Trends in storm surges along the Dutch coast derived from an ensemble of regional climate model simulations for the period 1951-2100

S.P. Aardoom (Sten)

May 11th, 2020

UNIVERSITY OF TWENTE.

Colophon

Title	Trends in storm surges along the Dutch coast derived from an ensemble of regional climate model simulations for the period 1951-2100	
Document type	Master thesis	
Version	1.0	
Date	May 11th, 2020	
Location	Enschede, the Netherlands	

Author	S.P. Aardoom (Sten)	
Student number	s1841246	
E-mail address	stennn@live.nl	
Study specialisation	Water Engineering and Management	
Institute	University of Twente	
Supervisors	Prof. dr. S.J.M.H. Hulscher (Suzanne)	University of Twente
	Dr. J.J. Warmink (Jord)	University of Twente
	Dr. Ir. E.M. Horstman (Erik)	University of Twente
	Dr. H. de Vries (Hylke)	KNMI

Cover image	Photo of a storm surge wave (retrieved from: van den Adel, 2013)

Preface

This preface is written in Dutch.

Voor u ligt het onderzoeksrapport “Trends in storm surges along the Dutch coast derived from an ensemble of regional climate model simulations for the period 1951-2100”. Hierin heb ik model output data, geleverd door het KNMI, gebruikt om trends tussen 1951 en 2100 in stormvloeden langs de Nederlandse kust inzichtelijk te maken. Dit onderzoek is uitgevoerd in het kader van mijn afstuderen voor de masteropleiding Civil Engineering and Management aan de Universiteit Twente, met als specialisatie Water Engineering and Management. Ik ben bij dit onderwerp gekomen door mijn interesse in klimaat verandering. Mijn doel was om een waterdicht onderzoek te realiseren. De eerste maanden waren vooral verkennende maanden, waarin ik onder andere mezelf de programmeertaal Python heb aangeleerd. Hier heb ik tijdens het onderzoek veel aan gehad en zal ik in de toekomst ongetwijfeld ook profijt van hebben.

Tijdens het onderzoek heb ik veel nuttige feedback en inspiratie ontvangen van mijn afstudeerbegeleiders. Allereerst wil ik Jord Warmink en Erik Horstman van de Universiteit Twente bedanken voor de wekelijkse overleg momenten en vele goede ideeën voor mijn onderzoek. Ook wil ik Hylke de Vries van het KNMI bedanken voor het aanleveren van de model output data, de vele goede ideeën en het overbrengen van belangrijke meteorologische kennis van het vakgebied. Verder wil ik Suzanne Hulscher bedanken voor de nuttige feedback tijdens de halverwege-evaluatie en het groen licht gesprek.

Ten slotte wil ik mijn vader, moeder, broertje, zus en alle vrienden die geholpen hebben bedanken voor het lezen, becommentariëren en steunen van het onderzoek.

Sten Aardoom

11 mei 2020, Schalkhaar

Abstract

The implications of climate change are nowadays a hot topic, such as mean-sea-level rise due to global warming. Besides mean-sea-level rise, flood safety is also affected by sea water level extremes. One of the main contributors to these extremes is storm surge, which is the rise of sea water level caused by storms above seas, also known as extra-tropical cyclones. In the past, storm surges have proven to be a major threat to low-lying countries in particular. For example, the North Sea flood of 1953 in The Netherlands. The magnitude of future storm surges and the role of storm surge clusters herein are relatively uncertain, which led to the motivation to conduct this research.

This research investigates storm surge trends in the period 1951-2100 along the Dutch coast for two locations, namely Hoek van Holland and Harlingen. For this purpose, the output data from the WAQUA/DCSMv5 storm surge model forced by RACMO2.2, a Regional Atmospheric Climate Model, were used. This output data consists of 16 members of hourly total water levels, surge height and tide water levels from 1951 to 2100. To simulate climate change, RACMO2.2 was forced with the RCP8.5 (Representative Concentration Pathway), which is the most severe defined greenhouse gas emission pathway.

The Eulerian approach was adopted, which includes the peak over threshold (POT) method and the block extremes method. The POT method was examined and eventually slightly customised to determine the definitions of a storm surge event and storm surge cluster event. The accuracy of the RACMO and WAQUA model combination was determined by calculating water level exceedance frequencies and comparing it to those based on measurement data. It was found that for the entire range of assessed exceedance frequencies (i.e. 1/0.2 till 1/4000 per year) the model combination resulted in systematically lower values. The water level differences between the model and measurement data based exceedance frequencies become larger for smaller exceedance frequencies. This especially holds true for Hoek van Holland. Therefore, it is questionable if this model combination is able to simulate the most extreme storm surge events. Storm surge trends were assessed for three criteria: frequency, intensity and duration of storm surge (cluster) events. For Hoek van Holland, no significant increasing trends were obtained. Only a significant decrease from 1951-1980 to 2071-2100 of the frequency of storm surge events of approximately 0-3 events per 30 years resulted from the trend assessment for Hoek van Holland. For Harlingen mainly an increase in the duration (above the 0.5 meter surge height) of storm surge events of approximately 2-6 hours from 1951-1980 to 2071-2100 is expected. However, the obtained trends are much smaller than the decadal and inter-decadal variability of storm surge activity, leading to the conclusion that the trends cannot be directly attributed to climate change.

Contents

List of terms and abbreviations

Term/abbreviation	Explanation/definition
AR5	Fifth assessment report of the Intergovernmental Panel on Climate Change
Astronomical tide	The tidal levels and characteristics which would result from gravitational effects (e.g. moon, sun and earth), without any atmospheric influences.
(Atmospheric) storm	Characterised by strong pressure gradients and high wind speeds.
Bathymetry	Underwater topography (seabed elevation).
ef	Exceedance frequency
Exceedance frequency	Number of times a water level or surge height is exceeded within one year.
GCM	Global Climate Model (collective term)
Harmonic tide	Theoretical tide divided into its harmonic tide constituents, such as the principal lunar semi-diurnal tide (M_2) and principal solar semi-diurnal tide (S_2).
IPCC	Intergovernmental Panel on Climate Change.
KNMI	Royal Netherlands Meteorological Institute (in Dutch: Koninklijk Nederlands Meteorologisch Instituut).
mb	Member
Member	A model run of a group of model runs.
MSLP	Mean-sea-level-pressure.
NAO	North Atlantic Oscillation (defined as the difference in normalized sea-level pressure anomalies between a southern node, located in the Azores, and a northern node, located in Iceland).
NAP	Normaal Amsterdams Peil, which is the Dutch reference sea water level.
POT	Peak over threshold (methodology).
RACMO(2.2)	The Regional Atmospheric Climate Model used in this research.
RCM	Regional Climate Model (collective term)
RCP	Representative Concentration Pathway
Still water level	The water level without waves.
Storm surge cluster event	A surge event comprising of multiple storm surge events in close succession.
Storm surge event	The occurrence of storm surge at one specific location induced by one atmospheric storm.
Threshold exceedance	An occurrence of the water level or surge height being higher than the POT threshold.
WAQUA(/DCSMv5)	The storm surge model used in this research.

1. Introduction

This chapter introduces the research topic. It includes the research motivation, objective, questions and scope. Furthermore, an overall research methodology is presented, which summarises the research question-specific methodologies described in the corresponding chapters. The chapter ends with a concise reading guide.

1.1. Background

During the existence of the earth, scientists believed the earth encountered natural fluctuations in climate characteristics, such as temperature. However, in the past decades, findings of a more abrupt temperature rise have been published, which is accompanied by a rise in mean-sea-water level. In their fifth assessment report (AR5), the Intergovernmental Panel on Climate Change (IPCC, 2013) states that the relationship between the rising concentrations of greenhouse gases in the atmosphere and the warming of the climate system is now well documented, well understood and beyond reasonable doubt. The global climate is now beyond the Holocene[1] envelope of variability (Steffen et al., 2015), meaning that the observed changes are not solely caused by the internal climate variability of our planet, but that anthropogenic[2] forces are also involved.

In addition to mean-sea-level rise, another implication of climate change is the expected increase in severity of extreme wind events (Feser et al., 2015), suggesting that storms could occur more often and with greater intensity in the future. The combination of mean-sea-level rise and storm events may even lead to more extreme high sea level occurrence, which could have a crucial impact on Dutch coastal defences. Storms are generated by atmospheric low pressure systems, that is extra-tropical storms (Weisse et al., 2012), or also known as extra-tropical cyclones. Storms are characterized by strong pressure gradients and high wind speeds, and they may be accompanied by heavy precipitation, hail, thunder and lightning (Feser et al., 2015). A variety of studies have shown that especially during the winter season a considerable amount of storm variability above the North Sea can be explained by the North Atlantic Oscillation (NAO) (e.g., Yan et al., 2004; Dangendorf et al., 2012). The NAO is traditionally defined as the difference in normalized sea-level pressure anomalies between a southern node, located in the Azores, and a northern node, located in Iceland (Cropper et al., 2015). Nevertheless, Mailier et al. (2006) state that the east Atlantic pattern, the Scandinavian pattern, the east Atlantic-western Russian pattern, and the polar-Eurasian pattern are also of significant influence for explaining the variability of cyclone counts in western Europe.

When storms are situated above seas, storm surges can develop. Storm surge is defined as the rise of the sea surface caused by high wind speeds and low atmospheric pressure (Weisse et al., 2012). In the past, storm surges have proven to be a major threat to coastal areas. Sterl et al. (2009) state that especially low-lying countries like The Netherlands are vulnerable as large areas can easily be flooded. During the last great flood in 1953, nearly 4% of the Dutch territory was inundated, and about 1835 people lost their lives (Sterl et al., 2009). Another factor playing a substantial role in the occurrence of extreme sea water levels is astronomical tides (Caires, 2011; Weisse et al., 2012). The total sea water level can be broken down into the tide water level and surge height, or in other words, a tidal contribution and a non-tidal residual (Pugh and Woodworth, 2015, as referred to in Ridder et al., 2018). When both constituents are simultaneously at their peak, the most extreme sea water level heights are observed. Tidal sea water level variations are caused by the gravitational forces exerted by the moon and the sun (Weisse et al., 2012). Weisse et al. (2012) also state that these forces cause complicated spatially and time varying patterns in ocean sea water level, but these patterns are generally regular and predictable. For large parts of the Dutch coast, north-westerly winds are the main contributor of high surge heights due to the angle of the Dutch coastline. This is because in this case the wind essentially pushes the sea water towards the coast. Surge heights are a result of two meteorological variables, namely mean-sea-level pressure (MSLP) and to a greater extent near-surface wind stress (Ridder et al., 2018; Weisse et al., 2012). They also describe that this MSLP has an inverse impact on sea level heights, known as the inverse barometer effect. This means that an increase in MSLP causes a decrease in sea water level and vice versa. Storm surges are characterized by considerable interannual-to-decadal variability (Dangendorf et al., 2014) and therefore difficult to predict over long timespans.

[1] The current geographical epoch (period) from about 11,700 years ago till the present.
[2] Originated from / caused by humans.

Multiple studies show the presence of storm clusters in western Europe and the North Sea region (e.g., Mailier et al., 2006; Pinto et al., 2014; Priestley et al., 2017). For example, Pinto et al. (2014) demonstrated that secondary cyclogenesis, i.e. new storms that develop on the trailing fronts of previous storms, further contributes to the occurrence of cyclone clusters arriving into Western Europe in rapid succession. Claassen (2018) observed that in many of the clustering events it is not three or more intense storms that hit the Netherlands, but instead they often represent several ordinary low-pressure systems/weak storms. These ordinary storms would on their own not have a large effect. However, when they occur in a cluster, they cause several high sea water level days in close succession (Claassen, 2018). Hence, the term storm surge cluster in this research arises.

Research in the field of storm surge is commonly conducted with the use of two types of data, namely measurement data or model output data. Measurement data result from (tide) gauge records dating back to the end of the 19th century for a large part of the measurement stations along the Dutch coast. A disadvantage of this type of data is that it is often prone to inhomogeneities. Inhomogeneities could cause trends and/or jumps in the data values due to sea level variations caused by global warming, dredging, coastal works and/or morphological changes (Caires, 2011). Water level data, including storm surge heights, from model output data often originates from a storm surge model. This type of model needs atmospheric input to calculate these water levels, which is commonly provided by regional or global climate models (RCM's and GCM's) or reanalysis data. RCM's generally have a much higher spatial resolution than GCM's and are forced at the lateral open boundaries (Feser et al., 2015). These climate models simulate changes based on a set of scenarios of anthropogenic forcing's. The most recent set of scenarios are the Representative Concentration Pathways (RCPs), which are divided into RCP2.6, RCP4.5, RCP6.0 and RCP8.5, in ascending order of severity (IPCC, 2013). In all RCPs, atmospheric CO_2 concentrations are higher in 2100 relative to present day (IPCC, 2013). For reanalyses data, measurement data are used as input for a global weather model in order to achieve a relatively homogenous atmospheric dataset, to gain additional meteorological variables, equal grid spacing and equal time intervals (Feser et al., 2015).

Water level data analyses are commonly done by the Eulerian approach. This approach is used to derive extremes by using block extremes or peak over threshold statistics can be computed (Visser & Petersen, 2012). In an Eulerian approach, block extremes can be obtained by taking the most extreme values in a block of measurements, like a season or year (Feser et al., 2015). For the peak over threshold method a certain threshold can be defined which marks, for instance, the number of storm days per year (Feser et al., 2015).

1.2. Research motivation

Nowadays, the consequences of climate change are becoming more and more notable, which indicates the urgency by which more insight about these consequences needs to be obtained. The documented future implications of climate change are broad. Mean-sea-level rise is a well-known concern for many countries, especially low-lying countries, such as the Netherlands, where great effort has been and still is spent in protecting the land (de Vries et al., 2014). Nevertheless, there is still significant uncertainty in future climate extremes, such as storms and therefore also storm induced flood risks. Feser et al. (2015) published a review in which storm studies over the North Atlantic and northwestern Europe are assessed. They found that future scenarios until about the year 2100 indicate mostly an increase in winter storm intensity over the North Atlantic and western Europe. One could argue that therefore an increase in storm surge intensity is a logical hypothesis. Nevertheless, Christensen state that confidence in future changes in windiness in Europe remains relatively low. Changes in atmospheric circulation and storminess are presently relatively uncertain (Weisse and von Storch, 2010, referred to in Dangendorf et al., 2014). All this indicates the inadequate scientific understanding about the magnitude of future storms and the corresponding extreme water levels. This is also the case for the extreme water level caused by storm clusters, which is a relatively new interest in the scientific research field. Especially in future storm surge clusters are hardly any scientific studies conducted.

Model output data from two models was made available by the Royal Netherlands Meteorological Institute (KNMI). These models are the Regional Atmospheric Climate Model, RACMO2.2 (from here on referred to as RACMO) and the storm surge model WAQUA/DCSMv5 (from here on referred to as WAQUA). This model combination is relatively new in the research field (ran in the period 2013-2017) and therefore not yet analysed completely.

Moreover, the Netherlands was recently in a phase in which flood safety laws were revised. As from the beginning of 2017 new flood safety standards were applied (Kok et al., 2017). However, Kok et al. (2017) also stated that “the job is never finished”, meaning that newly discovered implications of climate change should be accounted for in the future. For this purpose, it is all the more interesting to gain insight in future extreme water level extremes. All reasons mentioned led to the motivation to conduct the present research.

1.3. Research objective and questions

By adopting the hypothesis that climate change results in an increase of storm intensity above the North Sea (see also Feser et al., 2015) and therefore also in an increase in storm surge intensity along the Dutch coast towards the end of the 21st century, the research objective is to test this hypothesis. Therefore, the research objective is to quantify and assess storm surge trends in the period 1951-2100 at the Dutch coast based on the model combination of RACMO and WAQUA. The results can then potentially be used as input for the determination of future Dutch storm related flood safety standards. The main objective entails three sub-objectives. Firstly, the definition of a storm surge event and storm surge cluster event need to be determined as a basis for the rest of the research. These definitions can then be used to perform the quantification and assessment of trends for three criteria: frequency, intensity and duration of storm surge (cluster) events. Secondly, for the purpose of interpreting the resulting trends, the model accuracy (also known as model bias) needs to be quantified. Thirdly, for the same reason physical processes in the WAQUA output data need to be identified. To achieve the research objective four research questions were formulated. These research questions are defined as follows:

1. *What is a suitable quantitative definition of a storm surge event and a storm surge cluster event?*
2. *How well does the, by RACMO forced, WAQUA model output data represent physical processes and observed water level exceedance frequencies?*
3. *What are trends in frequency, intensity and duration of storm surge (cluster) events (WAQUA) in the period 1951-2100?*
4. *What are the future implications of the obtained storm surge trends on the Dutch coast?*

The first three research questions are extensively elaborated on in the corresponding chapters. The fourth research question is only elaborated on in the discussion.

1.4. Research scope

The scope of the present research is described by five factors. Firstly, the research focusses on storm surge and therefore atmospheric processes are mostly only briefly addressed. This entails that only the WAQUA output data is analysed. The reason for this is that this research is conducted from a flood safety perspective and thus not from a climatological perspective. Secondly, the research is limited to analysing still water level. Still water level is defined as the water level without waves. A study of trends in wind-driven waves is another discipline and waves are also not included in the WAQUA model. Thirdly, also not included in the WAQUA model is mean-sea-level rise. The effect of mean-sea-level rise on storm surge heights is therefore not analysed but only discussed by consulting literature. Fourthly, the presence and influence of astronomical tides on the sea water level is analysed, but the astronomical background of tides is only briefly addressed. Fifthly and finally, the geographical case study of this research is the Dutch coast. While the model domain includes at least locations around the entire North Sea basin, no other regions are analysed. The reason is the limited time resources available for this research.

1.5. Overall research methodology

In this section the overall research methodology is briefly addressed. For the answers to each research question the used methodologies can slightly differ. These methodologies are in more detail described in the chapters corresponding to these research questions. All scripts/codes in this research are written in the programming language Python.

The data used in this research is produced by a combination of two models: Regional Atmospheric Climate Model (RACMO) (van Meijgaard et al., 2008, 2012) and a storm surge model (WAQUA). For a more detailed description of these models is referred to chapter 2. The data from both models for this research was made available by the KNMI. The output data of the WAQUA model consists of water level time series of surge, tide and surge plus tide (total water level). For this research, the WAQUA output of two out of 41 locations, namely Hoek van Holland and Harlingen, are used. For each location 16 members (runs) of 151 years (1950-2100) of these data are available. These 16 members are a result of the forcing of 16 corresponding RACMO members with 16 different initial and lateral boundary atmospheric conditions. Therefore, the WAQUA data can be considered as 16 independent members of 151 years each. In this research, the initial year (1950) of these data are kept unused.

The general approach used to derive extremes in this research is the Eulerian approach, which is commonly used to derive extremes from water level or discharge data sets and to make statistical statements about these extremes (Caires, 2011; Feser et al., 2015; Tijssen & Diermanse, 2009; van den Brink & Können, 2011). Two methods from the Eulerian approach are used to obtain extremes, namely the peak over threshold (POT) method and the block extremes method. The POT method is used to determine the definitions of a storm surge event and storm surge cluster event.

To quantify the model accuracy, the from the POT method obtained total water level extremes were used to calculate corresponding exceedance frequencies. For this purpose, the Generalised Pareto Distribution (GPD) is fitted to these selected extremes. Thereafter, the calculated exceedance frequencies were compared to exceedance frequencies of measurement data based exceedance frequencies. In addition to the model accuracy analysis, the POT method output is examined to identify and visualise physical processes, such as tide cycles and surge-tide interaction.

To quantify and assess storm surge trends, obtained surge height and total water level extremes were used. The trend assessment is done for three criteria: frequency, intensity and duration of storm surge. The entire data timespan (1951-2100) was divided into periods of 30 years, in which successive periods overlap with 29 or 15 years (running mean), depending on the criterion. For the intensity and duration criteria the results are visualised in boxplots. Subsequently, linear trend lines were fitted to the frequency results and to the intensity and duration boxplot variables. The obtained trend lines are tested on statistical significance.

1.6. Reading guide

The present research report is organised as follows. In chapter 2 the available model output data, the case study (locations) and the measurement data based information are described. In chapters 3, 4 and 5 the methodology and results of the research questions (section 1.3) are presented, so each of these chapters include a methodology and a results section. Of these three chapters, chapter 3 describes the process leading to the definitions of a storm surge event and a storm surge cluster event, chapter 4 describes the physical processes identified in the data and the model accuracy, and chapter 5 describes the storm surge trend assessment. Finally, in chapter 6 the research is discussed and the conclusions are presented.

2. Data availability and case study

This chapter describes the two models (RACMO and WAQUA) and their output data. Furthermore, the chosen case, two locations along the Dutch coast, are presented. Finally, water level exceedance frequencies based on measurement data are given and how these exceedance frequencies are calculated is described.

2.1. Available model output data

The KNMI have been using several models to simulate climate change and related developments from 1950 until 2100. They provided the output data of two models for this research, namely RACMO2.2 and WAQUA/DCSMv5. This section provides a description of these two models.

2.1.1. RACMO2.2

The first model is RACMO2.2 (van Meijgaard et al., 2008, 2012), which is an acronym for Regional Atmospheric Climate Model version 2.2 (from here on referred to as RACMO). It simulates changes in climate based on a set of scenarios of anthropogenic forcing's, namely the Representative Concentration Pathways (RCPs), which represent greenhouse gas emission projections. RACMO simulates atmospheric processes, such as wind, pressure, temperature, precipitation, etc., for the period from 1950 until 2100 for a region including western Europe and a part of the North Atlantic Sea. The exact model domain is shown in Figure 2-1. To simulate the atmospheric and physical processes RACMO runs on the ECMWF CY23r4 physics package and the HIRLAM dynamics (Lenderink et al., 2003). Since RACMO is a Regional Climate Model (RCM), it needs input from the lateral boundaries over the entire timespan it is ran. Additionally, an initial condition over the entire grid should also be provided as input for RACMO. An overview of the functioning and output of RACMO is provided in Table 2-1.

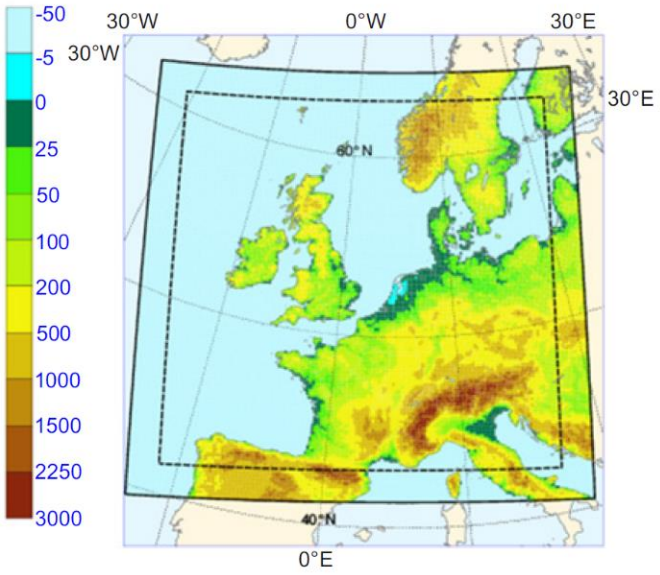

Figure 2-1. Model domain (outer box) of RACMO. The colours denote the land surface altitude in meters. (van den Hurk et al., 2015).

