value decreased again.

$$C_d = -\alpha(V_p - 33)^2 + c = -0.00215(V_p - 33)^2 + 2.797 \tag{B.4}$$

Holthuijsen et al. (2012) investigated the drag coefficients for wind and waves during tropical cyclones. This study included the effects of white capping and cross swell to determine the wind drag during high winds. The research is based of a large number of observations with drop-sondes to determine the drag coefficients. The research found systematically lower drag coefficients compared to earlier studies that were also based of drop-sonde measurements. The difference in results was possibly explained by the effect of cross-swell on the drag. The study found evidence that the swell direction influences the drag coefficient. Furthermore, the reason that the drag coefficient decreases or stays the same at high wind speeds is also investigated. They conclude that the reason for this might be white capping and the forming of foam on the surface (Holthuijsen et al., 2012). However, the impact of swell on drag coefficient is still a topic of investigation today. Finally, Zijlema et al. (2012) proposed a new method to improve the SWAN wave model results with new drag-wind parameterizations. For wave modelling often different values are used for the wind drag coefficients and the bottom friction. This study proposed a new drag parameterization with lower values for drag and bottom friction to improve the SWAN model results. The study concluded that a lower bottom friction values than the default one used in SWAN is preferred for both swell and locally generated waves. Furthermore, the proposed wind drag parameterization (Equation B.5) fit the observations much better than the default coefficients used in SWAN, especially at high wind speeds.

$$\begin{aligned} C_D &= (0.55 + 2.97\tilde{U} - 1.49\tilde{U}^2) * 10^{-3} \\ \tilde{U} &= U_{10}/U_{ref} \;\; with \; U_{ref} = 31.5 m/s \end{aligned} \tag{B.5}$$

Delft3D allows the user to input three values for drag coefficients for different wind speeds. These values are then linearly interpolated to all wind speeds. As discussed previously, some studies have suggested the use of parabolic relationships. However, this is not possible with Delft3D. Consequently all drag coefficients are linearly interpolated between the three input points. Table.. shows the different values used for the sensitivity analysis. Due to the limitations of Delft3D the drag relationship of Bi et al. (2015),Peng and Li (2015) and Zijlema et al. (2012) cannot be reproduced exactly. As a consequence, the linear interpolated drag coefficient values are slightly underestimated compared to the parabolic values.

	A		B		B	
	ws (m/s)	C_d (-)	ws (m/s)	C_d (-)	ws (m/s)	C_d (-)
Bi et al. (2015)	7	0.0013	18	0.002	33	0.0016
Peng et al. (2015)	0	0.0005	30	0.0027	60	0.0012
Zweers et al (2010)	0	0.001	30	0.0025	60	0.0015
Holthuijsen et al. (2015)	20	0.001	40	0.0022	60	0.0008
Zijlema et al. (2012)	0	0.0006	30	0.0018	80	0.0008

Table B.2: Relation drag coefficients and wind speed

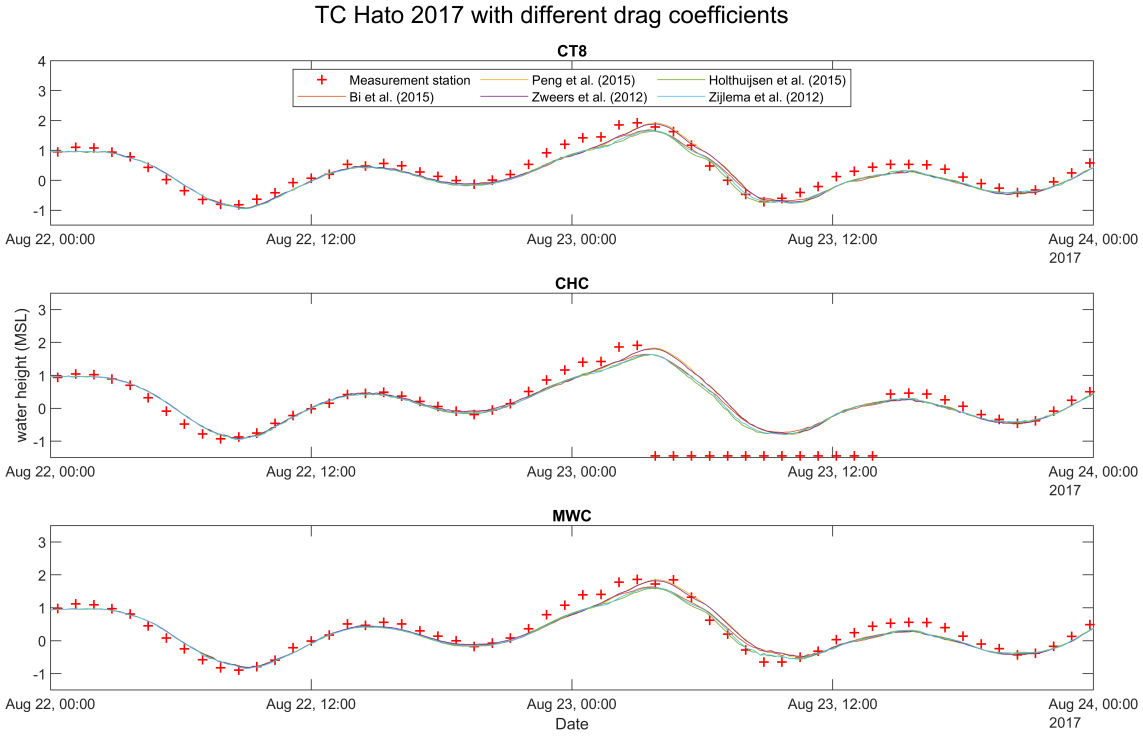


Figure B.4: Hato Cd influence

B.3. Tropical cyclone parameters

The tropical cyclone track data is provided by multiple agencies participating in the WMO tropical cyclone programme. Some of the agencies that provide TC track data are the Joint Warning Typhoon centre (USA), RMSC Tokyo (Japan), CMA (Chinese meteorological agency) and Hong Kong Observatory. The data provided by these agencies can be visited from: http://ibtracs.unca.edu/ .These agencies all provide slightly different parameters for the TC's. For example, the JWTC is the only agency that also provides the RMW besides the locations, maximum wind speed and eye pressure. The Tokyo agency provides values for the radius of 50 knot winds (R50) instead of the RMW. Furthermore, the measured wind speeds and eye pressures also slightly differ between the different agencies. The sensitivity analysis for the TC parameters aims a determining the influence of data from the TC agencies and different method for estimating the maximum radius of winds on the model results.