2.1.2. WAQUA/DCSMv5

The second model is the WAQUA/DCSMv5 model, from here on referred to as WAQUA. This is a storm surge model that computes total water levels, storm surge heights and astronomic tide water levels (Illustrated in Figure 2-2). It is developed by the RAND corporation (USA) in cooperation with the Dutch public works authority (Rijkswaterstaat), Deltares, and the KNMI (van der Grinten, 2011). The WAQUA model solves the two-dimensional shallow water equations to predict water levels at several locations along the coast of the North Sea (Ridder et al., 2018; Svasek Hydraulics, n.d.). The bathymetry[1] used in WAQUA is based on nautical maps (Figure 2-3), which were subsequently calibrated (Ridder et al., 2018). The model is

[1] Underwater topography (seabed elevation).

operated at a resolution of $\frac{1}{12}^{\circ} \times \frac{1}{8}^{\circ}$ (latitude and longitude respectively), which is roughly 8 km × 8 km, and the model calculates water levels at 10-min intervals in a two-step process (Ridder et al., 2018). First, the model solves the shallow water equations for the tidal component of total water level only. For this the model uses ten harmonic constituents to predict the water level at the open boundaries of the model domain and calculates their progression through the basin (Sterl et al., 2009). The second step adds the meteorological forcing by wind (at 10 meters above mean-sea-level) and mean-sea-level pressure (MSLP) to the calculations of the harmonic tide (Ridder et al., 2018). This meteorological forcing is provided by the wind and MSLP output from RACMO. The provided forcing fields are interpolated bilinearly in space and linearly in time to match model requirements, with a forcing frequency of multiples of 3 h (Ridder et al., 2018). The provided wind forcing is translated into wind stress through the application of a wind speed-dependent drag coefficient, proposed by Charnock in 1955, applying a constant Charnock coefficient: $\alpha_{Ch} = 0.032$ (Mastenbroek et al., 1993). Although the model is operated on a horizontal grid, the output data is stored for a selection of locations along the coastlines of the North Sea basin. An overview of the functioning and output of WAQUA is provided in Table 2-1.

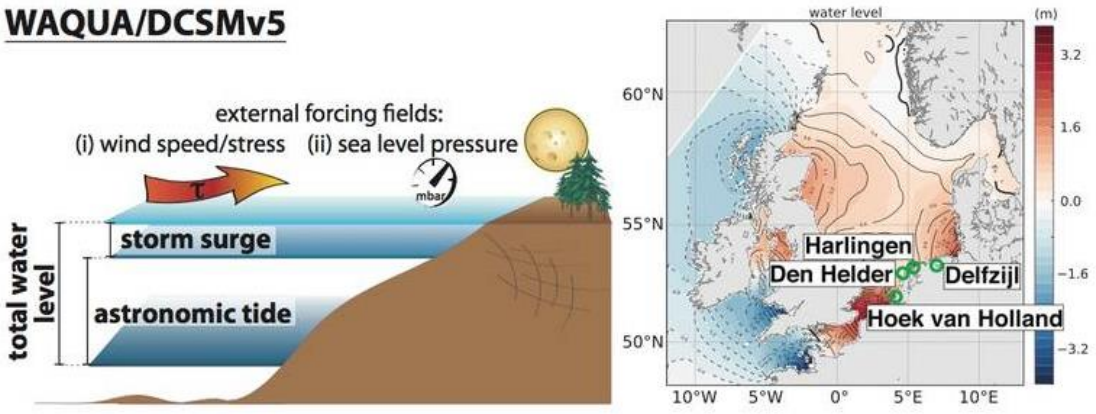

Figure 2-2. Illustration of WAQUA, which calculates total water levels, astronomic tides and storm surge heights. The latter is a result of external forcing fields of wind and mean-sea-level pressure. (KNMI, n.d.).

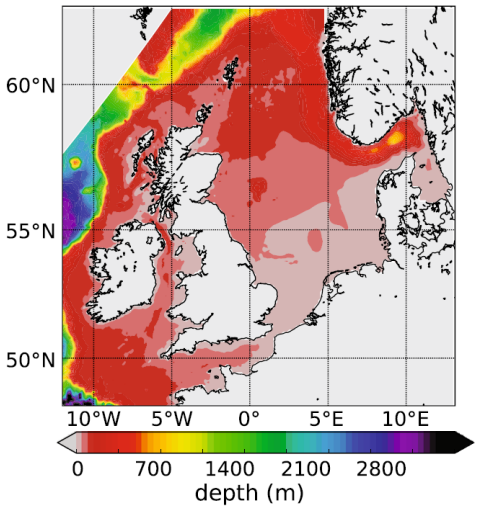

Figure 2-3. Bathymetry applied in the WAQUA model based on calibrated nautical maps. Shown is the North Sea and its surroundings. The colours denote the water depth in meters. (Ridder et al., 2018).

2.1.3. Model runs and output data

The KNMI ran RACMO 16 times for the period 1950-2100. Each of these 16 ensemble members was forced at the lateral boundaries with 16 corresponding runs of EC-Earth, a global climate

model system (European EC-Earth consortium, 2019). This resulted in different initial atmospheric states and different forcing at the lateral boundaries for each of the 16 RACMO members. Therefore, the weather and climate developed in a different manner for each of the 16 RACMO members, resulting in 16 independent members. This also means that the RACMO output data of past years does not represent the real climate which actually occurred in those years. For example, extreme events, such as storms, which happened in reality cannot be found in the output data. The 16 model runs are carried out in the period 2013-2016 at the KNMI. All 16 model runs of RACMO are ran with RCP8.5, which is the most severe Representative Concentration Pathway. This represents a hypothetical pathway of greenhouse gas emissions, which are a global driver of climate change. RACMO calculates atmospheric quantities in the spatial three-dimensional field, but only the wind at 10 meters above sea water level and MSLP are relevant as input for the WAQUA model.

The WAQUA model is ran with these 16 RACMO members as input, thereby also providing 16 corresponding members of output data for the period 1950-2100. Since the 16 RACMO members are independent, the 16 WAQUA members are also independent. The data consist of time-series at 10-min and 1-hour intervals for 41 different locations along coastlines around the North Sea basin. The WAQUA runs are carried out in the period 2016-2017.

Table 2-1 provides an overview of information about the models and their output data relevant for this research.

Table 2-1. Information about the models (RACMO and WAQUA) and their output data relevant for this research.

	RACMO	WAQUA
Input		
Forcing	At lateral boundaries with information obtained from EC-Earth RCP8.5 (Most severe Representative Concentration Pathway)	RACMO variables: wind speed, wind direction and mean-sea-level-pressure
Perturbation	Each run is slightly perturbed in atmospheric state	-
Bathymetry	-	Based on nautical maps
Functioning		
Physics & dynamics	Runs on the ECMWF CY23r4 physics package and HIRLAM dynamics	Solves the two-dimensional (horizontal) shallow water equations
Includes	Roughly all atmospheric parameters, such as precipitation, sea level pressure, wind, etc.	Storm surges and astronomic tides
Output data		
Space dimension	2D (grid)	0D (separate locations)
Time interval(s)	3 hourly and daily	10 min. and hourly
Time span [years]	1950-2100	1950-2100
Location(s)	North Atlantic and large part of Europe	41 locations along the coastlines of the North Sea basin
Amount	16 members (runs)	16 members (runs)
Quantities	Wind speed, wind direction and mean-sea-level-pressure	Total water levels, surge heights and tide water levels
Other information		
		Not included: mean-sea-level rise, waves and temporal changes in bathymetry

2.2. Case study: Dutch coast

In this research storm surge and corresponding atmospheric conditions are examined at the Dutch coast. Two locations along the coastline are chosen to represent the Dutch coast most efficiently, being Hoek van Holland and Harlingen (Figure 2-4). These locations are chosen because they are nationally considered as main stations for measurements and other studies have been conducted for these locations. Therefore, more information for these locations is available. Furthermore, both locations are situated very differently. Hoek van Holland is directly adjacent to the North Sea, whereas Harlingen is located in the Wadden Sea basin. Basins like the Wadden Sea could have significant different results in terms of tides and storm surges due to influences of resonance and smaller water depths. Also, the coastline orientation to the North Sea is different. Hoek van Holland is facing northwest, whereas the North Sea coastline near Harlingen is facing more towards the north.

Figure 2-4. Locations of Hoek van Holland and Harlingen in The Netherlands.

2.3. Available measurement data based information

The combination of RACMO and WAQUA is an approximation of reality and therefore a model accuracy analysis is conducted (chapter 4). In this model accuracy analysis, the WAQUA model output is compared with water level exceedance frequencies based on measurements. This section provides the water level exceedance frequencies and calculation method as obtained from and described in Dillingh (2013) and Rijkswaterstaat, (2013).

Table 2-2. Exceedance frequencies of water levels in Hoek van Holland and Harlingen applicable to the beginning of the year 2011 and based on measurement data of the period 1900-2010 for Hoek van Holland and 1933-2010 for Harlingen. Water levels corresponding to exceedance frequencies of 1/500, 1/1000, 1/2000 and 1/4000 are rounded to increments of 0.10 m. The other water levels are rounded to increments of 0.05 m. Water levels corresponding to higher exceedance frequencies (grey part) are calculated from peak values from threshold exceedance peaks (non-independent storm surges) and water levels corresponding to lower exceedance frequencies (white part) are calculated from peak values of independent storm surges. (Dillingh, 2013; Rijkswaterstaat, 2013).

Exceedance frequency per year	Water level [m +NAP]	
	Hoek van Holland	Harlingen
1/4,000	4.70	4.90
1/2,000	4.50	4.80
1/1,000	4.30	4.70
1/500	4.10	4.50
1/200	3.80	4.30
1/100	3.60	4.15
1/50	3.40	4.00
1/20	3.15	3.75
1/10	3.00	3.55
1/5	2.85	3.30
1/2	2.60	3.05
1/1	2.45	2.80
1/0.5	2.30	2.55
1/0.2	2.10	2.25

Table 2-2 shows the water levels at the locations of Hoek van Holland and Harlingen corresponding to the exceedance frequencies ranging from once in 4000 years to 5 times per year (1/0.2 per year in the table). This range is considered to be sufficient for the model accuracy analysis, because the WAQUA model output consists of 2416 years in total (16 members of 151 years), meaning that a water level corresponding to an exceedance frequency of once in 4000 years cannot be directly calculated from the WAQUA data. This would require an extrapolation method. The measurement data are of the period 1900-2010 for Hoek van Holland and 1933-2010 for Harlingen. However, the values are calculated such that they are applicable to the beginning of the year 2011 for both locations. This is realised by eliminating mean-sea-level trends, NAP[2] changes and the nodal tidal cycle. For example, the measurement data contain mean-sea-level rise which is corrected for. Furthermore, the method how the water levels are calculated differs for lower exceedance frequencies (white part of table) and higher exceedance frequencies (grey part of table). The lower exceedance frequencies are partly calculated directly from the data and partly with the use of a Generalised Pareto Distribution (GPD), which is the best performing extrapolation method to calculate the more extreme water levels resulting from the POT method (Dillingh, 2013). For this the peak water levels of selected storm surge events (independent peak water levels) are used. The higher exceedance frequencies are calculated from the peak water levels of all threshold exceedances (non-independent peak water levels). The threshold exceedance peak water levels were sorted from highest to lowest. Subsequently the exceedance frequencies were calculated by taking the number of the peak water level in the sorted list and dividing by the period of the used data in years (Dillingh, 2013). For example, the third highest peak water level from a data set of 100

[2] Normaal Amsterdams Peil, which is the Dutch reference sea water level.

years would have an exceedance frequency of 3/100, which results in about once every 33.3 years. To ensure a smooth transition between the resulting values from these two calculation methods a pragmatical approach is applied. This smooth transition is realised visually by plotting both calculation results in one graph and connecting the resulting lines of both calculation methods.

An important remark on the water level values in the table, is that the water levels are rounded to not suggest unjustified precision (Dillingh, 2013). This is due to uncertainties which result from the two different calculation methods mentioned in the preceding paragraph. Water levels corresponding to exceedance frequencies equal to or lower than once in 500 years (i.e. 1/500, 1/1000, 1/2000 and 1/4000) are rounded to increments of 0.1 meters and all other shown water levels are rounded to increments of 0.05 meters. Furthermore, the water levels shown in Table 2-2 are relative to NAP. Therefore, it is necessary to take the mean water levels relative to NAP into account to be able to perform the comparison of measurement data based and model data based exceedance frequencies with the same reference water level. Table 2-3 shows these mean water levels relative to NAP for both locations applicable to the year 2011. The values are both slightly higher than 0, which can be attributed to mean-sea-level rise.

Table 2-3. Mean water levels relative to NAP in Hoek van Holland and Harlingen applicable to the year 2011. (Dillingh, 2013).

	Hoek van Holland	Harlingen
Mean water level [m +NAP]	0.09	0.07

3. Defining and selecting storm surge events

The main focus of this chapter is finding definitions of a storm surge event and storm surge cluster event. The former is defined by analysing the output of the general peak over threshold (POT) method, which is originated from the Eulerian approach, on the WAQUA data and by considering applied definitions in related studies. In addition, the POT method output that is used in the remaining part of this research is described.

3.1. Methodology

The POT method is used to define the definition of a storm surge event and storm surge cluster event. The following sections describe the process leading to the implemented POT method, including adaptions to make it suitable for the desired purpose.

3.1.1. The general peak over threshold method

The POT method makes use of pre-defined water level threshold values to select threshold exceedances from the concerning variable in time. Figure 3-1 provides an illustration of the POT method. Shown is a water level time series at a certain location (e.g. surge height or surge plus tide water level). In this illustration the water level threshold, from here on referred to as the POT threshold, is set on 1.90 m. Three threshold exceedances are selected in the shown time span. A threshold exceedance is defined as the occurrence of the water level time series being higher than the POT threshold. For every exceedance the peak value is calculated, which results in three peak values (red crosses in Figure 3-1). Note that, because of this, the first exceedance does not contain two peak values. By obtaining peak values, threshold exceedance frequencies can be calculated (i.e. the number of peaks within a certain period).

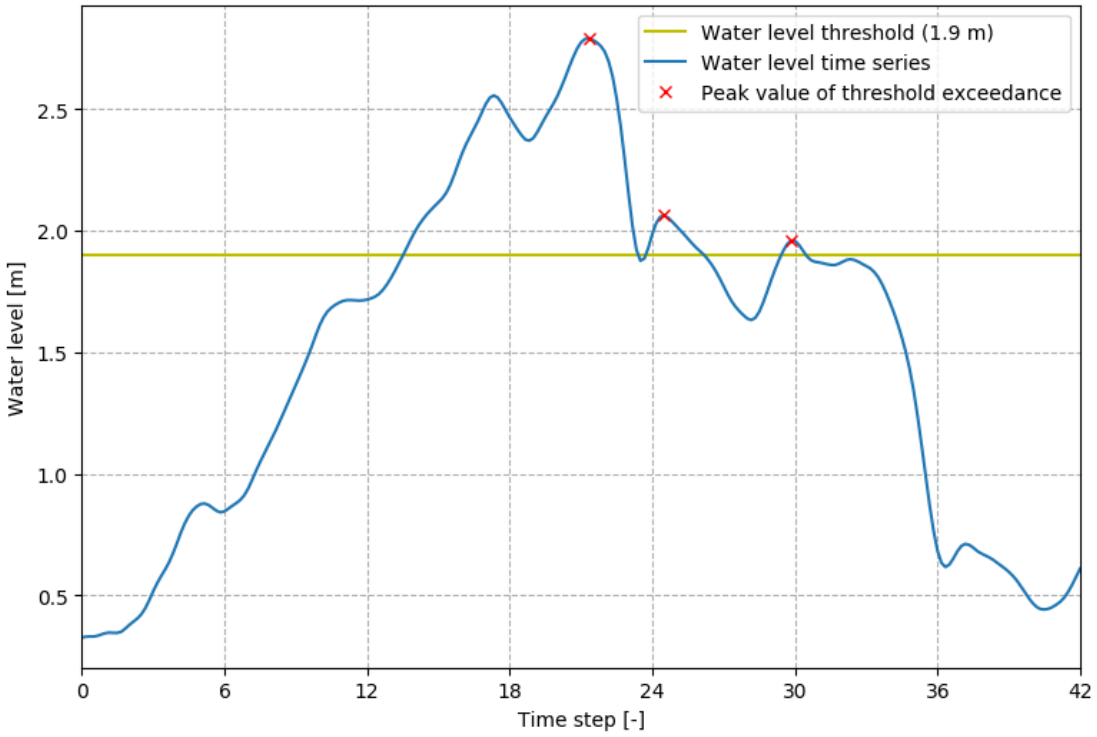

Figure 3-1. Illustration of POT method showing a water level threshold of 1.90 m (yellow), synthetic water level time series (blue) and the peak values of the threshold exceedances (red). The unit of time used in this research is hours.

3.1.2. Used water level variable and storm surge event independency

To find the most suitable definition of a storm surge event two factors are considered. Firstly, the decision is made of which water level variable will be used in the POT method. Secondly, the factor of successive threshold exceedance dependency is addressed.

A water level variable is chosen on which the POT method is applied. Literature points out there are two variables commonly used to analyse storm surge activity, which are the storm surge residual and the skew storm surge (Caires, 2011; Dillingh et al., 1993; Tijssen & Diermanse, 2009). The storm surge residual is defined as the total water level minus the astronomical tide. The skew storm surge is defined as the difference between the astronomical high tide water

level and the nearest experienced total high water level for each semi-diurnal[1] period (Caires, 2011; Dillingh et al., 1993). The choice has been made to analyse the storm surge residual, because the duration of storm surge is one of the criteria in this research. By analysing the storm surge residual, the duration of threshold exceedances can be calculated. This is rather difficult and cumbersome for the skew storm surge variable due to the fact that the peaks in total water level and tide water level do not have to coincide in time (Tijssen & Diermanse, 2009). In addition, the storm surge residual is calculated by the WAQUA model and can thus be directly used to perform the POT method on. From here on, the storm surge residual is referred to as the surge height. The POT method is applied on the hourly data, because this led to acceptable script run times.

The second factor that influences the definition of a storm surge event is successive threshold exceedance dependency. Statistical conclusions of the selected threshold exceedances can only be made if successive threshold exceedances are occurring virtually independent from each other. Firstly, successive water levels are autocorrelated (Dillingh et al., 1993), meaning that a water level at a certain time has influence on the water level at the next time step. Secondly, the first runs of the general POT method over the WAQUA surge height data contain many successive threshold exceedances within a relatively short time interval. When successive threshold exceedances are occurring in a too short time interval, they could belong to the same atmospheric storm event. To ensure that the resulting successive threshold exceedances are virtually independent two options are explored. The first option is using a time interval threshold between successive threshold exceedances and the second option is using a certain surge height threshold, not to be confused with the POT threshold, between successive threshold exceedances. For the first option the time interval between successive threshold exceedances has to be larger than a time interval threshold to ensure virtual independency. For the second option the surge height between successive threshold exceedances has to be lower than this threshold to ensure virtual independency. Literature shows the first option is more common (e.g., Caires, 2011; Van Den Brink & Können, 2011) and is therefore also chosen in this research. To find the most suitable time interval threshold value a time interval analysis is conducted.

3.1.3. Time interval analysis

For storm surge events to be classified as virtual independent, a time interval analysis is conducted. Herein, the number of occurrences of different time intervals between successive threshold exceedances resulting from the POT method are visualised and quantified. Additionally, the resulting threshold exceedances for some higher POT thresholds were visually examined to obtain information about the shape of the threshold exceedances, i.e. the combination of duration, peak surge height and inclination of surge heights in time. This is done by generating plots of the surge height in time at the times of threshold exceedances. The following output variables are calculated by using the POT method:

- Start time of each threshold exceedance
- End time of each threshold exceedance
- Time interval between each two successive threshold exceedances

Figure 3-2 illustrates how the enumerated POT method output variables are calculated. The start time of each threshold exceedance is calculated at the first time step when the water level variable is larger than the POT threshold. The end time of each threshold exceedance is calculated at the first time step when the water level is below the POT threshold. The time intervals between each two successive threshold exceedances are calculated at the POT threshold by subtracting the start time of the threshold exceedance from the end time of the previous threshold exceedance for all successive threshold exceedances. Figure 3-3 shows the result for a POT threshold value of 1.90 meter at Hoek van Holland. This result shows that the largest part of intervals between 0 and 72 hours are below 12 hours (i.e. 94.9%), suggesting that the successive threshold exceedances corresponding to these time intervals have the

[1] A semi-diurnal period is about 12.41 hours and thus occurs approximately twice a day (see also section 4.2.1).

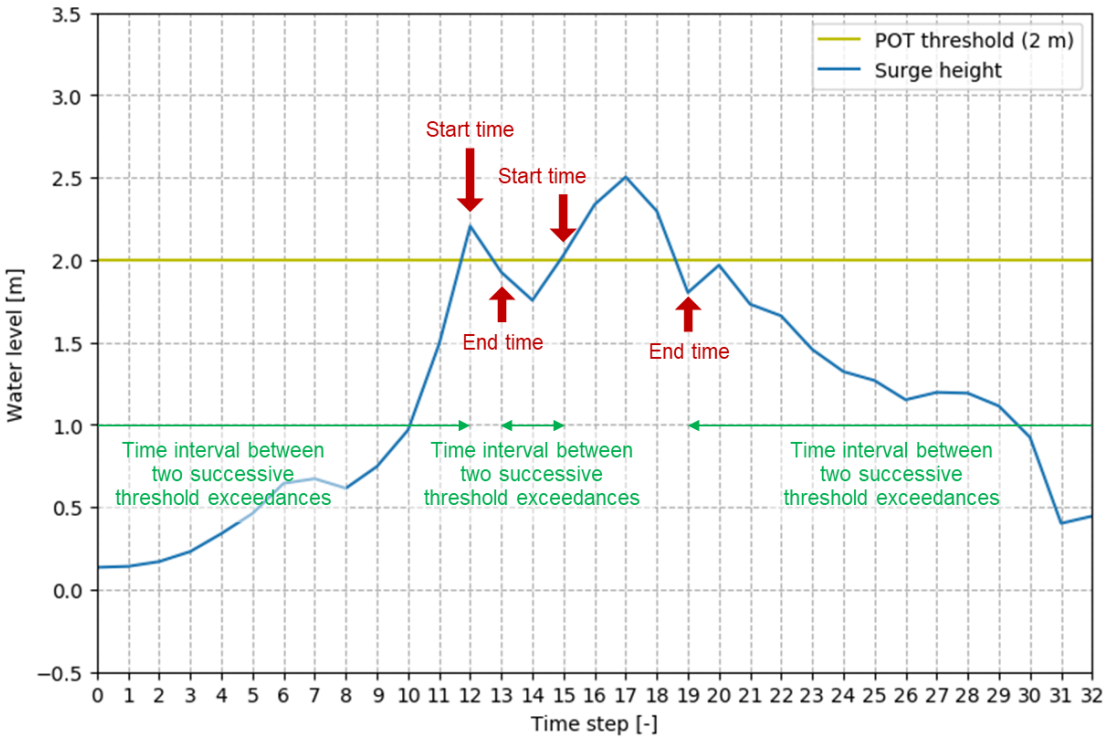

Figure 3-2. Calculation method illustration of POT output variables: start time and end time of threshold exceedances (red) and time interval between each two successive threshold exceedances (green).

same underlying cause (i.e. belonging to the same atmospheric storm). This result is compared with the threshold exceedance plots. Figure 3-4 shows an example of these plots in which three successive threshold exceedances within a time span of less than 24 hours are present. Based on the surge variable (green line) all three threshold exceedances are part of the same 0.5 m water level exceedance suggesting that all successive threshold exceedances in this example are part of the same atmospheric storm. For Harlingen a similar result as for Hoek van Holland was found. Figure 3-5 shows that a large part of intervals are below 12 hours (i.e. 89.0%), which also suggests that this part of intervals are successive threshold exceedances which belong to the same atmospheric storm. Here a POT threshold of 2.35 meters was chosen, because this resulted in a similar number of occurrences in total as for Hoek van Holland. Literature shows that implemented time interval thresholds are in a range of 24 hours (Tijssen & Diermanse, 2009) to 3 days (van den Brink & Können, 2011). Dillingh et al. (1993) state that the average

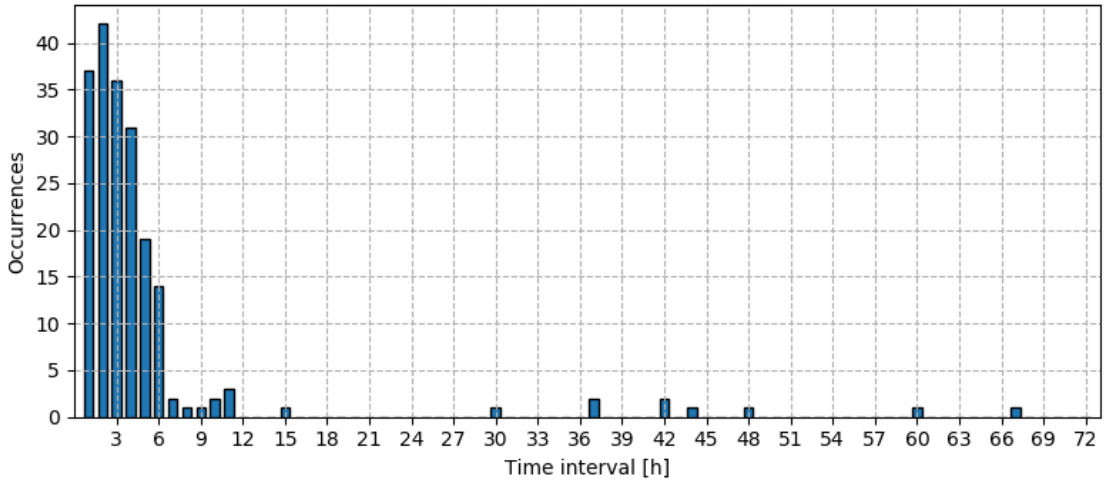

Figure 3-3. Occurrences of time intervals between successive threshold exceedances as a result of the POT method on the WAQUA surge height time series with a POT threshold of 1,90 m at Hoek van Holland.

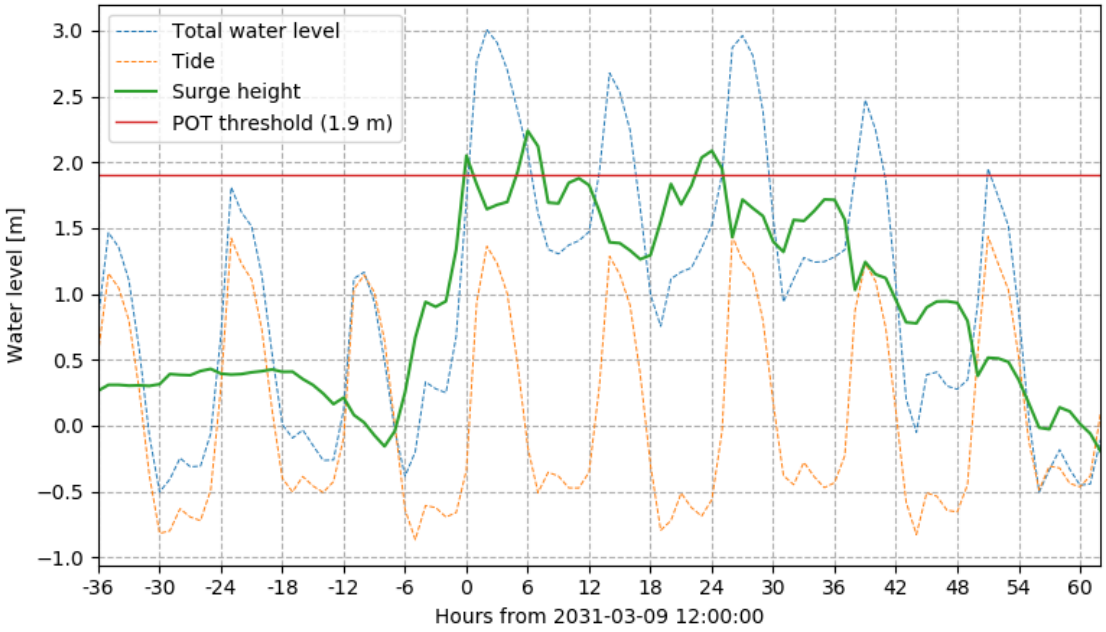

Figure 3-4. Visualisation of an example of successive threshold exceedances in Hoek van Holland. The POT method, with a POT threshold of 1.90 m, is used on the WAQUA surge height time series to select the threshold exceedances.

duration of North Sea storms is of about 2.41 days. By taking these findings into account, the decision is made to implement a time interval threshold of 24 hours. This value is chosen because it is a compromise between the results of the conducted time interval analysis and the values found in literature. The implemented time interval threshold operates such that successive threshold exceedances occurring within a time interval shorter than the time interval threshold are aggregated as one threshold exceedance, from here on defined as a storm surge event. Hence, one storm surge event can consist of multiple threshold exceedances. For example, the threshold exceedances shown in Figure 3-4 are part of the same storm surge event.

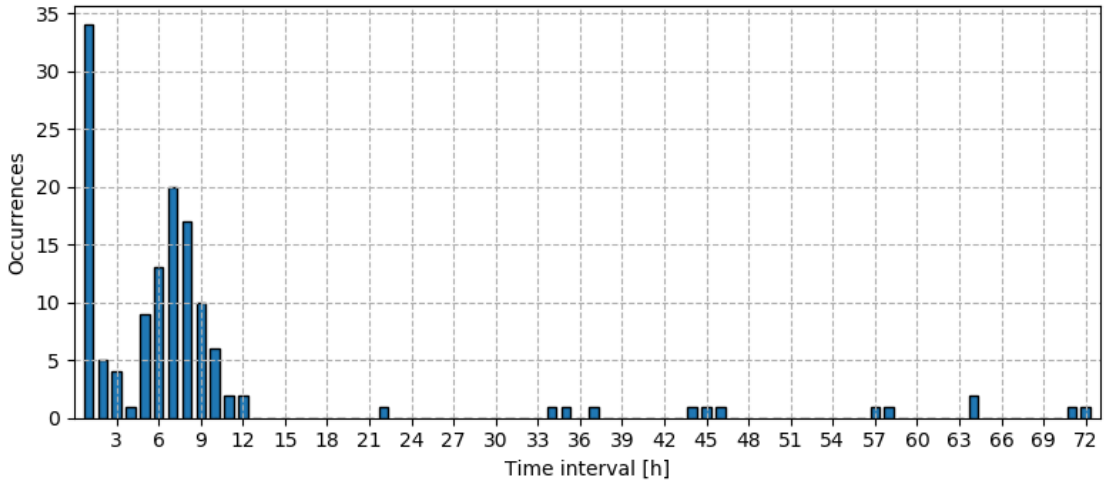

Figure 3-5. Occurrences of time intervals between successive threshold exceedances as a result of the POT method on the WAQUA surge height time series with a POT threshold of 2.35 m at Harlingen.

3.1.4. Storm surge quantification

Since one of the objectives of this research is to examine trends in frequency, intensity and duration of storm surge events, output from the POT method to quantify these criteria is needed. For this purpose, the following output variables are calculated by using the POT method:

- Number of storm surge event in certain periods
- Peak surge height of each storm surge event
- Peak surge plus tide water level (i.e. total water level) of each storm surge event
- 0.5 meter surge height duration of each storm surge event

Figure 3-6 illustrates how the enumerated POT method output variables are calculated. The peak surge height output is used to quantify the intensity criterion of the trend analysis and is calculated by finding the maximum surge height value between the start and end time of each storm surge event. The peak surge plus tide water level value output is also used to quantify the intensity criterion of the trend analysis and is calculated by finding the maximum between 12 hours before the start time and 12 hours after the end time of each storm surge event. The value of 12 hours is chosen because of two reasons. Firstly, it is necessary to implement such an interval because the maximum total water level caused by the atmospheric storm is not necessarily occurring within the start and end time of the storm surge event. This is due the timing of the high tide, which is highly determinative in when the maximum total water level is occurring. An interval of 12 hours ensures that at least one entire tide cycle period is included at both temporal sides of each storm surge event. Secondly, since the time interval threshold of 24 hours is applied in the definition of a storm surge event (section 3.1.3), an interval of 12 hours is the maximum to be sure that two successive storm surge events cannot have the same peak total water level. In other words, the temporal search area of two successive storm surge events could overlap if an interval larger than 12 hours is applied. The WAQUA surge plus tide water level is from here on referred to as the total water level. The 0.5 meter surge height duration is used to quantify the duration criterion of the trend analysis and is calculated by searching backwards along the WAQUA surge time series until it finds the first value above the 0.5 meter (start) and similarly by searching forward until it finds the first value below the 0.5

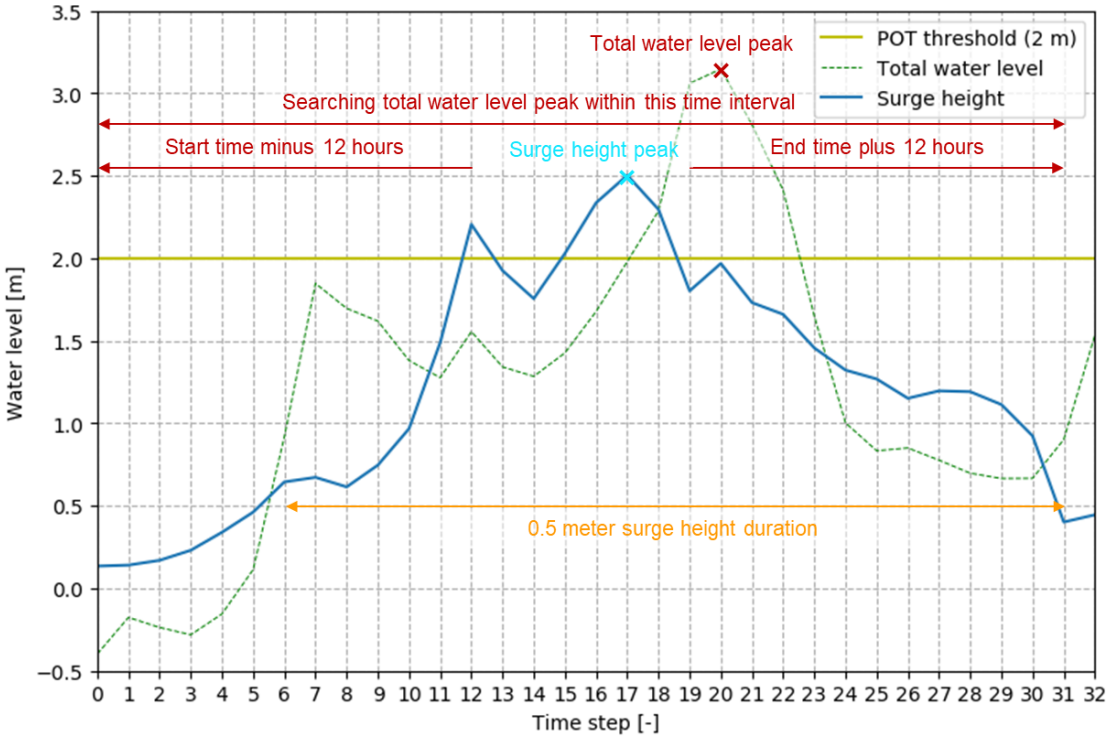

Figure 3-6. Calculation method illustration of POT output variables: peak surge height (cyan), peak total water level (red) and 0.5 m surge height duration (orange) of storm surge events.

meter (end). Subsequently, these values are subtracted and the duration value is calculated. This 0.5 meter surge height is used, because this corresponds with the research done by (Tijssen & Diermanse, 2009). They also state that a direct duration calculation at the surge height of 0.0 meter would not work due to the natural fluctuation of the surge height. In other words, “there is a lot of ‘noise’ that cannot be attributed directly to the storm” (Tijssen & Diermanse, 2009). Therefore, the duration at 0.5 meter surge height is considered to best represent the duration criterion.

Frivolous to mention is that in the POT method applied on Harlingen one storm surge event occurred directly from the start time of the data set. This storm surge event is removed from the POT results, because no duration could be calculated due to the missing start time of the storm surge event.

3.1.5. POT threshold stability analysis

Since the POT method requires a POT threshold, it was considered necessary to examine which POT thresholds are suitable for the analysis of storm surge events. The POT thresholds need to be in a certain range to ensure the validity of the results. This is because a too high POT threshold value leads to a too small sample size, resulting in a high variance (Caires, 2011). On the other hand, a too low POT threshold value will shift the focus from analysing extreme values to more common values. Therefore, a POT threshold stability analysis is conducted in which the standard error of a range of POT thresholds is calculated.

The threshold stability analysis indicated that for Hoek van Holland POT thresholds higher than 2.3 meters are showing instable results due to too low sample sizes. The lower bound of the POT threshold for Hoek van Holland is set on 1.5 meters to ensure extreme values are examined and not the more common values. For Harlingen the upper bound is set on 2.7 meters and the lower bound is set on 1.9 meters.

3.1.6. Defining a storm surge cluster event

The definition of a storm surge cluster event is determined in the same manner as a single storm surge event, i.e. by using certain threshold and time interval conditions. Since a storm surge event is a result of a storm, the definition of a storm surge event is used to define a storm surge cluster event. Hence, a storm surge cluster event is the result of multiple storm surge events occurring in close succession. To find the most suitable definition of a storm surge cluster event the definition of close succession and a cluster are determined. This ensued partly from the definition of a storm surge event and partly from literature. For example, Claassen (2018) defines storm clusters as three or more storms which hit in close succession. She also states that close succession is defined as a maximum of three non-high water level days (72 hours) in between two high water level days. Finally, a cluster event is defined as a sequence of at least three days where high water levels occur in close successions, i.e. with a maximum of three days in between each occurrence of high water levels, and a minimum of one non-high water level day in between the three (or more) days where the threshold is exceeded (Claassen, 2018). As a result of the definition the frequency of storm surge cluster events is added to the storm surge quantification output variables from section 0.

3.2. Results

The results presented in this section are the results of the definition of a storm surge event and a storm surge cluster event.

3.2.1. The definition of a storm surge event

In this section the definition of a storm surge event is given. It is based on the analyses conducted in the preceding sections of this chapter. The final definition is described as follows and examples are illustrated in Figure 3-7.

Definition:

In this research a storm surge event is defined as the maximum possible number, one or multiple, of surge height POT threshold exceedances wherein the maximum time interval between each two successive threshold exceedances is 24 hours. occurring within a 24 hour time interval between each two successive exceedances. Herein, POT threshold is defined as a surge height threshold which can vary per analysis, and the time interval is calculated from the end of the first exceedance to the start of the second exceedance at the POT threshold.

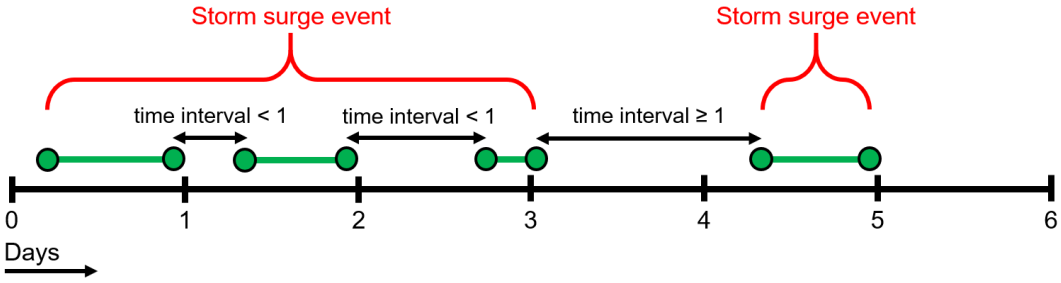

Figure 3-7. Illustration of examples of storm surge events. Shown is a time line with steps of one day, threshold exceedances (green) occurring within the time line, if the time interval threshold of one day (24 hours) between threshold exceedances are satisfied, and finally storm surge events (red) consisting of one or multiple threshold exceedances.

3.2.2. The definition of a storm surge cluster event

The definition of a storm surge cluster event is based on the definition of a storm surge event and the definition from (Claassen, 2018) as described in the methodology. It requires a minimum and maximum time interval threshold between successive storm surge events and a minimum number of storm surge events in a cluster. Because in the definition of a storm surge event a maximum time interval threshold of 24 hours is used, this is also chosen as the minimum time interval threshold for storm surge cluster events. The maximum time interval threshold is based on the one used in (Claassen, 2018), which is 72 hours. The minimum number of storm surge events in a storm surge cluster event is determined to be 2 events, because as Pinto et al. (2014) stated that, in the case of storm clusters, new storms develop on the trailing fronts of previous storms, suggests that storm surge cluster events could also consist of just 2 storm surge events. However, next to a minimum value of 2, also 3 and 4 are analysed to obtain insight in the sensitivity of the storm surge cluster event definition. The final definition is described as follows and an example is illustrated in Figure 3-8.

Definition:

A storm surge cluster event is defined as a minimum of 2 storm surge events occurring within a 24 hour and a 72 hour time interval between each two successive storm surge events. Herein, the time interval is calculated from the end of the first storm surge event to the start of the second storm surge event at the POT threshold.

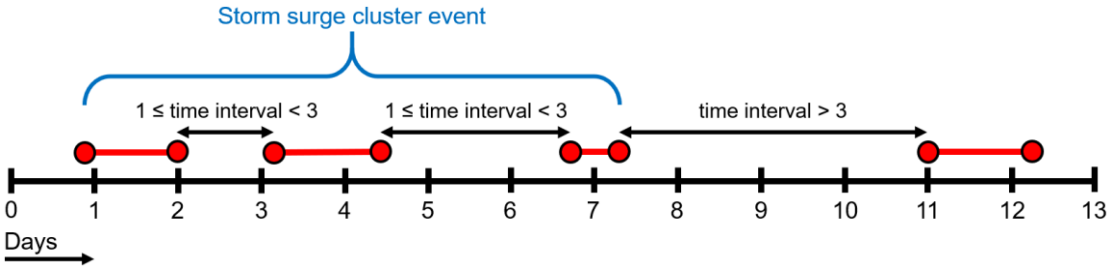

Figure 3-8. Illustration of an example of a storm surge cluster event. Shown is a time line with steps of one day, storm surge events (red) occurring within the time line, if the time interval thresholds between storm surge events are satisfied, and finally a storm surge cluster event (blue) consisting of three storm surge events.

4. Physical processes in the data and model accuracy

To interpret results regarding trends and statistics it is essential to obtain an adequate understanding of what physical processes are included in the model output data. Furthermore, the RACMO and WAQUA model combination is an approximation of reality and could thus have a bias in comparison to measurement data based water levels. This comparison is made to represent the model accuracy. This analysis is done to inform readers that linear comparisons with studies based on measurements should not be made.

4.1. Methodology

4.1.1. Physical processes

Identifying physical processes from both models' output data is essential to obtain an adequate understanding of what the data represents. Physical processes, such as the tide and its cycles and the interaction between the surge and the tide are examined. Known physical processes derived from literature are identified in the data by generating graphs which provide visual confirmation. These graphs originate from the output of the peak over threshold (POT) method described in chapter 3.

4.1.2. WAQUA data corrections for calculation of model accuracy

As mentioned before, the WAQUA and RACMO model combination is an approximation of reality and therefore a model accuracy analysis is conducted (also known as model bias). In this model accuracy analysis, the WAQUA output is compared with exceedance frequencies of water levels based on measurements, which are assumed to represent reality. For information about the origin and calculation method of the measurement data based values is referred to section 2.3.

In this section the WAQUA data correction for the calculation of the model accuracy is presented. Corrections are applied to ensure correctness in the comparison between measurement data based values and WAQUA model values. This refers to the fact that for the measurement data based values the mean-sea-level trends, NAP changes and the nodal tide cycle are corrected for. These factors are eliminated and a reference time corresponding to a mean-sea-level is chosen, that is the beginning of the year 2011. These steps are also done for the WAQUA water level data to determine if corrections need to be applied and to what extent. No corrections for mean-sea-level trends needed to be applied, since this phenomenon is not included in WAQUA.

The first step is the calculation of the mean water level of the reference year/period. For this purpose, the mean of the period 2006-2015 of the WAQUA data is calculated. This period is chosen because it is considered to best represent the mean water level for the year 2011 (i.e. the measurement data based reference period), since this year is exactly in the middle of the period. To calculate such a mean water level the effect of the nodal cycle within this period is analysed, because this cycle has an effect on the mean water level. To determine if corrections for the nodal tide signal should be applied, the mean over the entire data timespan (1951-2100) of the annual mean tide water levels is calculated. Subsequently, the annual mean tide water levels are subtracted from the entire data timespan mean for each of the years within the period 2006-2015. Thereafter, these nodal tide signal correction values were added to the mean total water levels of each of the years in the period 2006-2015. This resulted in corrected annual mean total water levels. Finally, one mean total water level was calculated from these annual mean total water levels.