A recent study by Ruiz-Salcines et al. (2019) investigated the use of different parametric wind models for TC modeling. The results showed that one wind model might be better for a particular event, but for modelling a large number of TC's the considered models performed basically the same. Due to the fact that the Holland wind model has been used extensively in the past and yielded good results, it is used during this Thesis research. The estimation of the radius of maximum winds (RMW) has been a research topic for a long time and the improvement of the parameterization of the TC's parameters is still researched today. One of the topics without widespread consensus is the estimation of the RMW parameter. The RMW represents to distance from the TC center to the bands with the strongest winds. For example the Holland (1980) model can estimate the RMW from the wind speed and eye pressure alone, but the RMW can also be used as input in the wind model. Other studies have suggested to estimate the RMW based of the diameter of the eye, maximum wind speed or central pressure alone. Recent insights however have led to the believe that the parameterization based on only the wind speed an eye pressure can lead to severe over or underestimations of the RMW because of variations in the data. To investigate the influence of the RMW on the surge heights in the model, a sensitivity analysis is conducted for multiple estimations of the RMW. For all the considered simulations, the Holland wind model is used to calculate the spiderweb wind input. Some agencies display the wind speed in 1 minute average while others use 10 minute average. For storm surge simulations a 10 minute average wind speed is required. The data in 1 minute average wind speed must therefore by converted. To do so, a so called wind conversion factor must be applied to the wind speed. According to the Delft3D flow user manual the wind speed should be multiplied with 0.9 to go from 1 minute to 10 minute average wind speed.

- RMW directly from USA (JWTC) track data IBTrACS.
- RMW calculated with Takagi from the Tokyo agency data. Takagi and Wu (2016) suggested a method to estimate the RMW from the radius of 50 knots wind (R50). A new relationship was created from meteorological TC data for 1990-2013 in the Western Pacific. The study concluded that the traditional RMW estimation methods from the maximum wind speed and eye pressures showed significant scatter. The method suggested to estimate from the R50 showed to improve the estimation of the RMW significantly. The new relationship is: $RMW = 0.23 * R_{50}$. Although this relationship improves the estimation of the RMW, there still are under and over estimation for storm surges. Therefore, this relationship must be used with care. Ultimately a variability of $0.15 * R_{50}\ to\ 0.35 * R_{50}$ should be considered to avoid under and over estimations. Furthermore it must be noted that the R50 parameter is not always given for the entire duration of the TC. Therefore, this relationship can only be used if sufficient data points of R50 are available in IbTRACs.
- RMW determined from the central pressure. $RMW = 0.676 * P_c - 578$ with the eye pressure in hpa and RMW in km.
- Standard value for RMW (25 nautical miles). One option to calculate the spiderweb in the wind enhancement scheme (WES) in Delft3D is to use a default value of 25 nautical miles for the RMW. This value is suggested when insufficient data is available.

All simulations have been conducted with wind speed and eye pressure data from either the Tokyo or United States TC agency with varying values for the RMW. The wind drag coefficients are the same

for all simulations. The results clearly shows that the RMW has a big influence in the maximum surge heights. The surge heights obtained from the simulations show that for the USA TC data, the surge height is overestimated. This can be explained by the fact that the wind speed and pressure provided by the USA has higher values than the data provided by the Tokyo agency. The RMW used in the simulation also influences the maximum surge height. The error in the tidal signal of measurement buoy CHC is due to the fact that it stopped recording for approximately 12 hours during the most intense part of the storm in the Pearl river delta.

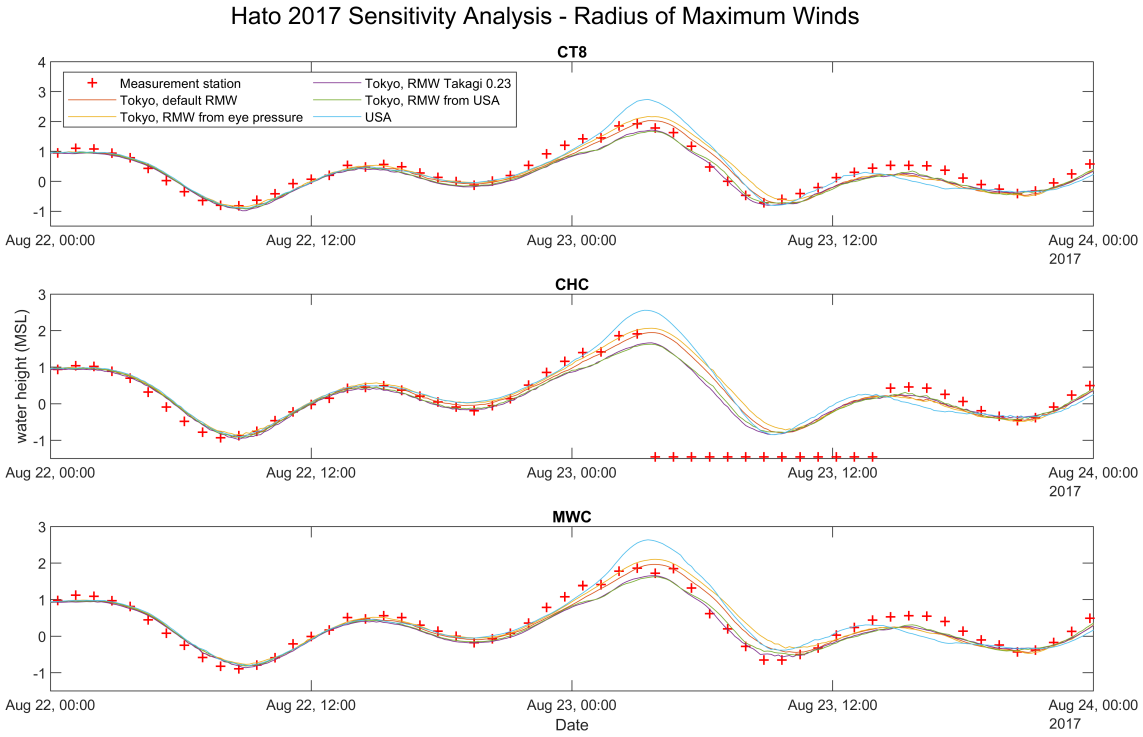


Figure B.5: Hato TC par influence

B.4. Grid cell resolution

As stated before, the grid cell resolution can be influencing the accuracy of the model. In general better results can be obtained for higher resolutions. However, the type of hydrodynamic processes to be simulated also influences the required grid cell resolutions.