The second step was to account for the difference in reference water level between the measurement data based values and WAQUA model based values. The mean water level relative to NAP based on measurement data are given in Table 2-3. The difference for both locations between these values and the mean water level values (2006-2015) was calculated. The entire WAQUA total water level time series were subtracted with these correction values to obtain water level data with the same reference as the measurement data based reference water level.

4.1.3. Model accuracy

As mentioned in the previous section, the model accuracy is represented by comparing the water level exceedance frequencies of WAQUA with those based on measurement data. For the calculation of the WAQUA water level exceedance frequencies the calculation method used for the measurement data based values is replicated as much as possible.

The WAQUA data set consists of 2416 years in total (16 members of 151 years). In this analysis the initial year of each member is not used, since this type of model is known for its incorrectness at initial time steps. To be sure this phenomenon is eliminated a whole year is kept unused. This results in a remaining total of 2400 years (16 members of 150 years). Hence the WAQUA data period is 1951-2100 versus the measurement data period of 1900-2010.

First, the POT method is applied on the corrected (i.e. corrected according to section 4.1.2) WAQUA total water level (surge + tide) time series. Total water level is used, because the measurement data based values also represent total water levels. A POT threshold of 2.00 meters was chosen, because this resulted in a large enough sample of peak water levels to be able to calculate water levels corresponding to higher exceedance frequencies. A different POT threshold would not affect the water level exceedance frequency values itself, but only the amount of values.

The exceedance frequencies were calculated in two ways. The first method is by first sorting the peak water levels from highest to lowest. Subsequently, the rank number of the peak water level in the sorted list is taken and divided by the period of the used data in years (i.e. 2400 years). For example, the third highest peak water level from the WAQUA data set would have an exceedance frequency of 3/2400, which results in once in every 800 years. The 16 members can be aggregated in this method, because the WAQUA output data members are independent (section 2.1.3). The water level exceedance frequencies are calculated with this method for the threshold exceedance peak water levels (non-independent peak water levels) and the storm surge event peak water levels (independent peak water levels), because both are used for the final result as is also the case for the measurement data based water level exceedance frequencies.

The second method by which the exceedance frequencies are calculated is by fitting the Generalised Pareto Distribution (GPD) to the selected peak water levels with the POT method. This distribution is commonly used and considered to be suitable in combination with the POT method (e.g. Caires, 2011; Dillingh, 2013). The GPD function which describes exceedance frequency as a function of water level is described as follows (Dillingh, 2013; Geerse & Wojciechowska, 2014; Philippart et al., 1995; Zhao, Zhang, Cheng, & Zhang, 2019):

$$q(x) = q_\mu \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}, \qquad \xi \neq 0, \qquad x - \mu \geq 0 \tag{4-1}$$

Where:

q $= exceedance\ frequency\ [/year]$

q_μ $= exceedance\ frequency\ corresponding\ to\ threshold\ \mu\ [/year]$

x $= water\ level\ [m]$

μ $= water\ level\ threshold\ or\ location\ parameter\ [m]$

ξ $= shape\ parameter$

σ $= scale\ parameter$

The function can also be written as the water level as a function of exceedance frequency:

$$x(q) = -\frac{\sigma}{\xi}\left(1 - \left(\frac{q}{q_\mu}\right)^{-\xi}\right) + \mu, \qquad \xi \neq 0, \qquad q \leq q_\mu \tag{4-2}$$

Eq.(4-2) can be used to extrapolate to water levels higher than the highest water level in selected peak water levels.

To use Eq.(4-1) and Eq.(4-2) the parameters $\mu, \xi\ and\ \sigma$ need to be determined. For the parameter μ Dillingh (2013) used the water level corresponding to an exceedance frequency of 0.5 per year (hence $q_\mu = 0.5$) for the measurement data based exceedance frequency calculation. Since, this method will be replicated as much as possible this same approach is applied. The parameters ξ and σ are estimated by using the Maximum Likelihood (ML)

estimation described in Zhao et al. (2019), because this method is also suggested by Choulakian & Stephens (2001) and Geerse & Wojciechowska (2014). Also the estimation method used for the measurement data based exceedance frequency calculation was not clear from the literature. Castillo and Serra (2015, referred to in Zhao et al., 2019) found a new and efficient estimation of the ML estimation for the parameters ξ and μ of the GPD. Eq.(4-3) and Eq.(4-4) are derived from this, which are the shape parameter ξ as a function of k and the corresponding profile-likelihood function respectively.

$$\xi(k) = \frac{1}{n}\sum_{i=1}^{n} ln\left(1 + \frac{x_i - \mu}{k}\right), \qquad 1 + \frac{x_i - \mu}{k} > 0 \tag{4-3}$$

$$l^*(k) = -n\left[ln\left(k * \xi(k)\right) + \xi(k) + 1\right] \tag{4-4}$$

Where:

$k \quad = \frac{\sigma}{\xi}$

$l^* \quad = profile\text{-}likelihood$

The ML estimation $\hat{k}$ of k is obtained by maximising the profile-likelihood function (Eq.(4-4) and Figure 4-1), so the shape and scale parameters can be estimated by $\hat{\xi} = \xi(\hat{k})$ and $\hat{\sigma} = \hat{\xi} * \hat{k}$ (Zhao et al., 2019). This is done numerically with a precision for $\hat{k}$ of 3 decimals. The resulting parameters were substituted in Eq.(4-1) and Eq.(4-2) by which the exceedance frequencies and corresponding water levels were calculated.

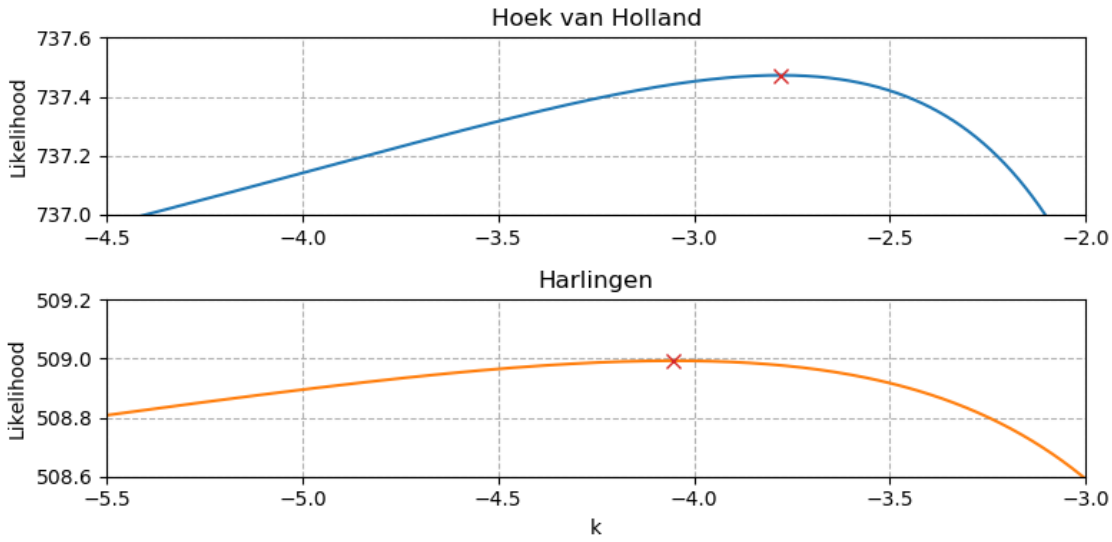

Figure 4-1. Maximising the profile-likelihood function to find $\hat{k}$.

The resulting water level exceedance frequencies from both calculation methods were plotted in the same figure for Hoek van Holland and Harlingen. In this way the results from both calculation methods could be compared and for each location one final relation between water levels and exceedance frequencies was determined. In this procedure, the transitions between the exceedance frequencies of the threshold exceedance peak water levels (non-independent peak water levels) and the storm surge event peak water levels (independent peak water levels) were constructed by connecting both results (lines) pragmatically.

4.2. Results

The results section of this chapter elaborates on the tide, surge and tide interaction, and finally presents the model accuracy.

4.2.1. Tide

The tide at the Dutch coast is approximately semi-diurnal and contains a spring-neap cycle and a nodal (or precession) cycle (Dillingh, 2013; The Open University, 1999). Semi-diurnal means that one day contains about two high and two low tides. The period of this cycle is approximately 12.41 hours. This is a result of the earth rotating around its own axis (24 hours) and the moon rotating around the earth. The spring-neap cycle is caused by the earth and the moon rotating around their common centre of mass and as a whole rotating around the sun. This full rotation period is about 29.5 days, which means the spring-neap cycle period is approximately 14.75 days (i.e. 29.5/2) (The Open University, 1999). The nodal cycle is a result of the declination of the moon, meaning that the orbit of the moon around the earth shifts in angle relative to the orbit plane of the earth around the sun (Dillingh, 2013; Kaye & Stuckey, 1973; The Open University, 1999). The period of this cycle is about 18.6 years. More detailed information about the astronomical background of tides is not provided, because it is beyond the scope of this research.

The mentioned three cycles are visible in the WAQUA data. An example of a randomly chosen time period of the tide water level in Hoek van Holland is given in Figure 4-2. In here the semi-diurnal and the spring-neap cycles are clearly visible. The semi-diurnal cycle can also be identified as a mixed diurnal cycle, since most successive high tides are not of equal magnitude. This is also valid for the low tide.

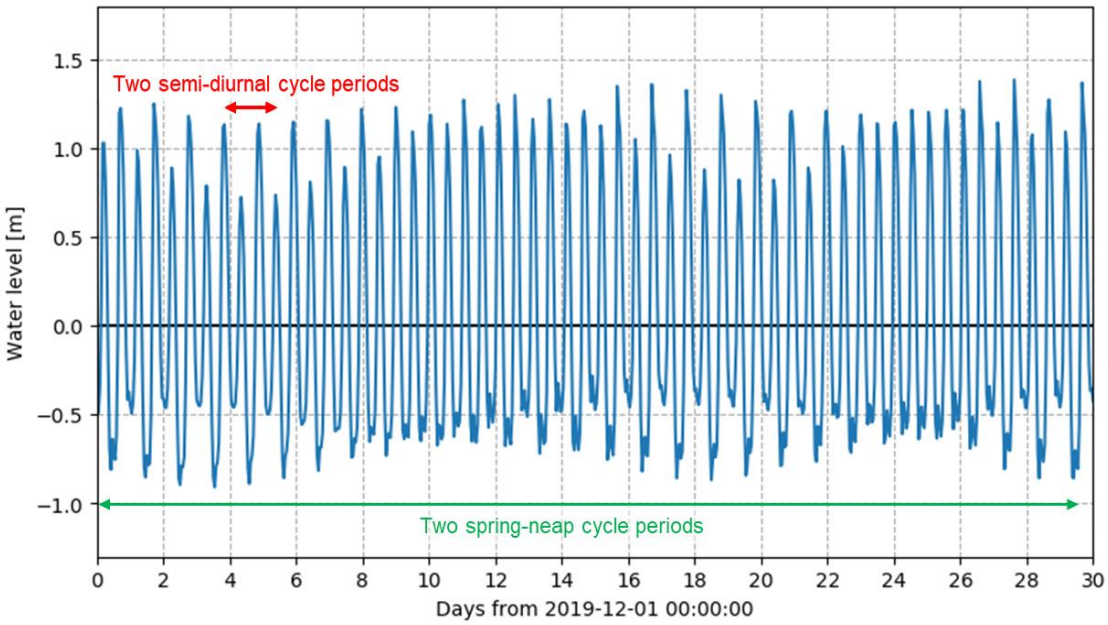

Figure 4-2. Visualisation of the WAQUA tide (blue) in Hoek van Holland. Also indicated are the duration of two semi-diurnal cycle periods (red) and two spring-neap cycle periods (green).

The nodal cycle has a much larger period and is depicted in Figure 4-3 for Hoek van Holland and Harlingen. Herein the shown lines represent the annual mean tide water levels. These water levels are calculated by taking the mean of all tide water levels of each calendar year. The minimum mean and maximum mean water levels differ approximately 0.004 m in Hoek van Holland and 0.009 m in Harlingen. These differences between the maxima and minima are rather small in comparison to the range of the semi-diurnal and spring-neap cycle. The annual maximum and minimum tide water levels are also calculated and examined via the same procedure. These results show a larger range in water level differences. For example, the

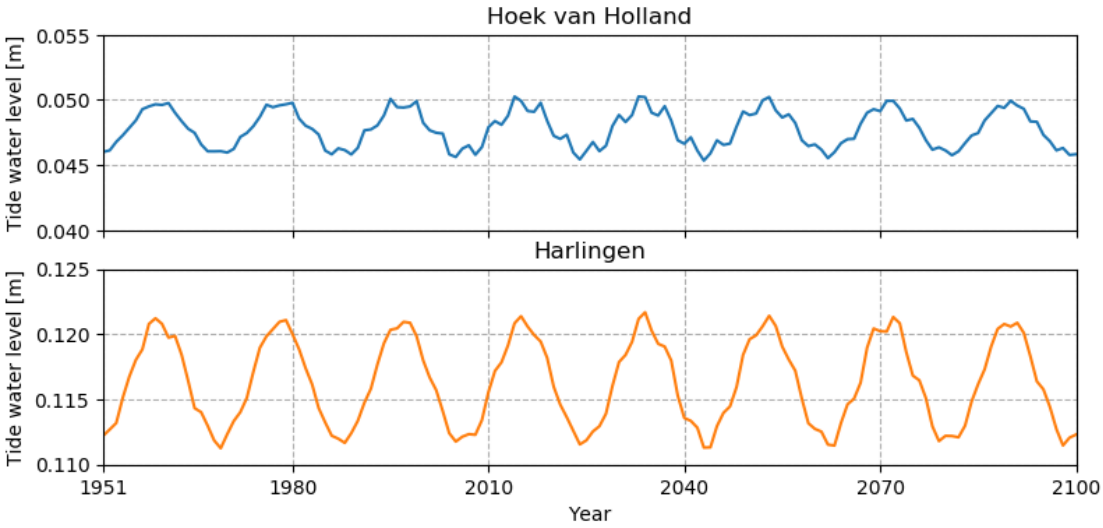

Figure 4-3. Annual mean tide water levels for Hoek van Holland and Harlingen of the period 1951-2100.

annual minimum and maximum tide water level both have a range of about 0.12 m in Hoek van Holland.

4.2.2. Surge and tide interaction

Surge heights are influenced by the tide. Caires (2011) states that surge occurring at the time of high tide tends to be damped, whereas surge on the rising tide typically is amplified. The latter is also mentioned by Horsburgh & Wilson (2007). This phenomenon is also found in the WAQUA data. Figure 4-4 depicts the WAQUA surge height from 36 hours before till 36 hours after an example of a storm surge event selected with the POT method in Hoek van Holland. The green line represents the surge height and clearly shows, apart from the multiple threshold exceedances, a peak every about six hours. Some of these peaks are synchronous with the rising tide (orange dashed line in the figure). Horsburgh & Wilson (2007) found that surge heights are significantly greater at low water levels than at high water levels. This could explain the peaks during low tide. Horsburgh & Wilson (2007) also states that the interaction between

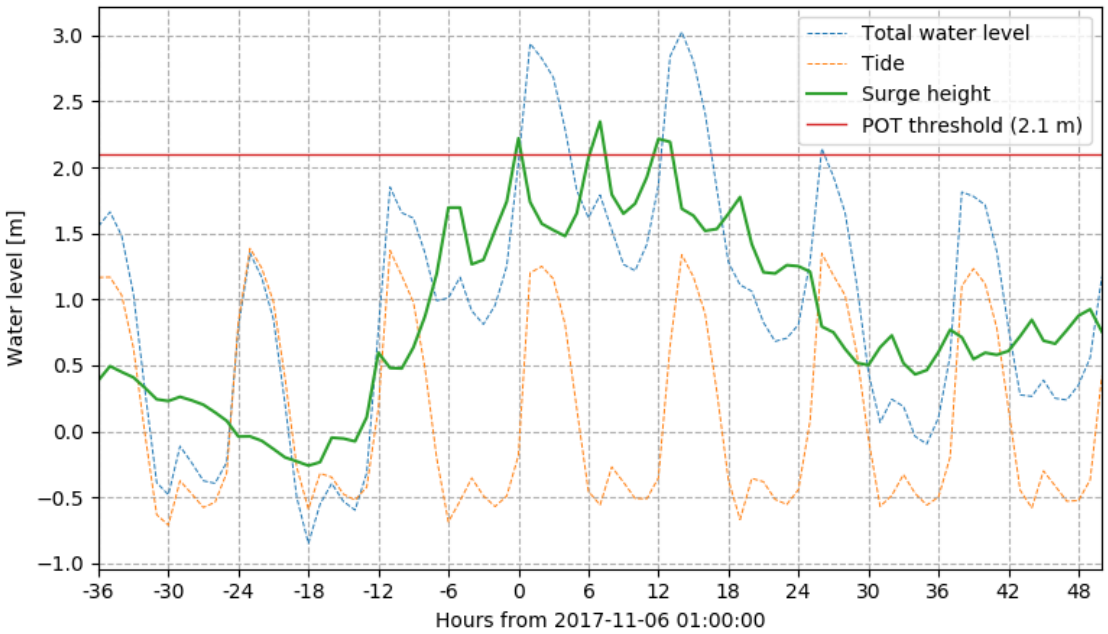

Figure 4-4. Visualisation of an example of a storm surge event at Hoek van Holland. The POT method is applied on the surge height (green).

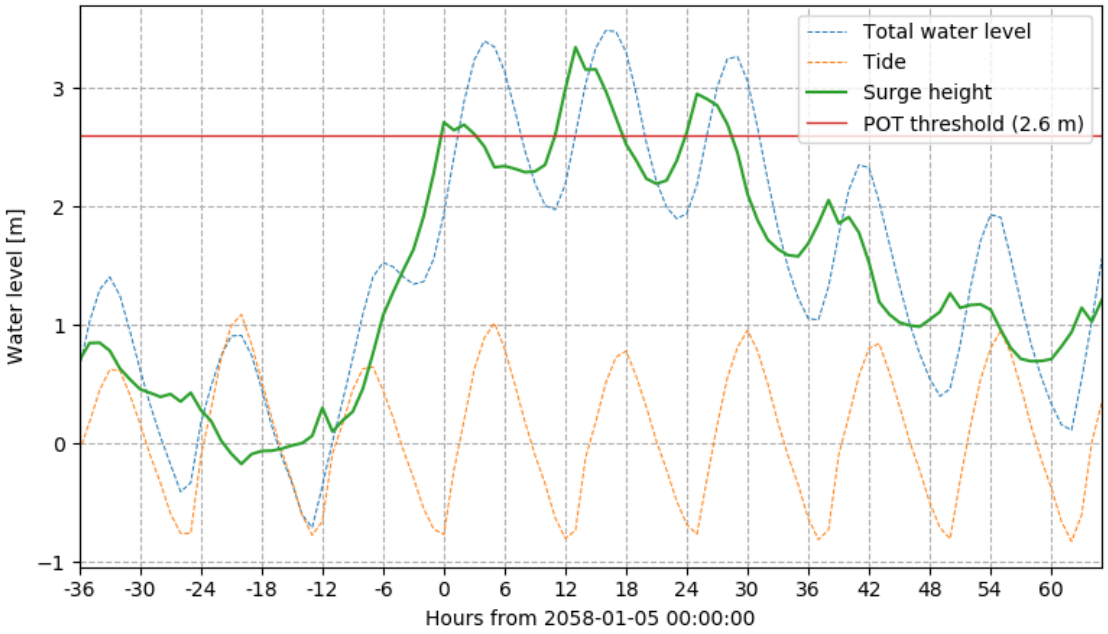

Figure 4-5. Visualisation of an example of a storm surge event at Harlingen. The POT method is applied on the surge height (green).

surge and tide is a dynamical effect, meaning that surge also affects tide water levels. This phenomenon is not found in the WAQUA data, because the tide water levels are the same for each of the 16 members.

Additionally, the time interval between successive threshold exceedances output from the POT method on the WAQUA surge height time series is used to confirm the influence of the tide on the surge heights. Figure 4-6 depicts the number of occurrences of time intervals between successive threshold exceedances in Hoek van Holland. Only time intervals between 0 and 72 hours are shown in this figure and a POT threshold of 1.00 meters is applied. The relatively low POT threshold is chosen for this analysis because it also clearly captures the influence of the tide on the surge heights, which is less the case for higher POT thresholds (Figure 3-3 and Figure 3-5). Observed are repeating peaks for about every half a semi-diurnal cycle period, i.e. approximately every 6.21 hours. This indicates that a large part of selected threshold exceedances contains peaks every about 6.21 hours or a multiply of this due to the influence of the tide. A similar phenomenon is found for the location of Harlingen (Figure 4-7), for which a POT threshold of 1.4 meters is applied. In Harlingen however, the peaks during rising tide are

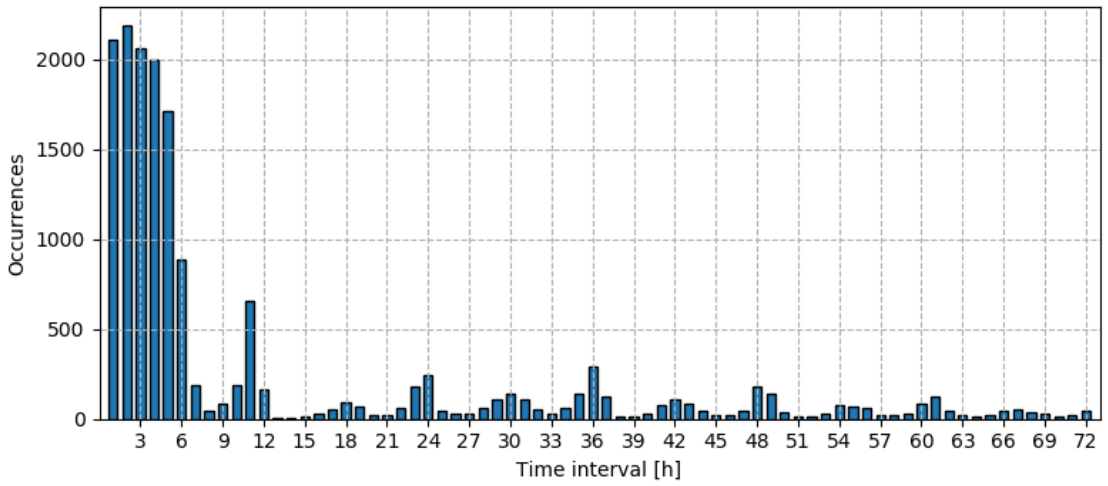

Figure 4-6. Occurrences of time intervals between successive threshold exceedances as a result of the POT method on the WAQUA surge height time series with a POT threshold of 1.0 m at Hoek van Holland.

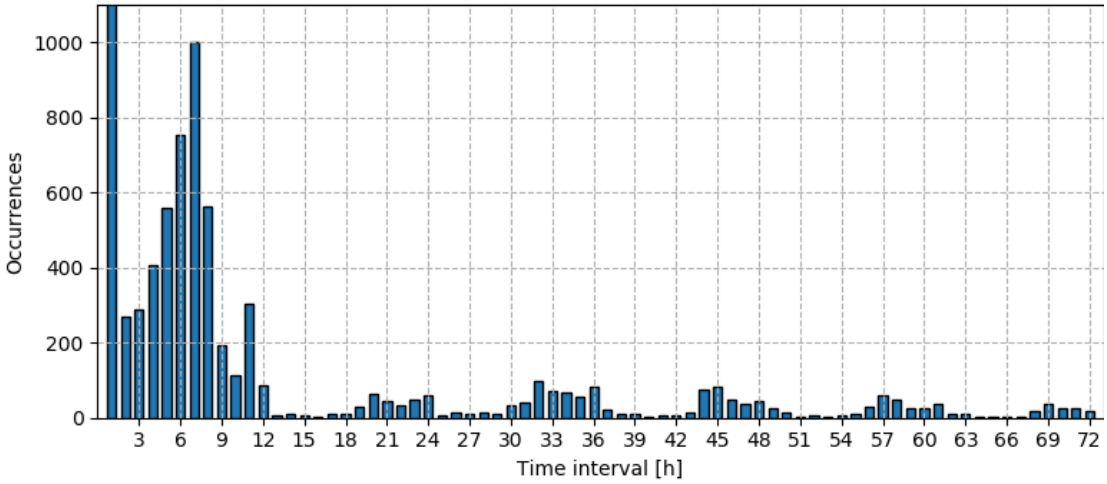

Figure 4-7. Occurrences of time intervals between successive threshold exceedances as a result of the POT method on the WAQUA surge height time series with a POT threshold of 1.4 m at Harlingen.

more dominant over the peaks during low tide. This is confirmed by the analysis on occurrences of time intervals between successive threshold exceedances, which shows repeating peaks for about every 12.41 hours (i.e. every semi-diurnal cycle period).

To summarize, the WAQUA data shows secondary peaks in Hoek van Holland during low tide and rising tide with comparable magnitude and shows secondary peaks in Harlingen mainly only during rising tide.

4.2.3. WAQUA data corrections for calculation of model accuracy

The results of the WAQUA data corrections for the calculation of the model accuracy are presented in this section. The first step is the calculation of the mean water level of the reference year/period. For this purpose, first the nodal tide cycle is analysed. The nodal signal for both locations is presented in Figure 4-3, in which the annual mean tide water levels are shown. Table 4-1 provides an overview of the correction values. Notable is that the correction due to the nodal tide signal is negligible (tide correction rows in the table), since the values are in the order of millimetres. However, the corrections are implemented because the correction values were already calculated which made the implementation practically effortless.

The second step was to account for the difference in reference water level between the measurement data based values (Table 2-3) and WAQUA model based values (mean water level rows in Table 4-1). This resulted in differences of +0.012 m for Hoek van Holland and -0.110 m for Harlingen. All WAQUA model output water levels were subtracted with these corrections to obtain water level data with the same reference as the measurement data based reference water level.

Table 4-1. Correction values for the model accuracy analysis resulting from the first step. Water level values are in meters and rounded to three decimals. Values represent annual means, unless specified differently (e.g. one value for all years is given). The calculation procedure was as follows (see also methodology section 4.1.2). First the annual mean total and tide water levels are calculated. Then the tide mean water level over 1951-2100 (entire data timespan) is calculated. The tide corrections are calculated by subtracting the annual mean tide water levels from the tide 1951-2100 mean. Subsequently, the corrected total water levels are calculated by adding the tide corrections to the annual mean total water levels. The mean water level is the mean of all corrected total water levels.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Hoek van Holland										
Total water level	0.083	0.073	0.067	0.079	0.070	0.077	0.083	0.080	0.081	0.088
Tide water level	0.046	0.047	0.046	0.046	0.048	0.048	0.048	0.049	0.050	0.050
Tide 1951-2100 mean	0.048									
Tide correction	0.002	0.001	0.002	0.001	0.000	-0.001	0.000	-0.001	-0.002	-0.002
Corrected total water level	0.084	0.074	0.069	0.081	0.070	0.076	0.083	0.079	0.079	0.085
Mean water level	0.078									
Harlingen										
Total water level	0.183	0.170	0.161	0.178	0.168	0.174	0.188	0.188	0.186	0.199
Tide water level	0.112	0.112	0.112	0.113	0.116	0.117	0.118	0.119	0.121	0.121
Tide 1951-2100 mean	0.116									
Tide correction	0.004	0.004	0.004	0.003	0.001	-0.001	-0.002	-0.003	-0.005	-0.005
Corrected total water level	0.187	0.174	0.165	0.181	0.169	0.173	0.186	0.186	0.181	0.194
Mean water level	0.180									

4.2.4. Model accuracy

Finally, the water level exceedance frequencies were calculated from the corrected WAQUA total water level data from section 4.2.3. Figure 4-8 shows the results of the two calculation methods. The figure shows four different results for each location (see figure legend). The threshold exceedances and storm surge events are the result of the first calculation method in which the rank number of a peak water level was divided by the total number of years considered. The result of threshold exceedances is for the higher exceedance frequencies till 0.5 per year and the result of storm surge events is for the lower exceedance frequencies from 0.5 per year, which is also done for the measurement data based exceedance frequencies (Dillingh, 2013). The GPD and GPD extrapolation are a result of the second calculation method for which Eq.(4-1) and Eq.(4-2) are used respectively. The shape (ξ) and scale (σ) parameters of the GPD fit are also shown in Figure 4-8. From the figure can be concluded that the GPD fits the resulting points of the first calculation method reasonably well. Therefore, the fitted GPD in combination with the threshold exceedances result are used to construct the final relation of water levels with exceedance frequencies of the WAQUA data.

Figure 4-9 shows the final result of the WAQUA data in combination with the measurement data based values. The measurement data based lines in the figure are constructed from the rounded values of Table 2-2, which is why these lines are not as smooth as the lines calculated from the WAQUA data. Note that the x-axis of the figure is inverted. Striking are the systematically lower values from WAQUA in comparison to the values based on measurement data. These differences are smaller for higher exceedance frequencies (i.e. left part of the graph) than for lower exceedance frequencies (i.e. right part of the graph). Comparing the two locations, the difference between the model and measurement data based water levels at Hoek van Holland is smaller than at Harlingen for higher exceedance frequencies, whereas this difference at Hoek van Holland is larger than at Harlingen for lower exceedance frequencies

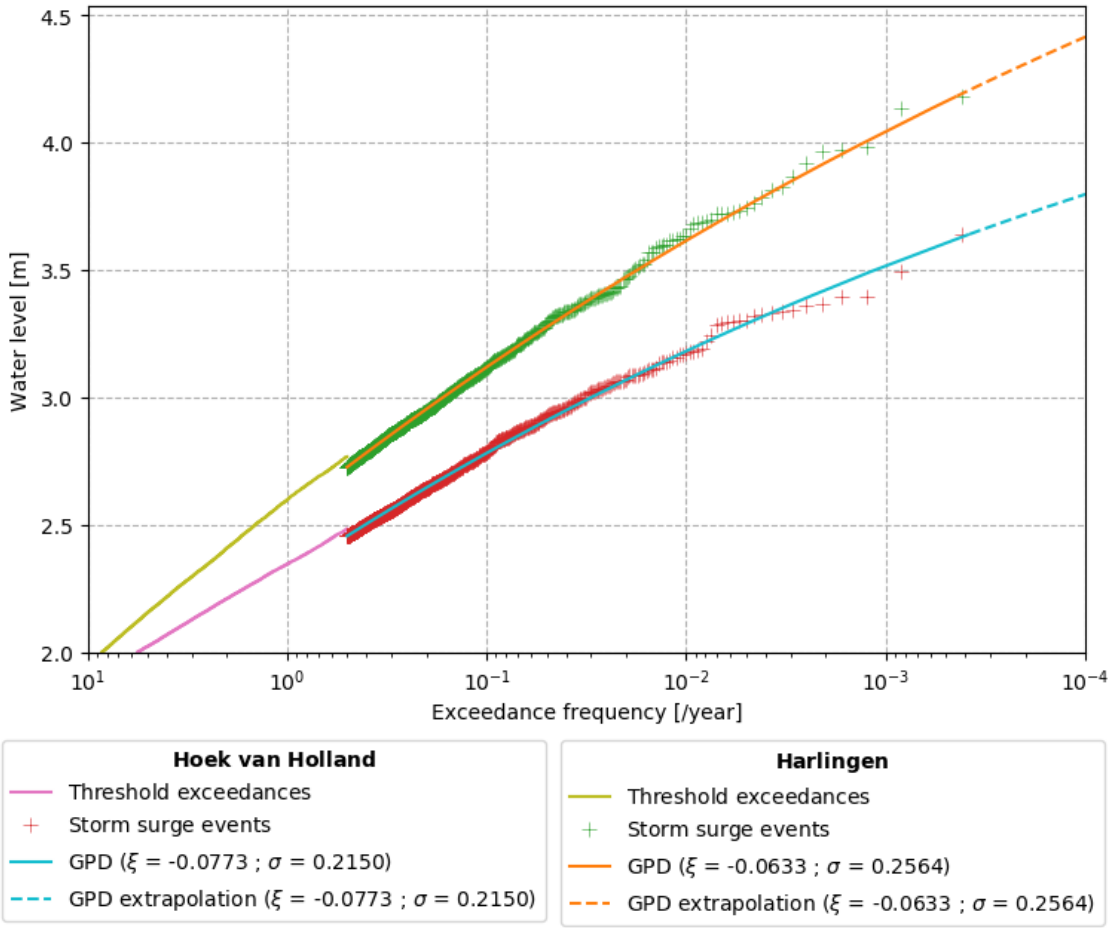

Figure 4-8. Water level exceedance frequency results of the different calculation methods. Exceedance frequencies till 0.5 per year are calculated from the threshold exceedance selected total water level peaks and from 0.5 till 10^{-4} per year are calculated from the storm surge event selected total water level peaks. For the latter, the GPD is fitted and extrapolated.

(see also lower panel of Figure 4-9). For example, from the water level of 2 meters at Hoek van Holland the difference is gradually increasing until around the exceedance frequency of 10^{-2} (1/100) where it diverges from the measurement data based line. For Harlingen, this 'diverging' phenomenon is also observed, however less predominant than for Hoek van Holland. The GPD fit in combination with the ML estimation of the GPD parameters results in a water level difference relative to the measurement data based water level of approximately -11% to -13% for Harlingen and up to -25% for Hoek van Holland. Important to mention is that confidence in these results decreases with a decrease in exceedance frequency. This is because the point (value) density is much lower for the lower exceedance frequencies. This results in that the estimated model lines in Figure 4-9 are much more likely to be accurate for the higher exceedance frequencies than for the lower exceedance frequencies.

Conclusive, from the lower panel of Figure 4-9 can be observed that the model underestimates the water level. This underestimation increases with a decreasing exceedance frequency. This especially holds true for Hoek van Holland.

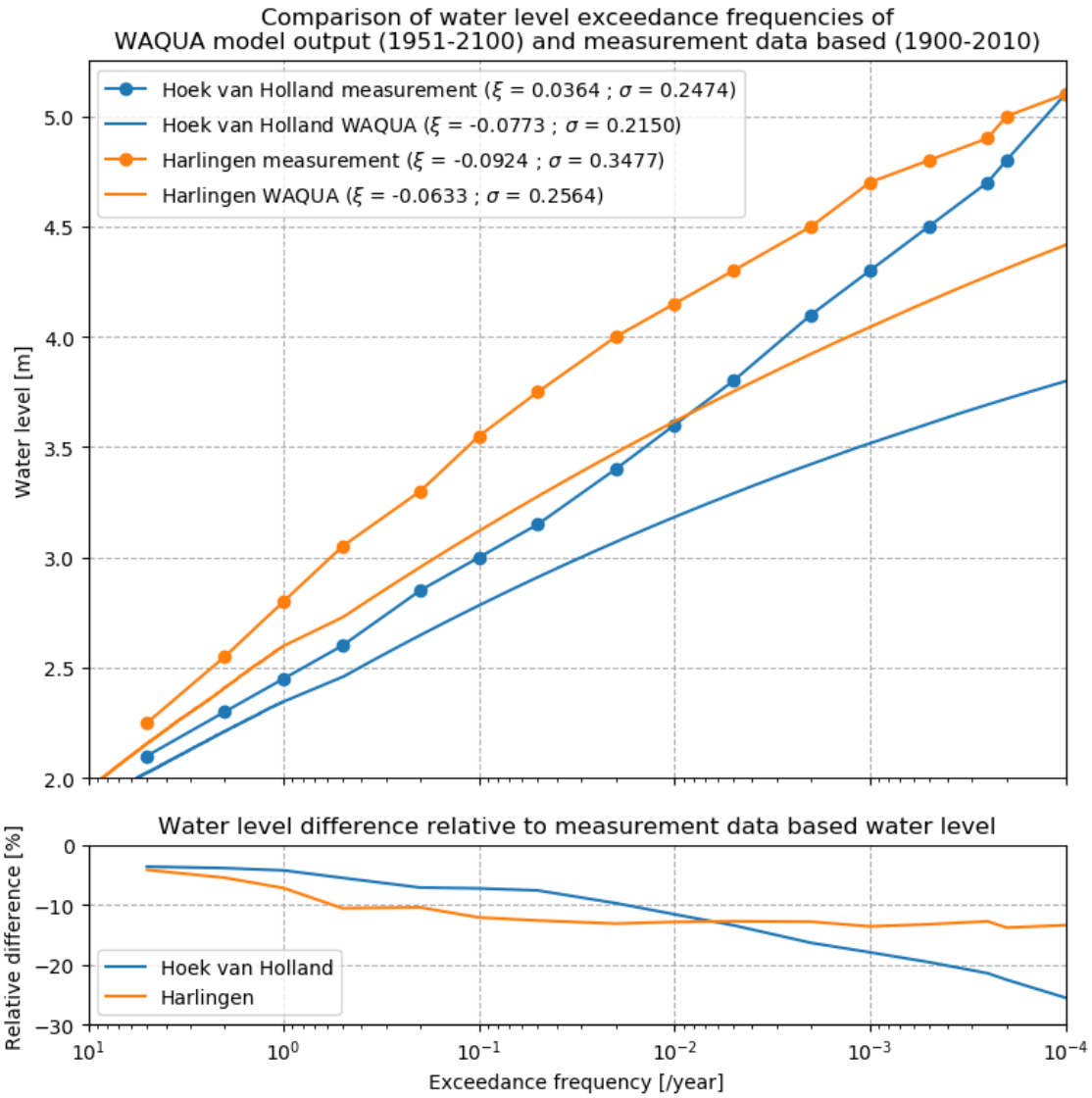

Figure 4-9. Water level exceedance frequencies based on measurement data (Dillingh, 2013) and based on WAQUA model output for Hoek van Holland and Harlingen (top figure) and the differences between them relative to the measurement data based water levels (bottom figure). The dots are the exceedance frequencies from Table 2-2. Values are applicable to the beginning of the year 2011.

5. Storm surge trends from 1951 to 2100

Three criteria are used to assess storm surge trends from 1951 to 2100, namely: frequency, intensity and duration of storm surge. The block extremes method, but mainly the peak over threshold (POT) method are used to perform this assessment. For the POT method, the definitions of storm surge event and storm surge cluster event are used (chapter 3). Trends in storm surge cluster events are only assessed for the frequency criterion. All criteria are analysed over overlapping time periods of 30 years.

5.1. Methodology

For the trend analysis is chosen to examine three criteria: frequency, intensity and duration of storm surge. Figure 5-1 provides an overview of how the methods led to the trend results. Two methods are used to analyse the trend assessment criteria over the period 1951-2100 (referred to as the entire data timespan) from the WAQUA model data, namely the peak over threshold (POT) method and the block extremes method. The POT method is extensively described in chapter 3 and makes use of the definitions of a storm surge event and storm surge cluster event. The block extremes method is added to this trend assessment, since this a relatively simple method and is used to make comparisons with the results of the POT method. The block extremes method only assesses the surge intensity and is not of influence on the definition of a storm surge event. For storm surge cluster events only the frequency of events is examined, because there is no clear definition of the intensity and duration of this type of event.

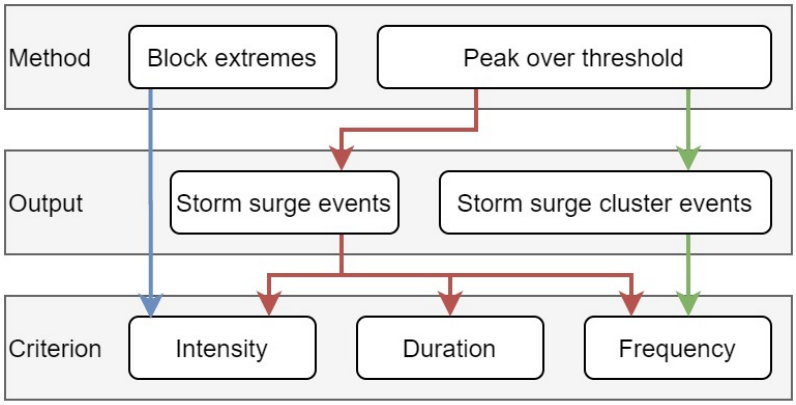

Figure 5-1. Overview of methods leading to trend results. Shown are the two methods used, the resulting output of these methods and the criteria which they address. Each arrow colour represents an assessment.

To obtain information about trends the WAQUA data period of 1951-2100 is divided in periods of 30 years, which overlap with 15 or 29 years depending on the criterion assessment. This overlapping (or running mean) methodology essentially functions as a smoothing algorithm. Overlapping periods are used, because this increases the total number of periods on which statistical tests can be carried out. In other words, overlapping results in more data points. Periods of 30 years are chosen because smaller periods (e.g. 20 years) lead to smaller sample sizes, and larger periods (e.g. 50 years) lead to less accurate trend assessments. Additionally, a period of 30 years is in climate research considered as the correct period length for the weather variations to be averaged out. These 30 year periods change forward in time with increments of 1 year (i.e. 1951-1980, 1952-1981, 1953-1982, etc.). For the criterion frequency of storm surge (cluster) events all these 30 year periods are shown relative to the mean of the entire data timespan, because this clearly shows the change over the entire data timespan and in this way the results of different POT thresholds and locations can easily be compared. For the intensity and duration criteria increments of 15 years are used (i.e. 1951-1980, 1966-1995, 1981-2010, etc.), because these results contain less “noise” than the frequency of events criterion.

5.1.1. Block extremes method

The block extremes method is only suitable to examine the intensity criterion, because the calculation only results in extreme surge heights (or total water levels). In this method the extreme values of surge height in the data are obtained by taking the maximum or maxima within time intervals. Examples of time intervals are daily, monthly or annual. Since the WAQUA data consists of 16 members, annual maxima result in a large enough sample size to examine the presence of trends in the criterion intensity. Therefore, the annual maxima of each member within the period 1951-2100 are calculated. The results are quantified and visualised in the form of boxplots.

5.1.2. Peak over threshold method

The peak over threshold (POT) method as implemented in this research is described in chapter 3. The POT threshold stability analysis (section 3.1.5) indicated that POT thresholds should be in the range 1.5 - 2.3 meters for Hoek van Holland and 1.9 - 2.7 meters for Harlingen. Within these ranges two POT thresholds are chosen for each location to be shown in the results section. For this purpose, sample sizes have to be large enough to be able to draw conclusions. Additionally, the POT thresholds are chosen such that sample sizes for both locations are approximately equal so that valid comparisons can be made. This resulted in the applied POT thresholds of 1.7 and 2.1 meters for Hoek van Holland and 2.1 and 2.6 meters for Harlingen.

Furthermore, storm surge cluster events are divided in events with a minimal number of 2, 3 and 4 storm surge events per cluster. This is done to directly address the sensitivity of the definition of storm surge cluster events. For storm surge cluster events the chosen POT thresholds are much smaller than for the single storm surge events due to the fact that storm surge cluster events are much more rare than single storm surge events. For these events the POT thresholds are chosen such that the resulting sample sizes for each used definition of a storm surge cluster event are approximately 100 events. This is approximately equal to an exceedance frequency of 0.2 per year.

5.1.3. Statistical significance of trends

For both the POT method and the block extremes method, linear trends are fitted to the various boxplot variables (e.g. medians, 75th percentiles). This is illustrated in Figure 5-2. Linear trends are also fitted to the frequency criterion results. This method of searching for significant trends is adopted, because it is most commonly used, it is computationally the most simple method, and there are visually no quadratic, exponential, or other patterns identified in the resulting data from the block extremes and POT method. Linear trend fitting is also known as a linear regression analysis, in which one linear function is used to search for a monotonic increasing or decreasing tendency in the data. The trends are fitted by the least squares procedure. From this analysis the following coefficients are derived: slope, intercept, correlation, p-value and standard error. If the linear trend line is described by the function $y = ax + b$, the coefficient a is the slope and b is the intercept. The correlation coefficient (r) describes the strength of the linear relationship between the linear trend line and the points to which this line is fitted. The correlation value is always between -1 and 1, where -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship and 1 indicates a perfect positive linear relationship. The p-value describes the statistical significance of the obtained trend lines and is calculated using a Wald Test with t-distribution of the test statistic. The p-value in this analysis is defined as the probability, assuming the null hypothesis is true, of the actual slope of the trend line being at least as extreme as the obtained slope value. The null hypothesis is set to be a slope of zero $(a = 0)$. The alternative hypothesis is then set to be a slope of not equal to zero $(a \neq 0)$. These hypotheses check whether a significant increasing or decreasing trend is found or if no significant trend is found. The alternative hypothesis states that the slope is positive or negative, indicating that the significance test is a two-tailed test. The p-value is thus also for both tails together. In this research, the significance level alpha (α) is set to be 0.05, which is commonly used in scientific studies. Therefore, an obtained trend is considered to be statistically significant if the corresponding p-value is smaller than the significance level α = 0.05. Important to mention is that the linear fitting procedure is done on overlapping 30 year periods, which is likely to result in lower p-values. For the interpretation of the statistical significance this is important to take into account. Finally, the standard error coefficient is that of the linear trend line. All five coefficients result from the linear regression function of the SciPy (Scientific Python) module in Python (The Scipy community, 2014).