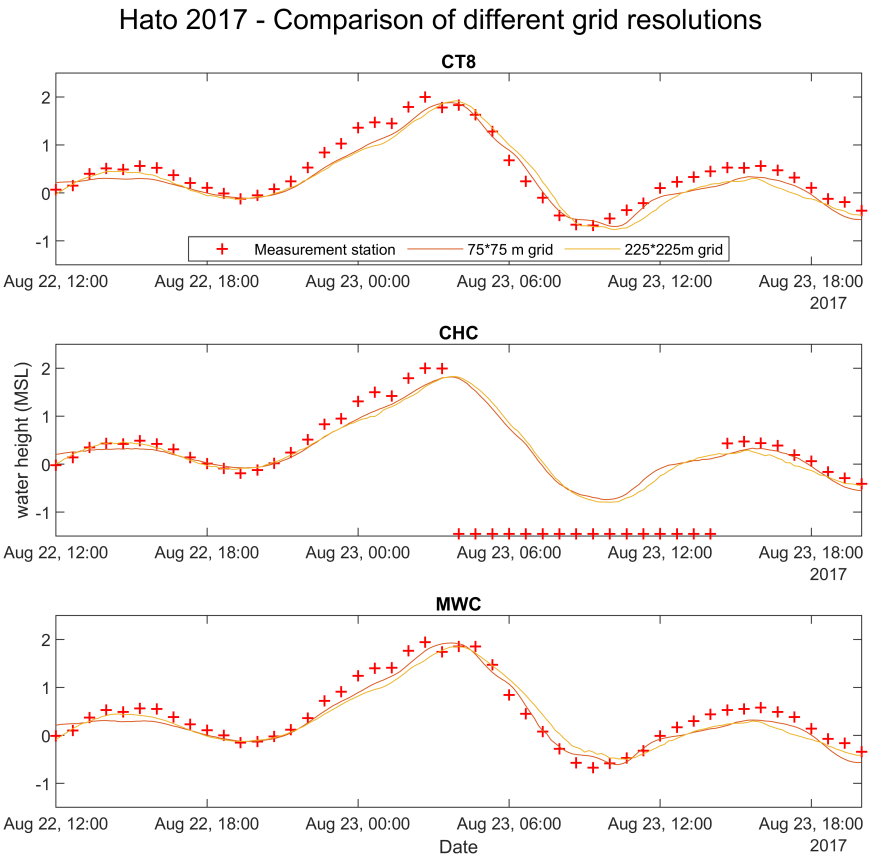


Figure B.6: Hato TC - Spatial resolution

The figure below shows that increasing the model resolution to 75 by 75 meter, does not significantly change or increase the performance of the model. On the other side, a significant increase in computational time is observed to compute this very fine domain. Based on the results, the 225 by 225 m grid is taken as finest grid.

C

Neural Network Output Results

C.1. Neural Network output maps

Training sample comparison

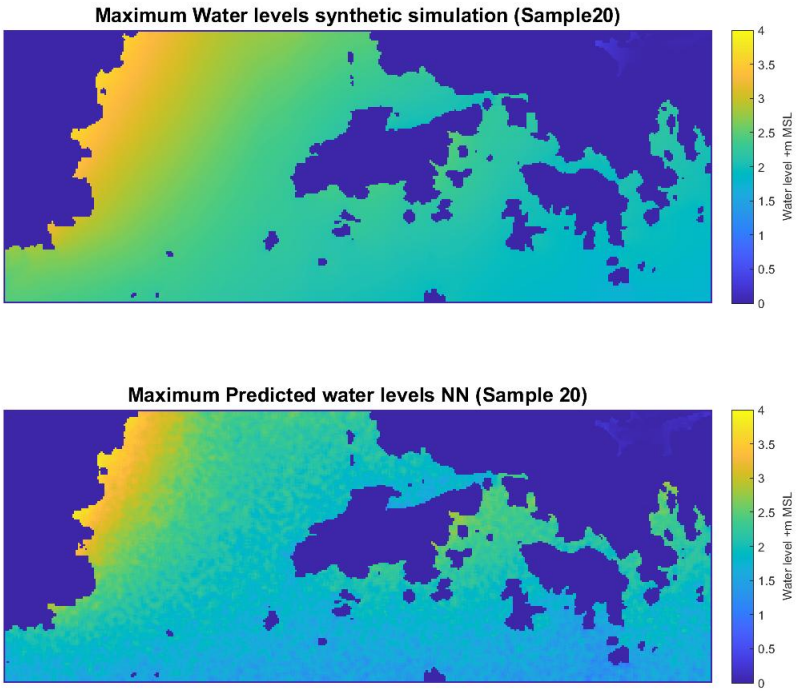


Figure C.1: Training sample 20

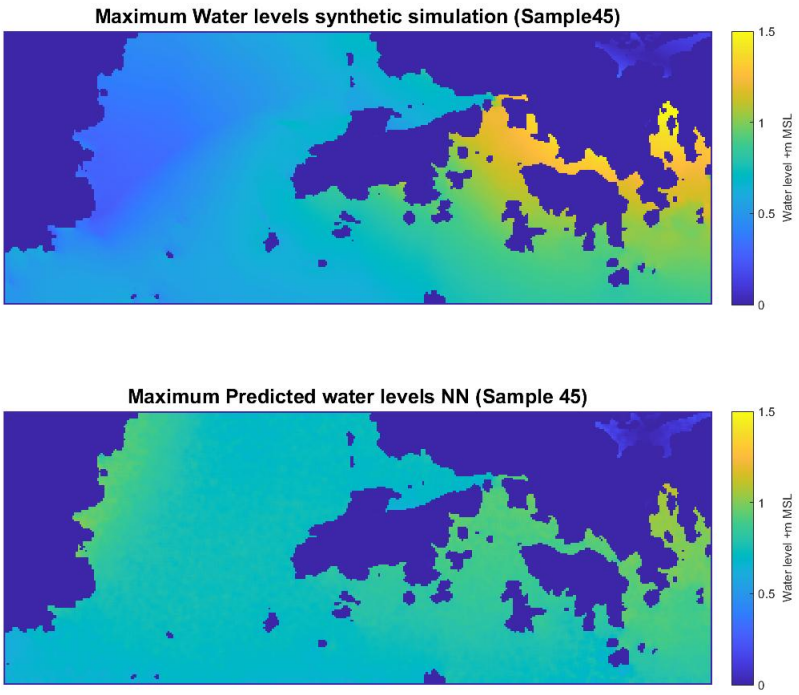


Figure C.2: Training sample 45

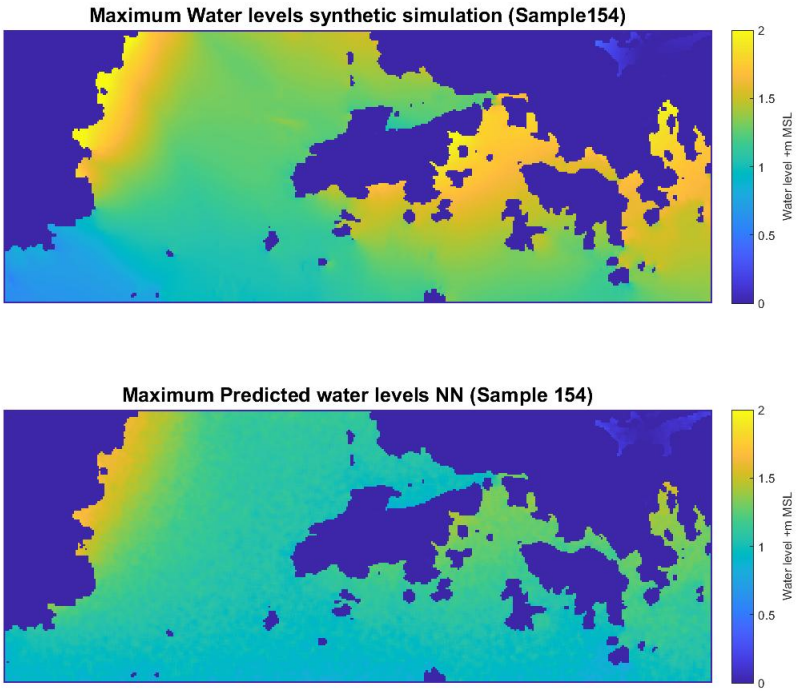


Figure C.3: Training sample 154

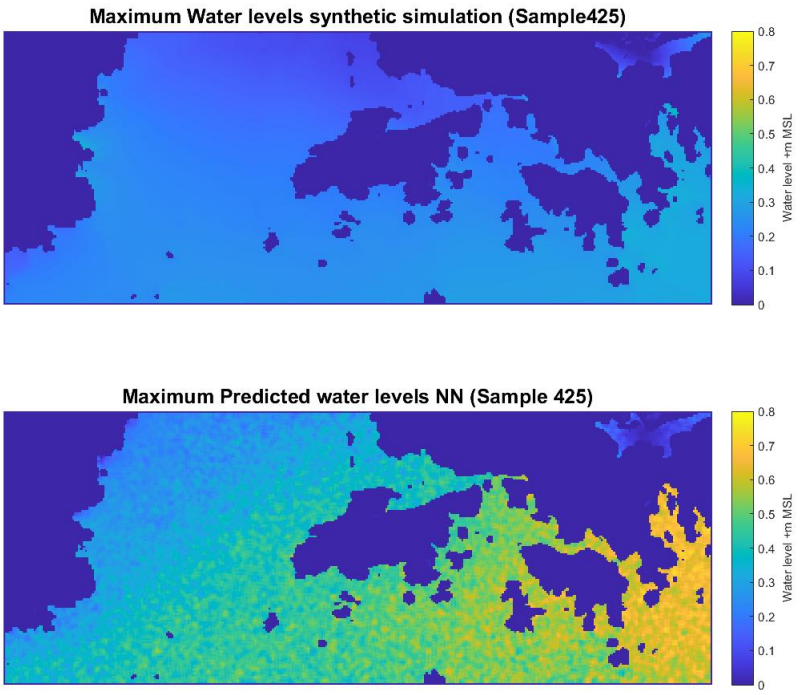


Figure C.4: Training Sample 425

Validation sample comparison

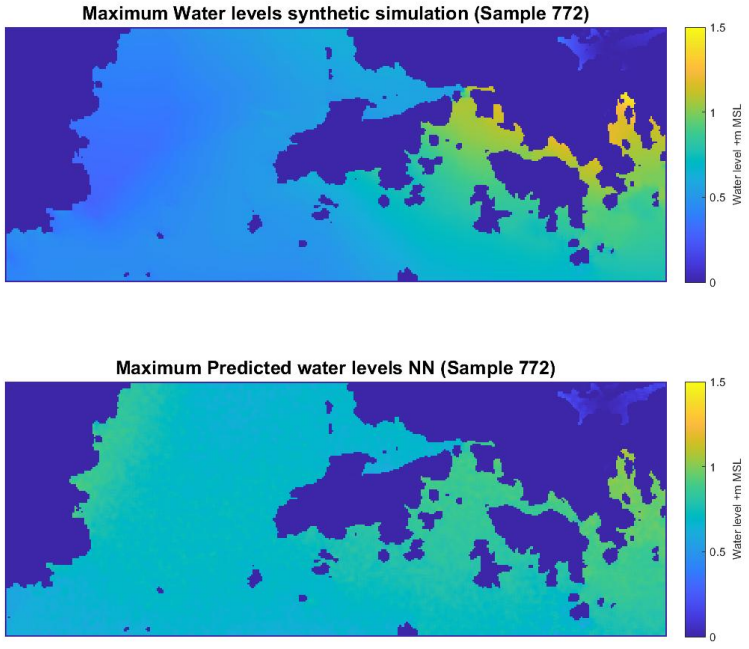


Figure C.5: Validation Sample 772

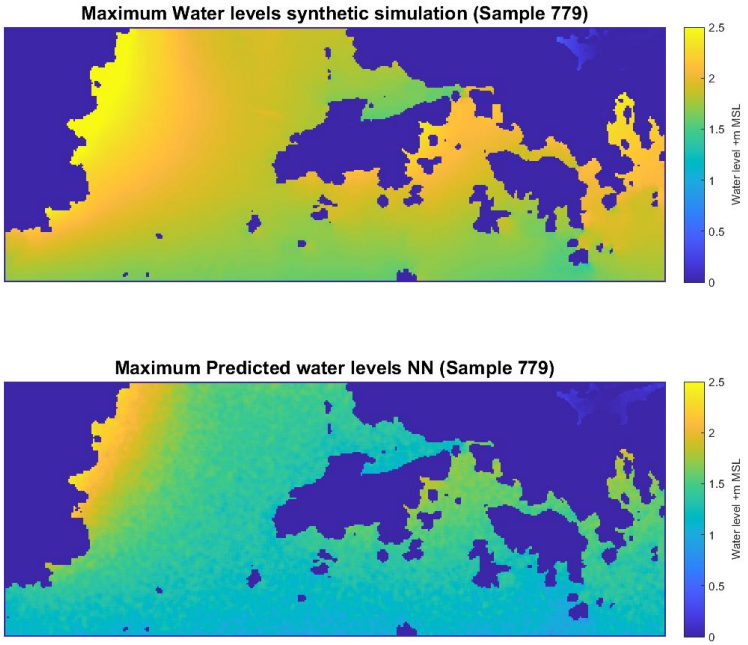


Figure C.6: Validation sample 779

Test sample comparison