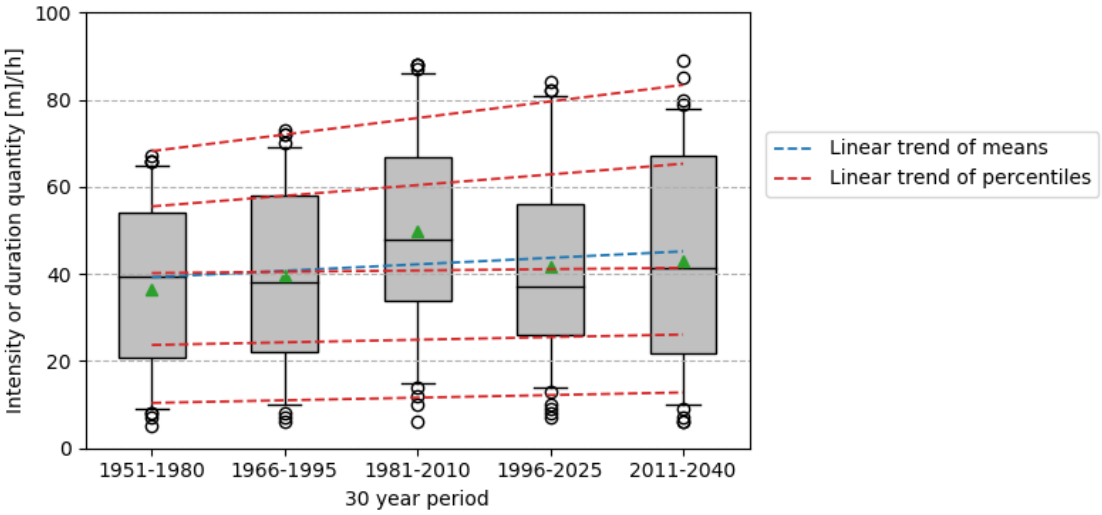

Figure 5-2. Illustration of linear trend fitting to the boxplot variables of synthetic data. The shown boxplot variables are the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), extremes (circles).

5.2. Results

This results section is structured per storm surge trend criterion in the order: frequency, intensity and duration of storm surge. The frequency of storm surge events criterion comprises of frequency of single storm surge events and frequency of storm surge cluster events. The intensity criterion is divided in the surge height intensity and total water level intensity. The assessment of surge height intensity is divided in the results of the block extremes method and POT method. For each criterion assessment the results are described in the order of Hoek van Holland, Harlingen, and a comparison between the two locations. Figure 5-1 provides a general overview of how the trends assessment is structured.

An explanation of the linear trend fitting coefficients can be found in the methodology in section 5.1.3. The resulting coefficients are rounded with the precision depending on the criteria and coefficient. The obtained p-values in the tables are marked green if the value is smaller than the significance level alpha (α = 0.05). Hence, green means the obtained trend line can be considered statistically significant.

5.2.1. Trends in frequency of storm surge events (1951-2100)

Table 5-1 shows the resulting mean numbers of storm surge events per 30 years. The member means in the table give an indication of the exceedance frequencies resulting from the applied POT thresholds. For example, for the POT threshold of 1.7 meters at Hoek van Holland the member mean is 15.6 storm surge events per 30 years. This results in 0.52 storm surge events per year, which is approximately the exceedance frequency of a surge height of 1.7 meters at Hoek van Holland. The sum of all members in this table give an indication of the sample sizes on which the assessment of the intensity and duration[1] criteria (section 5.2.3 and 5.2.4 respectively) are based on.

Table 5-1. Resulting mean frequencies of storm surge events for Hoek van Holland and Harlingen for each POT threshold. The values are numbers of events per 30 years averaged over the entire data timespan. Member means are calculated by dividing by number of members (i.e. 16). The estimated mean exceedance frequency is the member mean divided by the length of one period (i.e. 30 years).

	Hoek van Holland		Harlingen	
POT threshold [m]	1.7	2.1	2.1	2.6
Sum of all members	249.5	47.0	249.8	45.6
Member mean	15.6	2.9	15.6	2.8
Estimated mean exceedance frequency [/y]	0.5	0.1	0.5	0.1

Figure 5-3 shows the result of the POT method for storm surge event frequency for Hoek van Holland. Herein, standardised numbers of storm surge events are presented (i.e. relative to the entire data timespan mean). The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-2. The obtained trend line of the 1.7 meter POT threshold is found to be significant, whereas the trend line of the 2.1 meter POT threshold is not. The p-value of the 1.7 meter POT threshold trend line is negligibly larger than zero, which indicates that the null hypothesis of a slope of zero can be rejected for this obtained trend line. The slope of the significant trend is -0.02 per year, which results in approximately 15% less storm surge events at the end of the 21st century in comparison to 1951-1980. The separate member results are also shown in the figure. From these results also a clear decrease for the 1.7 meter POT threshold can be observed, whereas the 2.1 meter POT threshold separate member results show merely large variation. Of the latter, a few of the members contain zero storm surge events for some of the 30 year periods.

Figure 5-4 shows the result of the POT method for storm surge event frequency for Harlingen. Herein, also standardised numbers of storm surge events are presented (i.e. relative to the

[1] The sample sizes for the duration criterion are in some occasions 0 to 1% smaller than the original sample sizes depending on the POT threshold (See section 6.1.1).

entire data timespan mean). The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-2. For the POT threshold of 2.6 meters, the obtained trend was found to be significant. Two separate member results of this POT threshold show large variations to the positive side which is why the y-axis scale for this POT threshold in Figure 5-4 is much larger than for the 2.1 meter POT threshold. The 2.1 meter POT threshold shows a peak in frequency at around 1996-2025. The slope of the significant trend is relatively small, namely +0.0014 events per year.

Comparing the two locations, the most striking difference is the sign of the trend line slopes. Hoek van Holland resulted in only negative slopes, whereas Harlingen resulted in only positive slopes. When observing the two significant trends, the magnitude of the slope for Hoek van Holland is much greater than that of Harlingen. Moreover, both results for Harlingen contain larger variance when comparing with those of Hoek van Holland.

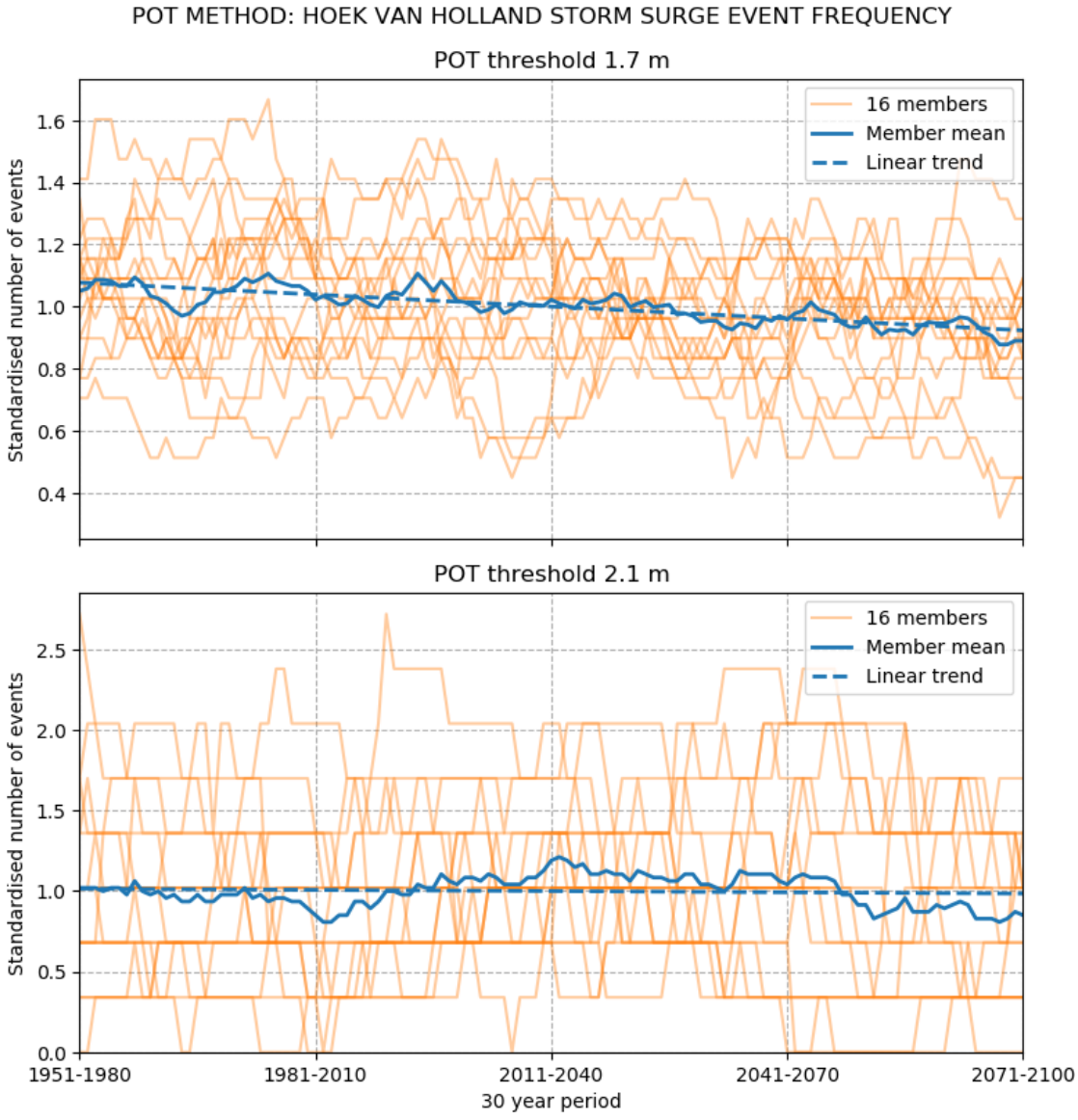

Figure 5-3. Frequency of storm surge events for Hoek van Holland. The frequency is shown for 30 year periods, which sequentially overlap by 29 years (also known as smoothing algorithm or running mean). Shown are numbers of storm surge events relative to the mean of the member mean over the entire data timespan for each POT threshold. The linear trend line is fitted through the member mean for each POT threshold. The scale of the y-axis is different for the two panels.

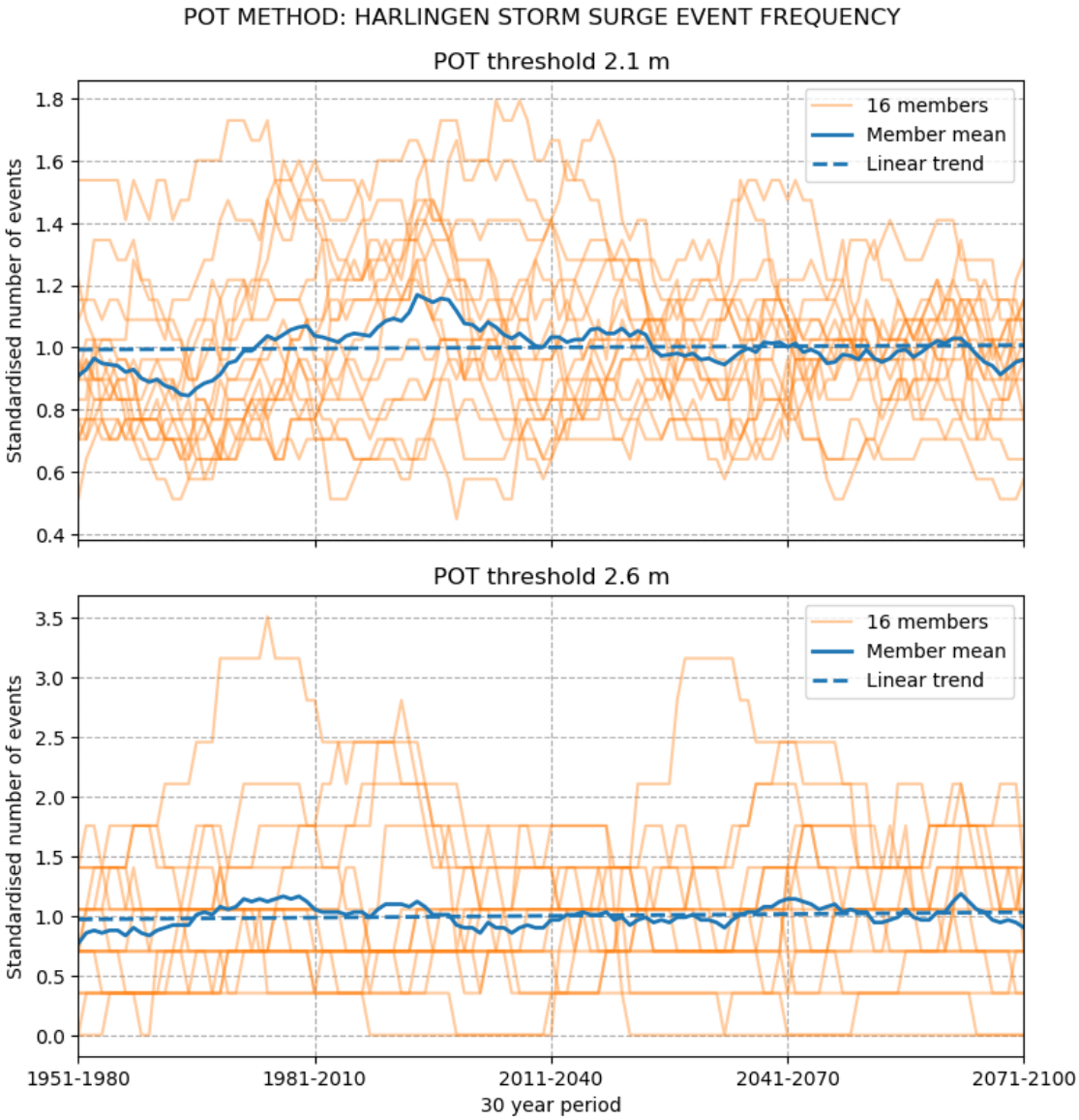

Figure 5-4. Frequency of storm surge events for Harlingen. The frequency is shown for 30 year periods, which sequentially overlap by 29 years (also known as smoothing algorithm or running mean). Shown are numbers of storm surge events relative to the mean of the member mean over the entire data timespan for each POT threshold. The linear trend line is fitted through the member mean for each POT threshold. The scale of the y-axis is different for the two panels.

Table 5-2. Linear trend fitting coefficients of the frequency of storm surge events for Hoek van Holland and Harlingen (corresponding to Figure 5-3 and Figure 5-4).

POT METHOD: FREQUENCY OF STORM SURGE EVENTS					
	Slope (a) [/y]	Intercept (b) [/30y]	Correlation (r)	p-value	Standard error [/30y]
Hoek van Holland					
POT threshold 1.7 m	-0.0200	16.8	-0.81	$7.91 * 10^{-32}$	0.0012
POT threshold 2.1 m	-0.0007	3.0	-0.09	0.317	0.0007
Harlingen					
POT threshold 2.1 m	0.0019	15.5	0.07	0.471	0.0027
POT threshold 2.6 m	0.0014	2.8	0.20	0.026	0.0006

5.2.2. Trends in frequency of storm surge cluster events (1951-2100)

Table 5-3 shows the resulting mean numbers of storm surge cluster events per 30 years. The member means in the table give an indication of the exceedance frequencies resulting from the applied POT thresholds, for which the values are also included in the table.

Table 5-3. Resulting mean frequencies of storm surge cluster events for Hoek van Holland and Harlingen for the corresponding POT thresholds. The values are numbers of events per 30 years averaged over the entire data timespan. Member means are calculated by dividing by number of members (i.e. 16). The estimated mean exceedance frequency is the member mean divided by the length of one period (i.e. 30 years).

	Hoek van Holland			Harlingen		
POT threshold [m]	0.8	1.0	1.3	1.0	1.2	1.6
Minimal number of storm surge events	4	3	2	4	3	2
Sum of all members	107.3	110.8	102.5	89.8	112.1	99.4
Member mean	6.7	6.9	6.4	5.6	7.0	6.2
Estimated mean exceedance freq. [/y]	0.2	0.2	0.2	0.2	0.2	0.2

Figure 5-5 shows the results of the POT method for storm surge cluster event frequency for Hoek van Holland and Harlingen. Herein, standardised numbers of storm surge events are presented (i.e. relative to the entire data timespan mean). The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-4. Except for one of the obtained trend lines, all are found to be significant with p-values that are much smaller than the significance level of 0.05. Even though most correlation coefficients are relatively small. For Hoek van Holland only the trend lines with a negative slope are significant. The difference between the three shown storm surge cluster event definitions is remarkable, because there is no clear agreement between the three solid lines for Hoek van Holland in Figure 5-5. The only clear agreement which can be observed is the decreasing tendency from 2041-2070 towards the end of the 21st century. From around the period 2041-2070 all three definitions show a clear decrease towards about 20% less events than in the first 30 year period. The fact that there is not much agreement shows the sensitivity of the condition of the minimum number of storm surge events within a storm surge cluster event. Nevertheless, from the significant trends slope signs can be concluded that a decrease in the frequency of storm surge cluster events in Hoek van Holland is more likely than an increase.

The obtained trends for Harlingen are all significant but are not in agreement concerning the slope directions. Two of the trends have a positive slope and one has a negative slope. This observation again emphasises the sensitivity of the used definition. Two of the three definitions show a clear peak around the period 2021-2050. After this period the frequency has mainly a tendency to decrease till the end of the 21st century.

Both locations show that the results are largely dependent on the minimum number of events in a storm surge cluster event. Furthermore, the overall results of all three definitions for both locations show a decreasing tendency from the period around 2021-2050 till the final 30 year period. For Hoek van Holland this observation is comparable with the results of single storm surge events in section 5.2.1. The variation over the entire data timespan is for both locations of similar magnitude.

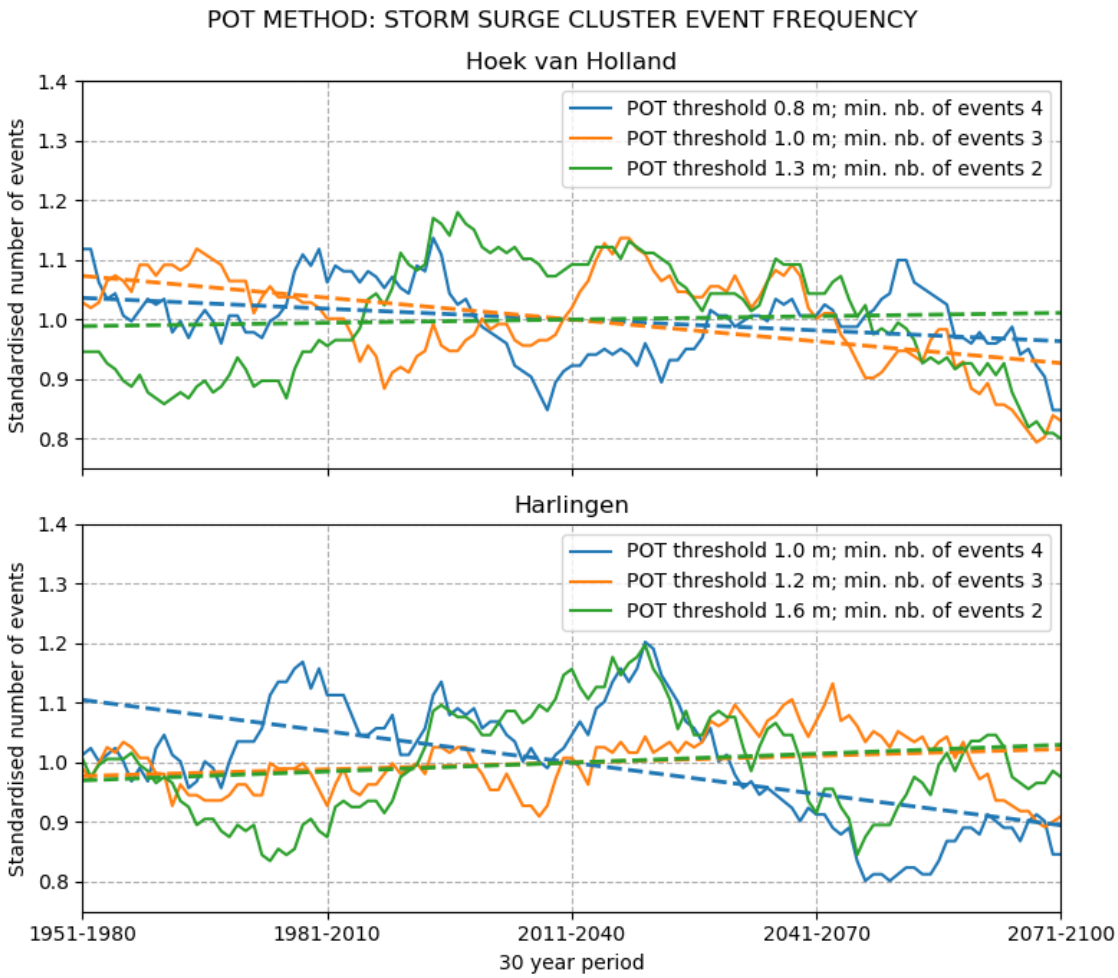

Figure 5-5. Frequency of storm surge cluster events for Hoek van Holland and Harlingen. The frequency is shown for 30 year periods, which sequentially overlap by 29 years (also known as smoothing algorithm or running mean). Shown are numbers of storm surge cluster events relative to the mean of the member mean over the entire data timespan for each storm surge cluster event definition. The linear trend lines (dashed lines) are fitted through the member mean for each storm surge cluster event definition.

Table 5-4. Linear trend fitting coefficients of the frequency of storm surge cluster events for Hoek van Holland and Harlingen (corresponding to Figure 5-5).

POT METHOD: FREQUENCY OF STORM SURGE CLUSTER EVENTS					
	Slope (a) [/y]	Intercept (b) [/30y]	Correlation (r)	p-value	Standard error [/30y]
Hoek van Holland					
POT threshold 0.8 m	-0.0041	7.0	-0.35	0.0001	0.0010
POT threshold 1.0 m	-0.0084	7.4	-0.54	$2.37 * 10^{-10}$	0.0012
POT threshold 1.3 m	0.0012	6.3	0.07	0.466	0.0016
Harlingen					
POT threshold 1.0 m	-0.0098	6.2	-0.61	$1.34 * 10^{-13}$	0.0012
POT threshold 1.2 m	0.0026	6.8	0.26	0.004	0.0009
POT threshold 1.6 m	0.0031	6.0	0.20	0.028	0.0014

5.2.3. Trends in storm surge intensity (1951-2100)

The intensity criterion is divided in the surge height intensity and total water level intensity. The assessment of surge height intensity is divided in the results of the block extremes method and POT method. The POT method resulted in four figures and four corresponding tables of which two provide information about surge height intensity and two about total water level intensity of storm surge events.

Figure 5-6 shows the result of the block extremes method for surge height intensity for both locations. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-5. The only significant trend is that of the 25th percentile of Harlingen. The p-value of this trend line of 0.011 is well below the significance level of 0.05. However, all other fitted trend lines are not significant. The trend lines of Hoek van Holland show mainly negative slopes, whereas those of Harlingen show both negative and positive slopes. The latter observation suggests that it is unlikely that there is a general significant trend at Harlingen. Moreover, the annual maxima at Hoek van Holland are within the range 0.7 till 2.9 meter (with the exception of one value) and the annual maxima at Harlingen are within the range 1.0 till 3.8 meter. This indicates that the variance at Harlingen is larger than at Hoek van Holland.