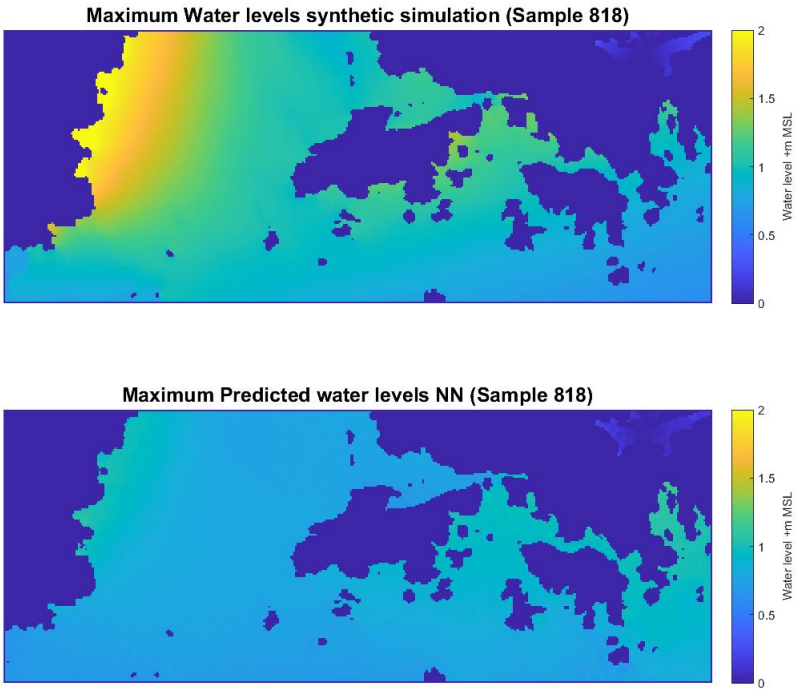


Figure C.7: Test sample 818

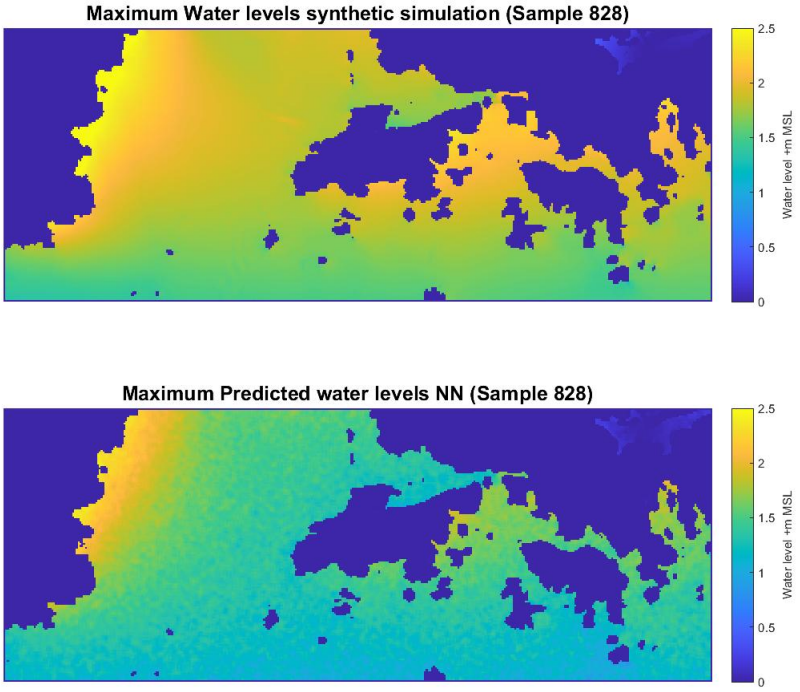


Figure C.8: Test sample 828

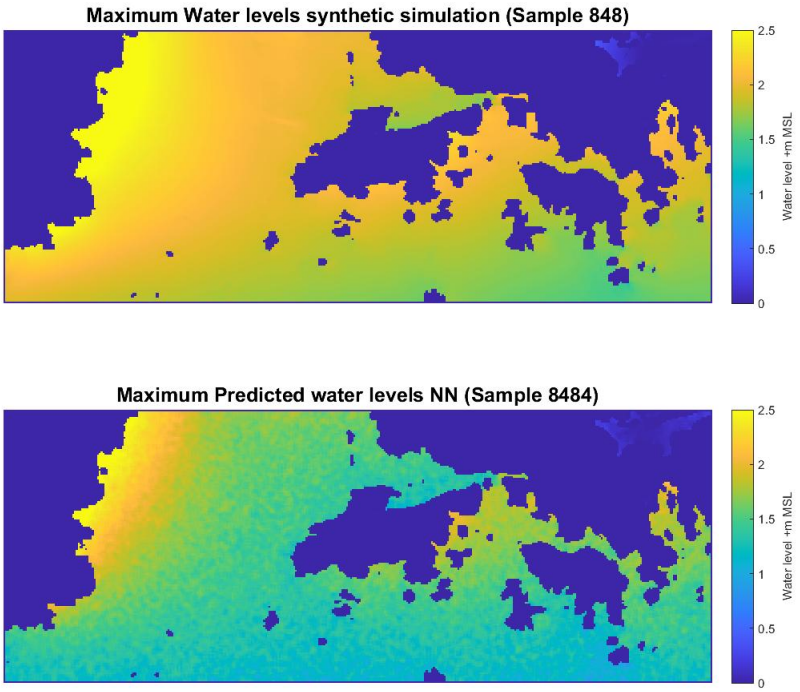


Figure C.9: Test sample 848

C.2. Neural network - Coastline

	Performance error (m)	Training error (m)	Validation error (m)	Test error (m)
50 Nodes	0.0329	0.0329	0.0306	0.0389
100 Nodes	0.0236	0.0247	0.0347	0.027
200 Nodes	0.024	0.0213	0.0388	0.023
500 Nodes	0.042	0.0413	0.0467	0.0402
1000 Nodes	0.0478	0.0416	0.0621	0.1037
1500 Nodes	0.0457	0.0403	0.0599	0.0833
2000 Nodes	0.0808	0.0761	0.1068	0.0921

Table C.1: Coastline network performance

error (m) Sample Number	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mt Davis	West lamma Channel	HK Airport	HK Lung Chau	Macau Airport	Macau City
7	-0.166	-0.170	-0.091	-0.145	-0.126	-0.099	-0.083	-0.048	0.027	0.077
20	-0.297	-0.272	-0.329	-0.272	-0.276	-0.307	-0.277	-0.322	-0.440	-0.392
45	-0.092	-0.111	-0.051	-0.156	-0.088	-0.092	-0.117	-0.042	0.113	0.183
154	-0.100	-0.110	-0.156	-0.074	-0.131	-0.142	-0.107	-0.084	0.272	0.057
425	0.138	0.129	0.075	0.179	0.120	0.090	0.043	0.129	0.075	0.111
772	-0.080	-0.045	0.043	-0.079	-0.026	-0.016	-0.085	-0.040	0.115	0.208
779	0.112	0.163	-0.004	0.112	0.095	0.053	-0.008	-0.020	-0.393	-0.315
796	-0.050	-0.050	-0.038	-0.057	-0.022	-0.059	0.046	-0.077	-0.035	-0.045
818	-0.245	-0.272	-0.284	-0.133	-0.207	-0.188	-0.305	0.019	-0.267	-0.440
828	-0.096	-0.105	-0.079	-0.134	-0.116	-0.094	-0.021	-0.073	0.028	-0.004
848	-0.105	0.004	-0.095	0.006	-0.033	-0.039	-0.153	-0.031	-0.451	-0.359

Table C.2: Sample errors (m) - Coastline Network, 100 Nodes

Relative Error Sample	HK Mawan	HK ct8	HK chc	HK Kowloon Bay	HK Mt Davis	HK West Lamma Channel	HK Airport	HK Lung Chau	Macau Airport	Macau City
7	0.076547	0.078	0.045	0.073	0.061	0.051	0.044	0.025	0.011	0.025
20	0.134238	0.123	0.152	0.135	0.130	0.149	0.129	0.162	0.150	0.115
45	0.077093	0.093	0.055	0.114	0.075	0.087	0.192	0.040	0.438	0.606
154	0.059282	0.065	0.100	0.045	0.079	0.092	0.084	0.054	0.265	0.033
425	0.776729	0.726	0.328	0.791	0.552	0.393	0.329	0.413	0.324	0.465
772	0.075001	0.042	0.056	0.068	0.027	0.018	0.149	0.042	0.386	0.619
779	0.055005	0.080	0.002	0.056	0.047	0.028	0.005	0.010	0.180	0.123
796	0.042397	0.042	0.033	0.056	0.020	0.055	0.036	0.076	0.020	0.021
818	0.202517	0.225	0.261	0.127	0.192	0.187	0.288	0.024	0.192	0.214
828	0.045467	0.050	0.041	0.066	0.057	0.049	0.012	0.040	0.014	0.001
848	0.050832	0.002	0.048	0.003	0.017	0.021	0.083	0.017	0.195	0.135

Table C.3: Relative errors - Coastline network, 100 Nodes

C.3. Neural Network output locations

Configuration		Performance error (m)	Training error (m)	Validation error (m)	Test error (m)
Nodes	Training algorithm				
10	BR	0.0501	0.0479	0.0396	0.0787
10	SCG	0.0775	0.077	0.0761	0.0833
10	LM	0.0494	0.0472	0.0625	0.0543
15	BR	0.0488	0.0428	0.0459	0.0598
15	SCG	0.0618	0.0643	0.0539	0.0489
15	LM	0.0497	0.0474	0.0373	0.0801
20	BR	0.0422	0.0361	0.0506	0.0807
20	SCG	0.0784	0.0726	0.1288	0.0746
20	LM	0.0461	0.0451	0.0527	0.0481
30	BR	0.042	0.0381	0.0562	0.0592
30	SCG	0.0487	0.0521	0.0389	0.0411
30	LM	0.0378	0.0306	0.072	0.0613
50	BR	0.0404	0.0377	0.0337	0.0692
50	SCG	0.0537	0.0506	0.0745	0.0576
50	LM	0.0436	0.0391	0.0775	0.0455
100	BR	0.0456	0.0430	0.0421	0.0704
100	SCG	0.0442	0.0385	0.0486	0.0853
100	LM	0.0317	0.0238	0.0336	0.0933

Table C.4: Neural network configurations performance for 10 output neuron network

Error (m)	**HK**	**HK**	**HK**	**HK**	**HK**	**West Lamma**	**HK**	**Lungchau**	**Macau**	**Macau**
Sample	**Mawan**	**ct8**	**chc**	**Kowloon Bay**	**Mount Davis**	**channel**	**Airport**		**Airport**	**City**
7	-0.28	-0.27	-0.18	-0.18	-0.21	-0.16	-0.24	-0.10	-0.18	-0.12
20	-0.25	-0.26	-0.32	-0.27	-0.30	-0.35	-0.28	-0.40	-0.48	-0.42
45	-0.10	-0.07	-0.04	-0.13	-0.06	-0.07	-0.04	-0.03	0.18	0.16
154	-0.09	-0.11	-0.15	-0.06	-0.12	-0.12	-0.07	-0.10	0.06	0.05
425	0.11	0.13	0.05	0.15	0.11	0.06	0.04	0.11	0.03	0.00
772	-0.13	-0.06	-0.03	-0.10	-0.05	-0.05	-0.09	-0.10	0.09	0.08
779	0.09	0.10	-0.03	0.10	0.07	0.01	-0.06	-0.05	-0.34	-0.32
796	-0.19	-0.22	-0.18	-0.25	-0.21	-0.21	-0.04	-0.31	-0.10	-0.12
818	-0.27	-0.27	-0.22	-0.18	-0.20	-0.18	-0.17	-0.07	-0.30	-0.25
828	-0.10	-0.11	-0.08	-0.13	-0.09	-0.08	-0.05	-0.04	0.04	0.03
848	-0.14	-0.09	-0.14	-0.05	-0.10	-0.11	-0.23	-0.04	-0.34	-0.34