Figure 5-7 shows the result of the POT method for surge height intensity of storm surge events at Hoek van Holland for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-6. None of the p-values of the obtained trends is close to the significance level of 0.05, hence there are no significant trends found. There is one extreme value that stands out (3.44 m), which is in the first 30 year period. This value is more than half a meter higher than the second highest surge height value.

Figure 5-8 shows the result of the POT method for surge height intensity of storm surge events at Harlingen for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-7. The 95th percentiles linear trend fit for the 2.6 meter POT threshold resulted in the only significant trend. The p-value and slope of this trend line are 0.001 and 0.0517 respectively, which indicate there is strong evidence for an increasing trend for this percentile. However, the trend lines of the median, 25th and 5th percentiles (i.e. lower percentiles) have negative slopes, whereas the 75th and 95th percentile trend lines have positive slopes. This suggests that the variance of peak surge heights of storm surge events increases towards the end of the 21st century, albeit with negligible magnitude. In contrary to this observation, the lower POT threshold of Harlingen resulted in only increasing trend lines. The overall observation of these results of Harlingen is that the most extreme surge heights of 30 year periods increases from approximately 3.0 – 3.2 meters to approximately 3.4 – 3.7 meters in the period 1951-2100, but that all selected surge heights above 2.1 and 2.6 meters does not increase on average.

The obtained trends for Hoek van Holland of both methods (block extremes and POT) show similar results, that is no significant trends are found. For Harlingen however, the 95th percentile trend fits of each method show a clear difference. For the block extremes method, the 95th percentile fit is far from a significant trend with a p-value of 0.653, whereas the p-value of 0.001 for the POT threshold of 2.6 meters is clearly significant. This again supports the observation that only the most extreme surge heights (i.e. approximately above 3.0 meters) increase. Comparing the surge height intensity of both locations, a relatively large difference in the magnitude of the extremes (circles) is observed. At Hoek van Holland the most extreme values are around 2.8 meters (except the 3.44 meter value), whereas at Harlingen these values are around 3.5 meters. Also, the 2.6 meter POT threshold at Harlingen resulted in relatively large variation over the entire data period in comparison to the higher POT threshold result of Hoek van Holland, considering both results have a similar sample size.

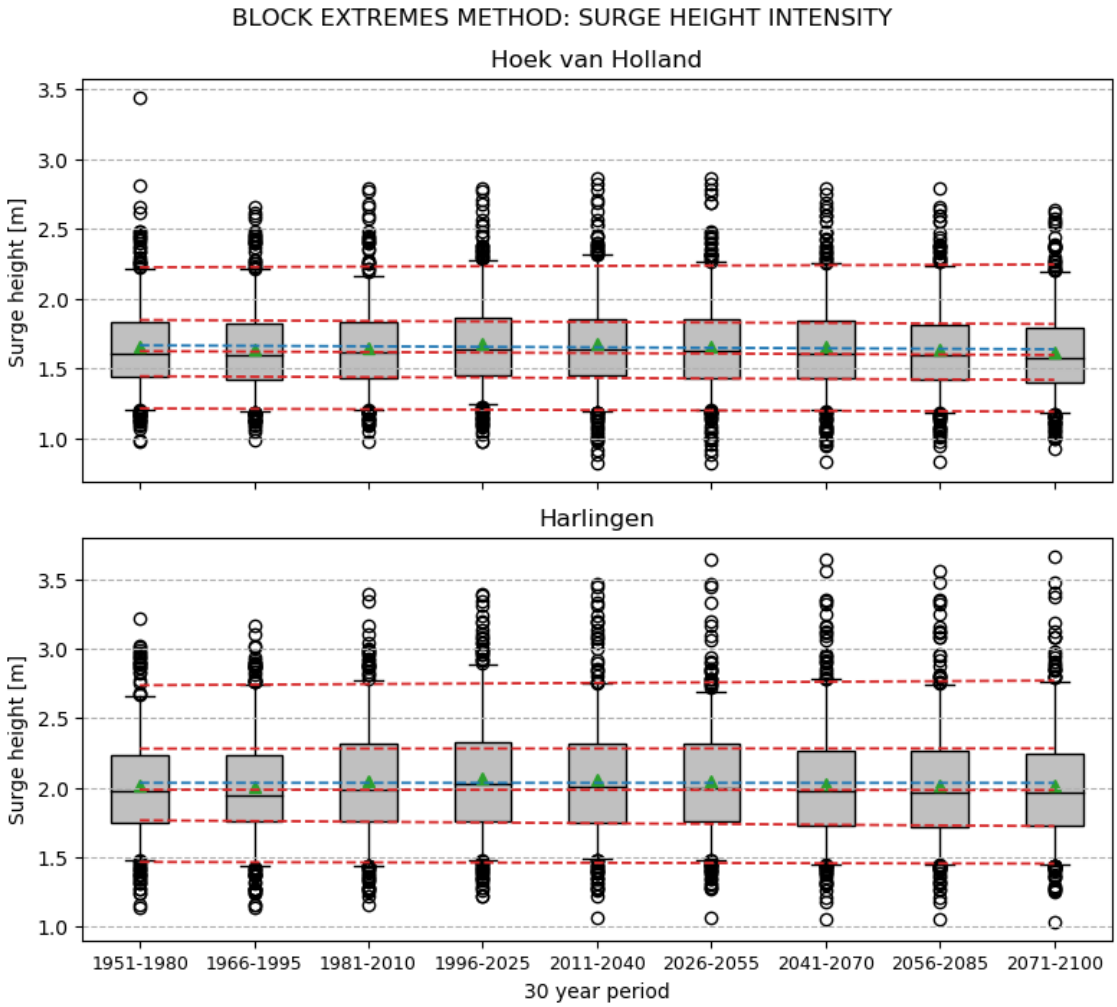

Figure 5-6. Surge height intensity resulting from the block extremes method for Hoek van Holland and Harlingen. The boxplots are created from the annual maximum surge heights of each member. The boxplots are for overlapping periods of 30 years from 1951 till 2100 and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-5. Linear trend fitting coefficients of the surge height intensity boxplots resulting from the block extremes method for Hoek van Holland and Harlingen (corresponding to Figure 5-6).

BLOCK EXTREMES METHOD: SURGE HEIGHT INTENSITY					
	Slope (a) [m/15y]	Intercept (b) [m]	Correlation (r)	p-value	Standard error [m]
Hoek van Holland					
Mean	-0.0035	1.67	-0.49	0.184	0.0023
95th percentile	0.0026	2.22	0.15	0.703	0.0064
75th percentile	-0.0037	1.85	-0.41	0.277	0.0031
Median	-0.0034	1.62	-0.47	0.202	0.0024
25th percentile	-0.0033	1.44	-0.59	0.092	0.0017
5th percentile	-0.0029	1.21	-0.38	0.316	0.0027
Harlingen					
Mean	-0.0001	2.03	-0.01	0.982	0.0028
95th percentile	0.0042	2.74	0.17	0.653	0.0089
75th percentile	0.0002	2.28	0.02	0.964	0.0052
Median	-0.0004	1.98	-0.04	0.916	0.0036
25th percentile	-0.0052	1.76	-0.79	0.011	0.0015
5th percentile	-0.0015	1.46	-0.20	0.609	0.0028

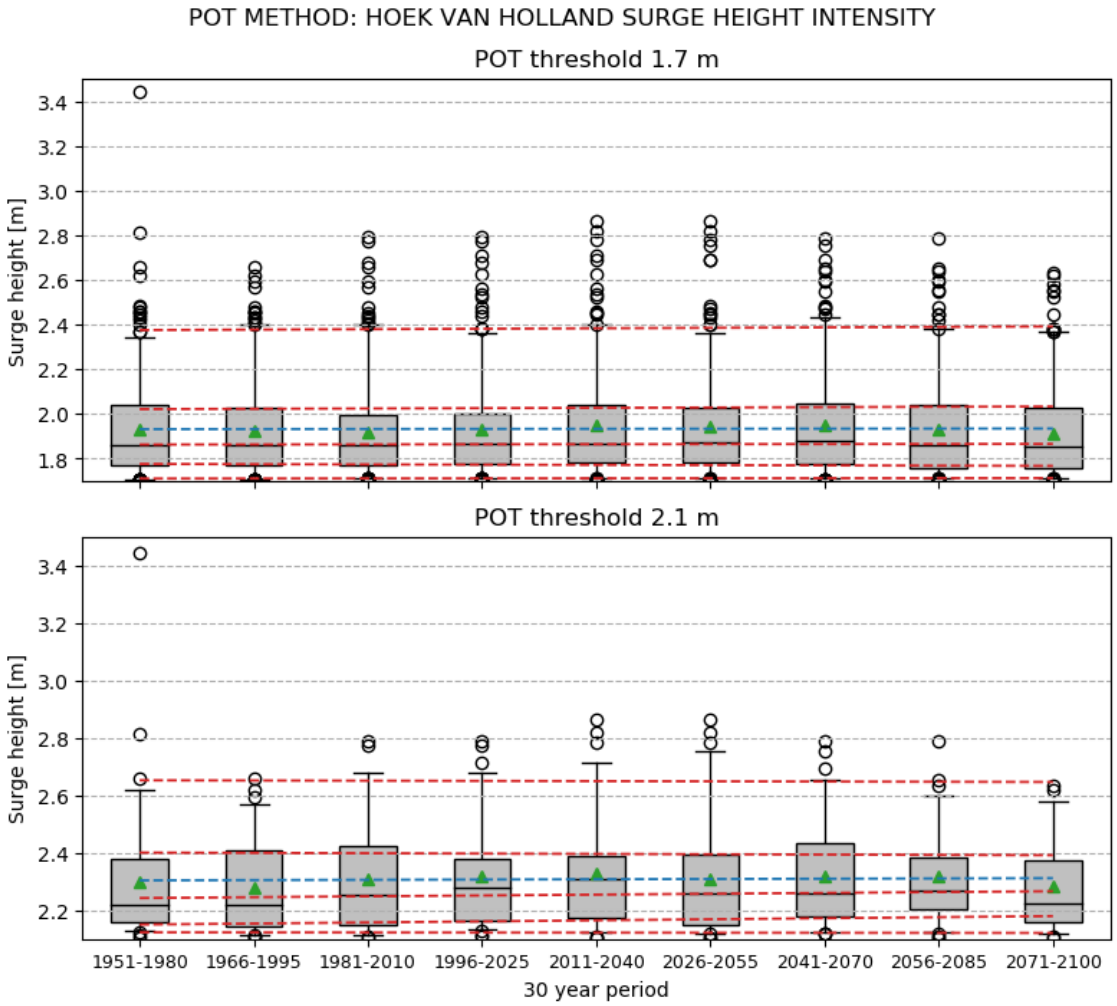

Figure 5-7. Surge height intensity resulting from the POT method for Hoek van Holland. The boxplots are created from the peak surge heights of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-6. Linear trend fitting coefficients of the surge height intensity boxplots resulting from the POT method for Hoek van Holland (corresponding to Figure 5-7).

POT METHOD: HOEK VAN HOLLAND SURGE HEIGHT INTENSITY					
	Slope (a) [m/15y]	Intercept (b) [m]	Correlation (r)	p-value	Standard error [m]
POT threshold 1.7 m					
Mean	0.0004	1.93	0.08	0.842	0.0017
95th percentile	0.0020	2.38	0.19	0.619	0.0039
75th percentile	0.0015	2.02	0.24	0.539	0.0023
Median	0.0003	1.86	0.12	0.760	0.0011
25th percentile	-0.0011	1.78	-0.31	0.421	0.0013
5th percentile	0.0003	1.71	0.28	0.464	0.0003
POT threshold 2.1 m					
Mean	0.0010	2.30	0.15	0.702	0.0024
95th percentile	-0.0007	2.65	-0.03	0.934	0.0086
75th percentile	-0.0011	2.40	-0.15	0.709	0.0030
Median	0.0030	2.24	0.26	0.497	0.0041
25th percentile	0.0037	2.15	0.55	0.122	0.0021
5th percentile	-0.0002	2.12	-0.09	0.825	0.0010

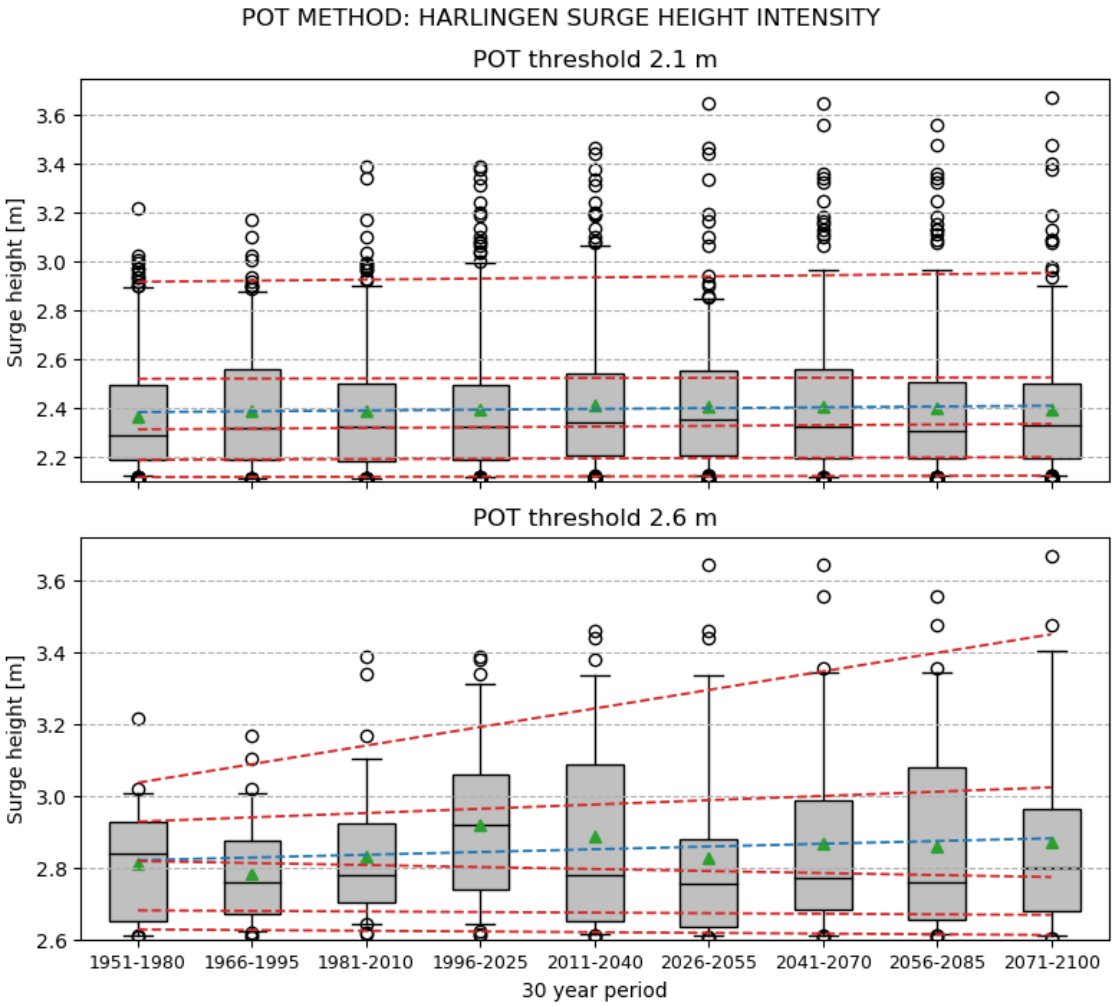

Figure 5-8. Surge height intensity resulting from the POT method for Harlingen. The boxplots are created from the peak surge heights of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-7. Linear trend fitting coefficients of the surge height intensity boxplots resulting from the POT method for Harlingen (corresponding to Figure 5-8).

POT METHOD: HARLINGEN SURGE HEIGHT INTENSITY					
	Slope (a) [m/15y]	Intercept (b) [m]	Correlation (r)	p-value	Standard error [m]
POT threshold 2.1 m					
Mean	0.0033	2.38	0.65	0.058	0.0015
95th percentile	0.0045	2.92	0.18	0.645	0.0093
75th percentile	0.0007	2.52	0.07	0.859	0.0039
Median	0.0028	2.31	0.40	0.281	0.0024
25th percentile	0.0014	2.19	0.42	0.256	0.0011
5th percentile	0.0007	2.12	0.44	0.235	0.0006
POT threshold 2.6 m					
Mean	0.0076	2.82	0.50	0.173	0.0050
95th percentile	0.0517	3.04	0.90	0.001	0.0097
75th percentile	0.0119	2.93	0.39	0.297	0.0106
Median	-0.0056	2.82	-0.29	0.449	0.0070
25th percentile	-0.0016	2.68	-0.14	0.725	0.0044
5th percentile	-0.0020	2.63	-0.40	0.286	0.0017

Figure 5-9 shows the result of the POT method for total water level intensity of storm surge events at Hoek van Holland for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-8. No significant trends were found at Hoek van Holland for the total water level intensity criterion. Except for the 5th percentile trend fit of the 1.7 meter POT threshold, all obtained trend lines have negative slopes. This observation suggests that a decreasing trend of the total water level intensity is more likely than an increasing trend.

Figure 5-10 shows the result of the POT method for total water level intensity of storm surge events at Harlingen for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-9. Three significant trends were found, which were obtained from the 2.1 meter POT threshold results. The 25th percentile, median and 75th percentile fitted trend lines have p-values of 0.021, 0.048 and 0.027 respectively. Also, the p-value of the mean trend line is close to being significant. In contradiction to the increasing tendency found for surge height intensity, these trend lines all have negative slope. This suggests that the magnitude of surge height intensity and total water level intensity is not necessarily correlated. The 2.6 meter POT threshold result contains too much interdecadal variation for trends to be significant, which can be concluded from the correlation and standard error coefficients.

For both locations the obtained total water level intensity trend lines all have negative slopes, except for the 5th percentile trend fit of Hoek van Holland for the 1.7 meter POT threshold. Also, the higher POT thresholds of both locations show more skewed distributions, which can be observed by the difference between the means and medians. The means in these results are systematically lower than the medians, which indicates that total water levels above the means are more likely to occur than below the means.

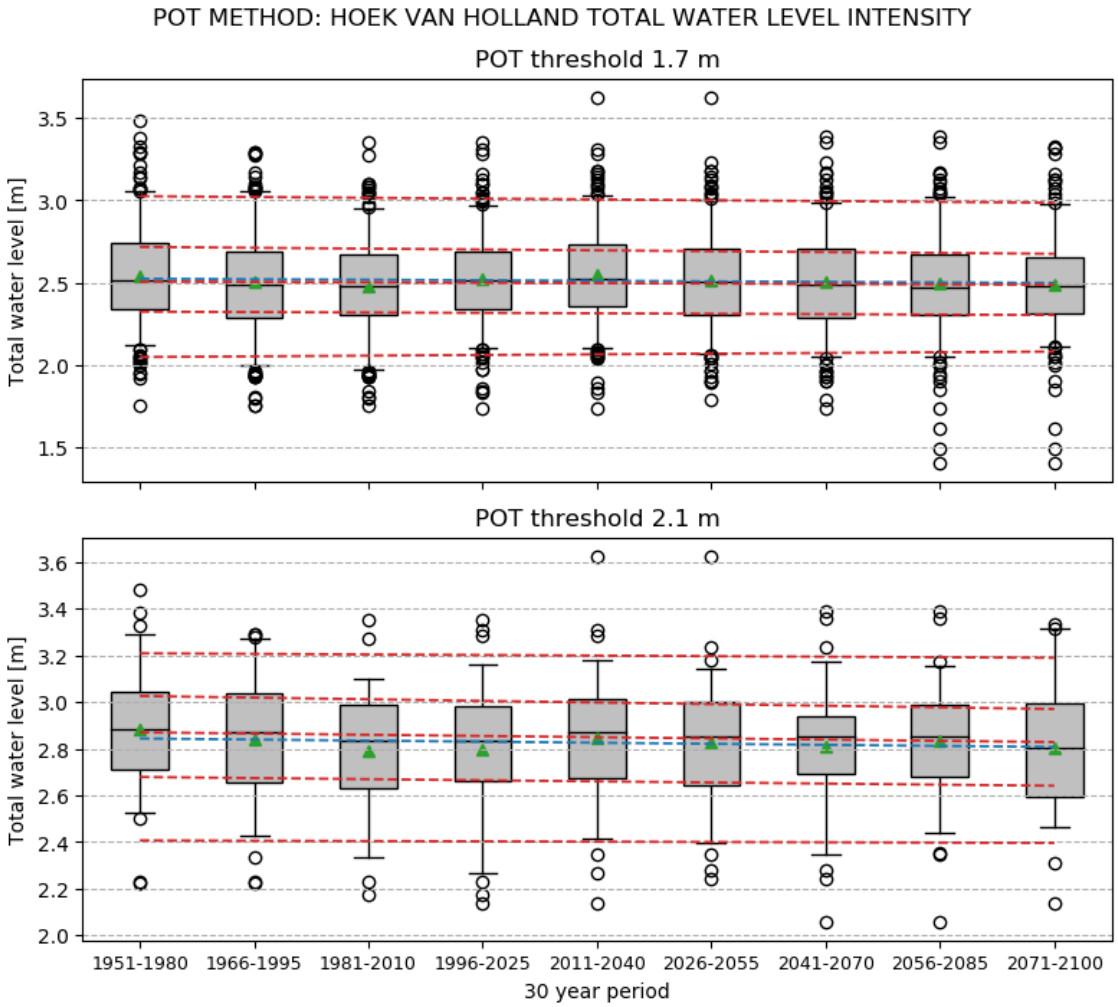

Figure 5-9. Total water level intensity resulting from the POT method for Hoek van Holland. The boxplots are created from the peak total water levels of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-8. Linear trend fitting coefficients of the total water level intensity boxplots resulting from the POT method for Hoek van Holland (corresponding to Figure 5-9).