Table C.5: Water level errors for 10 locations - Single layer, 20 hidden nodes network, LM training algorithm

Relative Error	HK	HK	HK	HK	HK	HK West	HK	HK	Macau	Macau City
Sample	Mawan	ct8	chc	Kowloon Bay	Mt Davis	Lamma Channel	Airport	Lung Chau	Airport	
7	0.132	0.124	0.088	0.090	0.102	0.082	0.125	0.052	0.072	0.044
20	0.113	0.120	0.147	0.134	0.142	0.168	0.127	0.201	0.162	0.134
45	0.086	0.055	0.043	0.096	0.052	0.069	0.066	0.029	0.661	0.550
154	0.055	0.062	0.096	0.037	0.072	0.078	0.055	0.064	0.046	0.033
425	0.622	0.651	0.219	0.652	0.502	0.249	0.310	0.345	0.117	0.000
772	0.126	0.055	0.039	0.088	0.052	0.060	0.162	0.107	0.293	0.245
779	0.045	0.048	0.015	0.050	0.035	0.005	0.035	0.025	0.156	0.137
796	0.162	0.194	0.156	0.243	0.191	0.195	0.031	0.308	0.057	0.064
818	0.228	0.226	0.200	0.174	0.186	0.182	0.164	0.088	0.193	0.139
828	0.049	0.051	0.041	0.064	0.044	0.042	0.028	0.022	0.020	0.013
848	0.069	0.043	0.071	0.026	0.051	0.058	0.124	0.022	0.146	0.138

Table C.6: Relative errors for 10 locations - Single layer, 20 hidden nodes network, LM training algorithm

D

Matlab Scripts

Over the course of this research, multiple Matlab scripts have been developed for data processing, file creation etc. In this appendix the purposes and processes of the different script will be discussed. All Matlab scripts are uploaded to the 4TU-data centrum (https://data.4tu.nl/info/).

D.1. Synthetic TC selection tool

This tool is developed to select unique synthetic TC events that fulfill specified selection criteria. Furthermore it edits the data set to make it suitable for use in Delft3D.The synthetic database with 10000 years worth of storms contains over 200000 unique event for the Western Pacific basin. The synthetic database is divided into 10 files with each 1000 years of storms. The files contain three hourly TC track data with the entries from figure D.1. For the purpose of this study we are interested in the storms that will reach close to the Pearl river delta and thus will have an effect on the water levels. To do so, the track data is selected of they meet the criteria based on location, landfall, category and minimal duration. If a storm meets the criteria set, the track data with all time intervals is selected and stored in a separate array.

Entry	Variable name	Unit	Notes on variable
1	Year	—	Starts at 0
2	Month	—	
3	TC number	—	For every year; starts at 0.
4	Time step	3-hourly	For every TC; starts at 0.
5	Basin ID	—	0 = EP, 1 = NA, 2 = NI, 3 = SI, 4 = SP, 5 = WP
6	Latitude	Deg	Position of the eye.
7	Longitude	Deg	Position of the eye. Ranges from 0-360°, with prime meridian at Greenwich.
8	Minimum pressure	hPa	
9	Maximum wind speed	m/s	
10	Radius to maximum winds	km	
11	Category	—	On the Saffir-Simpson scale[19]
12	Landfall	—	0 = no landfall, 1 = landfall
13	Distance to land	km	

Figure D.1: Synthetic storm database entries. Adapted from Bloemendaal et al. (2020)

As can be seen from the figure the TC number or ID starts again at 0 at the start of a new year. This means that if the entire catalogue is considered there are no unique IDs for the storms. This is a problem since Delft3D requires a unique name or ID for each individual simulation. To overcome this problem the tool can automatically assign a unique ID to each TC that is selected on the criteria. The unique IDs are added as an entry to the array.

The fact the we are working with synthetic simulations means that the there are no real dates and times assigned to the data sets. On the other side Delft3D requires a start and stop time/date in order to operate. To make the track data suitable for Delft3D input a time and date must be assigned to each time step. The final product is an array that only contains the selected track data. The track data itself is complemented with unique IDs and times/dates.

D.2. Synthetic Storm Delft3D file creator tool

This Matlab tool can be considered the main script for creating all the required files associated with running a synthetic simulation in Delft3D. It works like the synthetic TC selection tool by selecting the track data of the TC's of interest and assigned them an unique ID. Say we have selected 100 synthetic storm form the first 1000 years of the synthetic catalogue. These 100 storms all require different files and unique names to be able to run in Delft3D. When executing the Matlab script the following processes occur for each unique TC (:

- The script selects that track data of TC i, and creates a new folder names after the TC ID. All files created for TC i will be saved into this folder.

- Creation of spiderweb files. For each computational domain (In this study 3: fine,middle and course domain), a spiderweb file is created from the track data. The spiderweb files contains the wind forcing used in the model to simulate the TC.

- Creation of MDF files. For each domain a MDF file is created. These files are mostly the same for each domain. Only the name ID and duration, stop time in the MDF file change per unique TC simulation. The physical parameters, and grid,bathymetry files all stay the same.

- Creation of.ddb file. The ddb file is used in Delft3D to couple the domains. Since the MDf file names change for each simulation, the content of the .ddb file should change as well. The names of MDF files are changed in the .ddb for every unique TC

- Creation of .sh file. The .sh file is needed for running simulations in the cluster. It is used to start the simulation and contains information on which files to execute. Again some names of the files that must be executed are set to the corresponding ID.

- i+1. After all the files have been printed in the folder, the process start again with a new synthetic TC, until all 100 TC's have their corresponding Delft3D files. The explained steps are repeated for the 10 files containing the 10000 years worth of storm data for the Western Pacific

In the end each folder contains 3 .web and .MDF files, 1 .ddb file and 1 .sh file. The Delft3D files that stay the same for each simulation (grid, depth, boundary files etc) can be manually copies to each folder. Each folder now contains all the required files for Delft3D simulation (figure D.2).

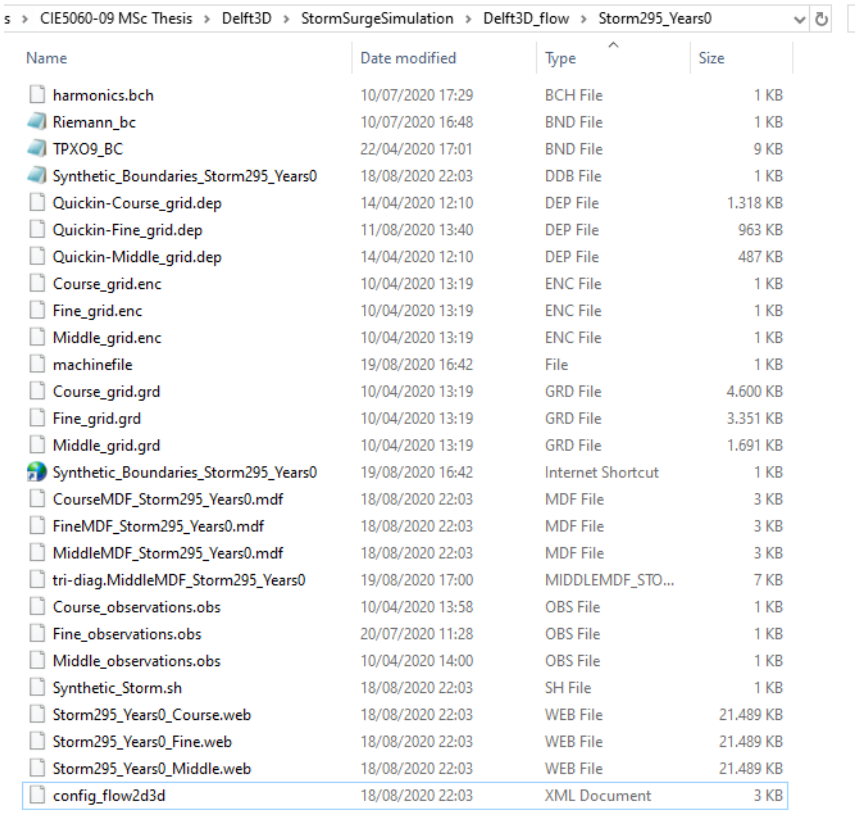


Figure D.2: All Delft3D files needed for cluster simulations - $Storm295\_Years0$

D.3. Delft3D output processing script

The main purpose of the output script is to efficiently process the Delft3D output data. After 1000 simulations, more than 1200 GB of simulation data has been collected. We are only interested in the water levels in the fine domain of the simulation. To process the Delft3D output, the user needs to download Open Earth Tools from Deltares (https://publicwiki.deltares.nl/display/OET/OpenEarth). This toolbox provides Matlab function that can export the required data from the Delft3D results. If OET is installed in the PC, it can be manually loaded into the Matlab program by typing the command below in the Matlab command window. Note that the path changes depending in which folder OET are installed.

```
run('C:\\Users\LucasWestrik\Documents\Deltares\OpenEarthTools\matlab\oetsettings')
```

The Delft3D output files are stored in separate file for each simulation. Therefore, each file must be loaded manually into Matlab. The following command are used and the ID number run up to 100 simulation per 1000 years

```
map_data_001=qpfopen('trim-FineMDF_Storm295_Years0.dat');
l001=qpread(map_data_001);
wl_map_data_001=qpread(map_data_001,l001(15),'griddata',0,0,0);
wl001=wl_map_data_001.Val;
```

These previously described functions, will load the water level for all time steps for one synthetic simulation. The size of the fine domain is 152*401 for each time step. To select the maximum water levels for each synthetic simulation the Matlab function max() is used.This function searches for the maximum values in the array and stores them in a new array. After selection of the maximum water levels and landmasses, the 152x401 array is reshaped into a single row array for NN input (1x60952

array). The final step is to combine the 100 arrays with maximum water levels into one. When this step is repeated for all years, one can combine and save the arrays into one containing all the maximum water level data (NN max water level array: 990x60952). Note that for Nn input the landmasses are removes and the number of output neurons is reduced from 60952 to 38189 output neurons for the NN map version.

D.4. Delft3D track data selector script

The sole purpose of this script is to select the track data corresponding to the maximum water levels. The combined tack data of the unique TC's selected are loaded in to the script. For each unique simulation 1 time step is selected based on the distance of the TC eye from Hong Kong. The time step with the shortest distance (calculated with Pythagoras) is selected. The result is an array with track data (990x7) of 990 simulations with 7 track parameters which is save for NN input.

D.5. Neural Network Training Script

The NN training script is used the Deep Learning Toolbox from Matlab. It requires the user to load the input and output data (Array with track data and array with maximum water levels). The script allows the user to change the configuration of the network. The training algorithm, activation function, normalization function, number of hidden nodes, epochs etc. can all be changed. When the script is executed, Matlab will open the NN training window (figure D.3). This windows displays the state of training and current performance of the network

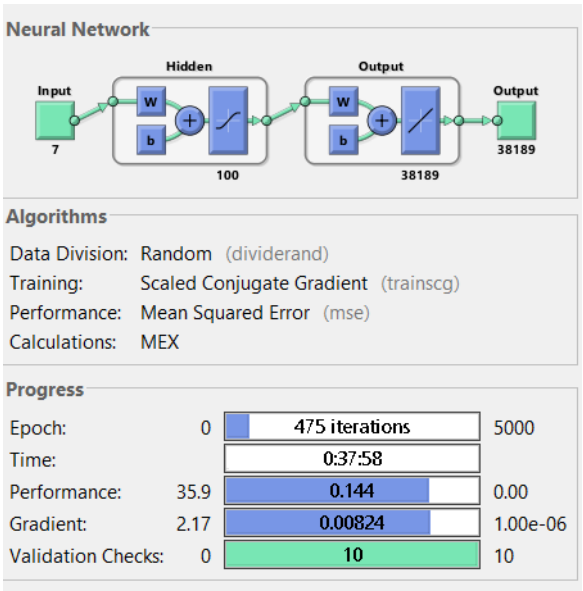


Figure D.3: Neural Network Training Window

When NN training is finished, Matlab creates a NN function for the trained network. This function has 7 input parameters (track data). The outputs are the predicted water levels.

D.6. Neural Network performance analysis Script

This script is developed to investigate the performance of the NN. With this script the user can compare the NN output visually with the output of the synthetic simulations. Furthermore, the difference (error) between the synthetic sims and NN output for specific location of interest are calculated. By checking the performance of multiple different samples, one can gain good insight in the network performance. The trained network function of the three different networks considered in this thesis are uploaded to the datacentre.