POT METHOD: HOEK VAN HOLLAND TOTAL WATER LEVEL INTENSITY					
	Slope (a) [m/15y]	Intercept (b) [m]	Correlation (r)	p-value	Standard error [m]
POT threshold 1.7 m					
Mean	-0.0032	2.52	-0.40	0.292	0.0028
95th percentile	-0.0050	3.03	-0.37	0.325	0.0047
75th percentile	-0.0052	2.72	-0.51	0.162	0.0033
Median	-0.0025	2.51	-0.35	0.355	0.0025
25th percentile	-0.0025	2.32	-0.26	0.497	0.0035
5th percentile	0.0041	2.05	0.22	0.570	0.0069
POT threshold 2.1 m					
Mean	-0.0045	2.84	-0.43	0.253	0.0036
95th percentile	-0.0024	3.21	-0.09	0.821	0.0104
75th percentile	-0.0070	3.03	-0.62	0.075	0.0033
Median	-0.0052	2.87	-0.57	0.111	0.0029
25th percentile	-0.0047	2.68	-0.38	0.310	0.0043
5th percentile	-0.0013	2.41	-0.05	0.903	0.0106

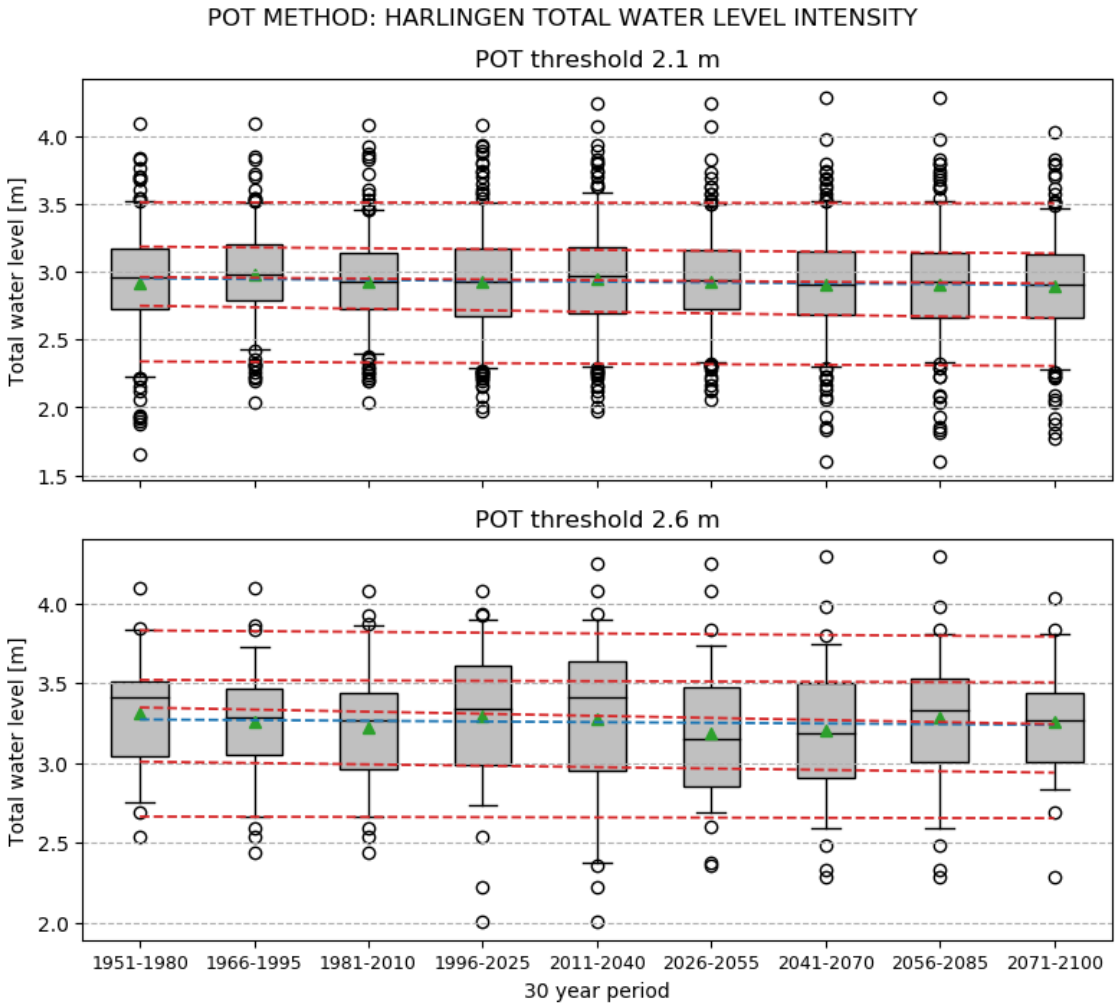

Figure 5-10. Total water level intensity resulting from the POT method for Harlingen. The boxplots are created from the peak total water levels of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-9. Linear trend fitting coefficients of the total water level intensity boxplots resulting from the POT method for Harlingen (corresponding to Figure 5-10).

POT METHOD: HARLINGEN TOTAL WATER LEVEL INTENSITY					
	Slope (a) [m/15y]	Intercept (b) [m]	Correlation (r)	p-value	Standard error [m]
POT threshold 2.1 m					
Mean	-0.0062	2.95	-0.66	0.053	0.0027
95th percentile	-0.0009	3.51	-0.06	0.869	0.0053
75th percentile	-0.0062	3.19	-0.73	0.027	0.0022
Median	-0.0060	2.96	-0.67	0.048	0.0025
25th percentile	-0.0113	2.75	-0.75	0.021	0.0038
5th percentile	-0.0042	2.34	-0.18	0.638	0.0084
POT threshold 2.6 m					
Mean	-0.0042	3.27	-0.28	0.459	0.0054
95th percentile	-0.0048	3.83	-0.20	0.606	0.0088
75th percentile	-0.0020	3.52	-0.08	0.843	0.0097
Median	-0.0130	3.35	-0.41	0.279	0.0111
25th percentile	-0.0086	3.01	-0.37	0.328	0.0082
5th percentile	-0.0013	2.66	-0.03	0.944	0.0180

5.2.4. Trends in duration of storm surge events (1951-2100)

The storm surge duration values mentioned in this section are durations above a surge height of 0.5 meters of storm surge events.

Figure 5-11 shows the result of the POT method for storm surge duration at Hoek van Holland for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-10. None of the obtained duration trend lines are found to be significant. The 2.1 meter POT threshold contains a lot of interdecadal variation in comparison to the 1.7 meter POT threshold. Storm surge events that reach a surge height of 1.7 meters or higher have generally a similar duration of those of 2.1 meters or higher, but the POT threshold of 1.7 meter results includes the highest durations, namely durations of more than 100 hours.

Figure 5-12 shows the result of the POT method for storm surge duration at Harlingen for both POT thresholds. The figure also shows the fitted linear trend lines, from which the obtained coefficients are given in Table 5-10. For the 2.1 meter POT threshold five significant trend lines were obtained. Only the 75th percentile trend fit for this POT threshold is not significant. For the 2.6 meter POT threshold one significant trend line was obtained. The mean trend fit of the 2.1 meter POT threshold resulted in a change over the entire data timespan of approximately +4.2 hours. The same value for the 2.6 meter POT threshold is approximately +4.9 hours. Storm surge durations at Harlingen can reach approximately 160 hours for the POT threshold of 2.1 meters. All significant trend lines have a positive slope, thus the general observation for Harlingen is that an increase in storm surge duration is likely.

Comparing the results of both locations, Harlingen shows clear increasing trends for specifically the lower POT threshold, whereas for Hoek van Holland no clear overall trend assessment could be made. The durations at Harlingen are systematically longer than at Hoek van Holland with a factor of approximately 1.5.

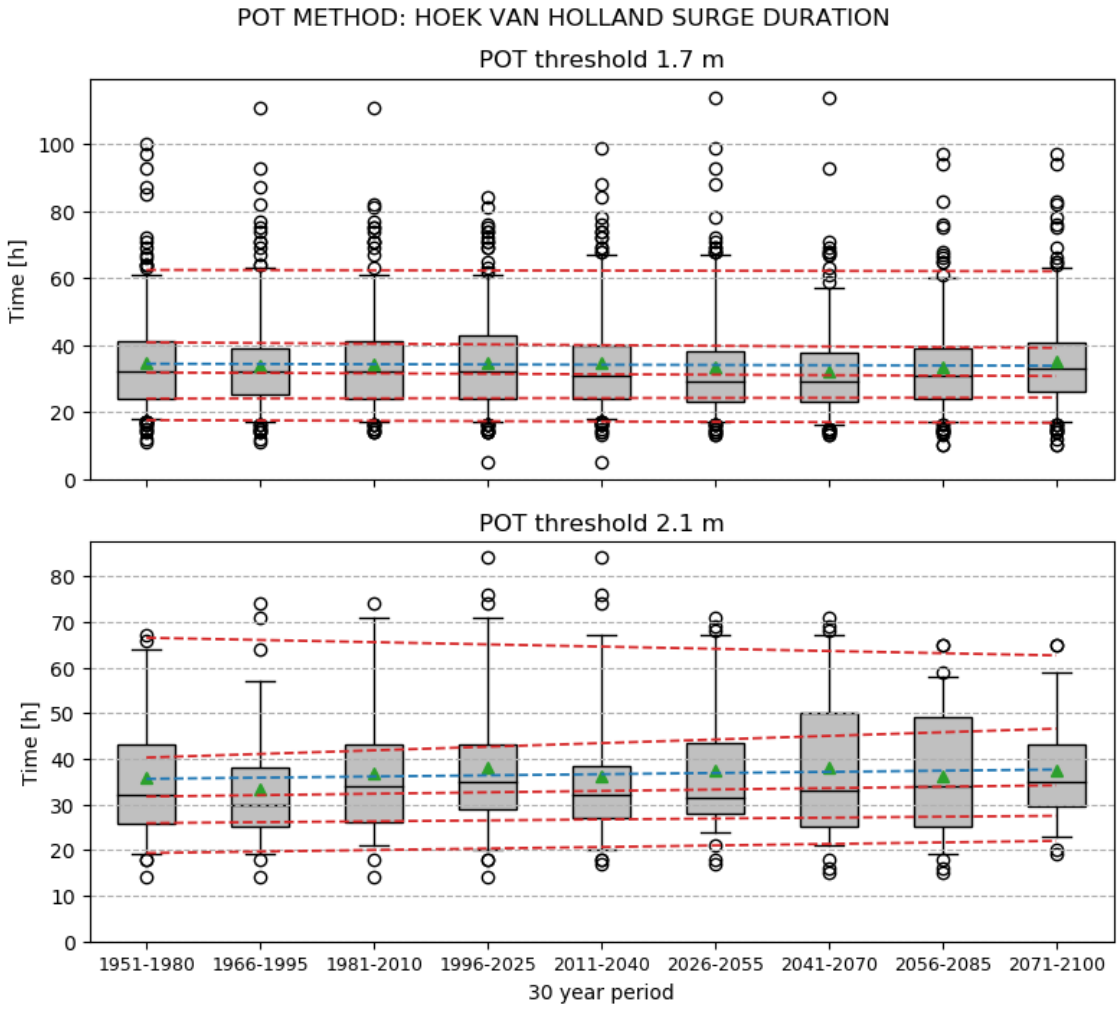

Figure 5-11. Storm surge event duration of Hoek van Holland. The boxplots are created from the duration above the 0.5 meter surge height of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-10. Linear trend fitting coefficients of the surge duration boxplots of Hoek van Holland (corresponding to Figure 5-11).

POT METHOD: HOEK VAN HOLLAND SURGE DURATION					
	Slope (a) [h/15y]	Intercept (b) [h]	Correlation (r)	p-value	Standard error [h]
POT threshold 1.7 m					
Mean	-0.073	34.4	-0.20	0.608	0.136
95th percentile	-0.05	62.4	-0.04	0.914	0.446
75th percentile	-0.208	40.8	-0.34	0.371	0.218
Median	-0.133	31.8	-0.26	0.496	0.186
25th percentile	0.033	24.0	0.10	0.801	0.127
5th percentile	-0.1	17.5	-0.46	0.218	0.074
POT threshold 2.1 m					
Mean	0.254	35.6	0.48	0.190	0.175
95th percentile	-0.483	66.5	-0.25	0.524	0.721
75th percentile	0.792	40.3	0.54	0.138	0.472
Median	0.308	31.7	0.50	0.175	0.204
25th percentile	0.2	25.9	0.31	0.417	0.232
5th percentile	0.333	19.3	0.51	0.164	0.215

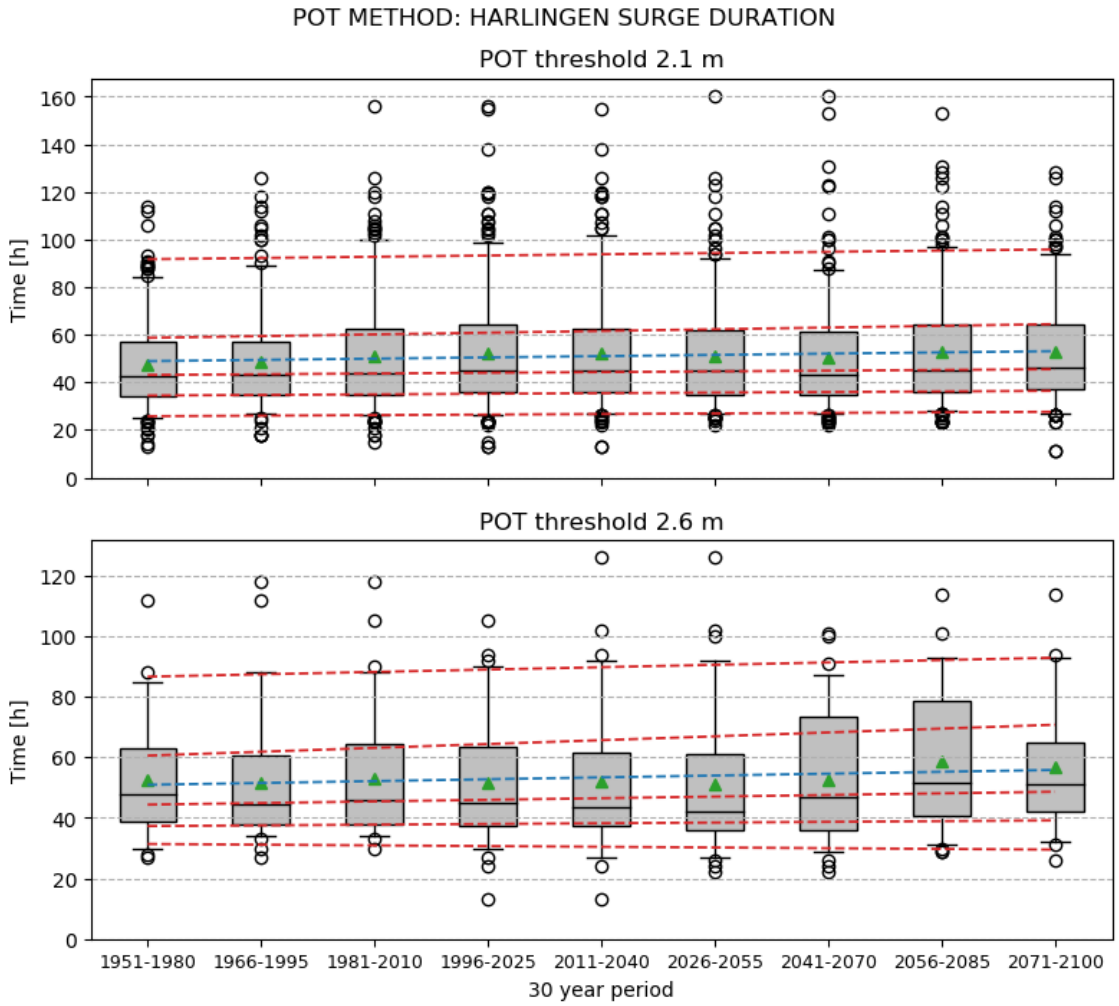

Figure 5-12. Storm surge event duration of Harlingen. The boxplots are created from the duration above the 0.5 meter surge height of storm surge events. The boxplots are for overlapping periods of 30 years from 1951 till 2100 for two POT thresholds and show the mean (green triangles), the median (middle box line), 25th and 75th percentiles (box edges), 5th and 95th percentiles (whisker caps), and extremes (circles). Linear trend lines are fitted through these percentiles (dashed red) and means (dashed blue).

Table 5-11. Linear trend fitting coefficients of the surge duration boxplots of Harlingen (corresponding to Figure 5-12).

POT METHOD: HARLINGEN SURGE DURATION					
	Slope (a) [h/15y]	Intercept (b) [h]	Correlation (r)	p-value	Standard error [h]
POT threshold 2.1 m					
Mean	0.526	48.9	0.73	0.025	0.185
5th percentile	0.517	91.7	0.23	0.557	0.839
25th percentile	0.721	58.7	0.71	0.031	0.271
Median	0.3	43.1	0.68	0.042	0.121
75th percentile	0.246	34.4	0.75	0.021	0.083
95th percentile	0.233	25.7	0.74	0.023	0.081
POT threshold 2.6 m					
Mean	0.614	50.9	0.66	0.052	0.262
95th percentile	0.783	86.6	0.74	0.023	0.271
75th percentile	1.275	60.6	0.57	0.108	0.691
Median	0.533	44.4	0.45	0.222	0.398
25th percentile	0.233	37.3	0.32	0.405	0.264
5th percentile	-0.233	31.4	-0.25	0.524	0.348

5.2.5. Summary of trend results and characteristic values

In the preceding sections of this chapter, storm surge trends were assessed for three criteria: frequency, intensity and duration of storm surge. This section provides an overview of these results per location (see also Table 5-12). In addition, characteristic values of storm surge (cluster) events are provided (Table 5-13).

For the location of Hoek van Holland mainly no significant trends were found, except for the frequency of storm surge events criterion. The frequency criterion resulted in one clear significant decreasing trend and one insignificant decreasing trend for the single storm surge events. This shows that a decreasing trend in the frequency of storm surge events at Hoek van Holland is more likely. From these linear trends the total change over the period 1951-2100 is estimated to be a decrease of 0 to 3 events occurring in 30 years. For the frequency of storm surge cluster events two significant decreasing trends were found. This result is consistent with the result of single storm surge events. However, the variation between the three tested definitions of storm surge cluster events was found to be large, which makes these results less reliable. The intensity criterion resulted in no significant trends for the intensity of surge height and total water level of storm surge events. Also, no significant trends were found for the duration criterion.

For the location of Harlingen mainly increasing trends were found. For the frequency criterion, one increasing significant trend for single storm surge events was obtained, although with a relatively small slope. Therefore, the change over the period 1951-2100 based on the linear trends is estimated to be an increase of just 0 to 1 events occurring in 30 years. The frequency criterion for storm surge cluster events resulted in only significant trends, of which two were increasing and one decreasing. Therefore, no unequivocal conclusion about trends could be drawn for the storm surge cluster events. The intensity criterion resulted in a mix of increasing and decreasing significant trends. The overall observation here is that the most extreme surge heights increases from approximately 3.0-3.2 meters in 1951-1980 to approximately 3.4-3.7 meters in 2071-2100, but that all selected surge heights above 2.6 meters do not increase on average. In other words, the variance of storm surge events with a surge height of 2.6 meter or higher seems to increase towards the end of the 21st century. The results of intensity of total water level shows only decreasing trends, of which three out of twelve trend lines are found to be significant. The total approximate change of total water level over the 1951-2100 period is a 0.0 to 0.1 meter decrease. Hence, the observation of increase in surge height intensity is not reflected in the total water level intensity of the same storm surge events. The duration criterion results showed mainly increases. Six significant trends were obtained, which suggests that an increase is very likely. The total increase of the fitted linear trends over the 1951-2100 period is in the range of 2 to 6 hours.

Table 5-12. Overall trend assessment results of all criteria for Hoek van Holland and Harlingen. Shown is whether a clear trend is found, and if so, the total approximate increase/decrease over the period 1951-2100 is given as a range.

	Hoek van Holland	Harlingen
Storm surge events (and block extremes method result)		
Frequency	0-3 events occurring in 30 years decrease	0-1 events occurring in 30 years increase
Intensity: surge height	No clear trend	0.4-0.5 m increase of extremes, but no clear overall trend
Intensity: total water level	No clear trend	0.0-0.1 m decrease
Duration	No clear trend	2-6 hours increase
Storm surge cluster events		
Frequency	No clear trend	No clear trend

In addition to the trend results, some characteristic values of storm surge (cluster) events are calculated and presented in Table 5-13. A first observation is that mean durations at Harlingen are approximately an order of 1.5 higher than at Hoek van Holland. Secondly, Harlingen has to

cope with extremer surge heights in general. For a similar frequency of storm surge events the surge heights at Harlingen are about 0.4 meters higher than at Hoek van Holland.

Table 5-13. Characteristic storm surge values resulting from the POT method. Shown are mean values calculated from the entire data timespan (from 1951-1980 till 2071-2100).

	Hoek van Holland			Harlingen		
POT threshold [m]	1.7	2.1		2.1	2.6	
Nb. of storm surge events [/30y]	15.6	2.9		15.6	2.8	
Surge height intensity [m]	1.9	2.3		2.4	2.9	
Total water level intensity [m]	2.5	2.8		2.9	3.3	
0.5 m surge height duration [h]	34	37		51	54	
POT threshold [m]	0.8	1.0	1.3	1.0	1.2	1.6
Nb. of cluster events [/30y]	6.7	6.9	6.4	5.6	7.0	6.2

6. Discussion and conclusions

In the first three sections the sensitivity of the definitions of a storm surge event and storm surge cluster event, the model accuracy GPD fit and the trend assessment method is respectively tested. In sections 6.1.4 and 6.1.5, the results of the research questions are discussed and possible explanations are addressed. The research limitations, an assessment of the impact on the Dutch coast and recommended future research are included in sections 6.1.6, 6.1.7 and 6.1.8 respectively. Finally, conclusions for the research questions are drawn.

6.1. Discussion

6.1.1. Storm surge (cluster) event definition sensitivity

The main method used in this research is that of the peak over threshold (POT) method. In this method the definition of a storm surge event and a storm surge cluster event play a major role in the obtained results. In the process leading to these definitions certain conditions are implemented. For example, the condition of successive storm surge event independency, which is defined as a minimum of 24 hours in between the end and start times of successive storm surge events. This value is determined partly by analysing the occurrences of time intervals between the threshold exceedances and partly by consulting literature, which makes it arguable. Other studies also used an interval of 24 hours (e.g., Caires, 2011; Tijssen & Diermanse, 2009), but longer implemented time intervals are also encountered, e.g. 3 days (van den Brink & Können, 2011). Furthermore, sometimes this interval is measured from the times of the successive peak values, instead of from the end and start times. Additionally, (Dillingh et al., 1993) stated that the average duration of a North Sea storm (in atmospheric terms) is 2.41 days. In this part of the definition there will most probably always be cases of resulting successive storm surge events which are part of the same atmospheric storm. However, the influence on the final results of this phenomenon are relatively small due to the relatively small number of these cases. One could also argue that a storm surge event should be defined with the condition of a minimal duration. This would lead, depending on the value of the minimal duration, to systematically longer durations. This type of condition is not implemented in this research, since consulted literature did not show such conditions.

To further discuss the used definitions a sensitivity analysis is conducted. This is done in two ways. Firstly, a check is done to inspect successive storm surge event independency from the water level perspective, instead of from the time perspective. Secondly, used variables in the definitions are varied to examine the effect of the definitions on the results. The first method is done to verify if the eventually defined storm surge events with a time interval threshold of 24 hours are virtually independent. For this the occurrences of the surge height being above 0.5 meter the entire time interval between successive storm surge events is calculated. This is expressed in percentage of total number of successive storm surge events for a range of POT thresholds (Figure 6-1). From the figure can be observed that the highest POT threshold at

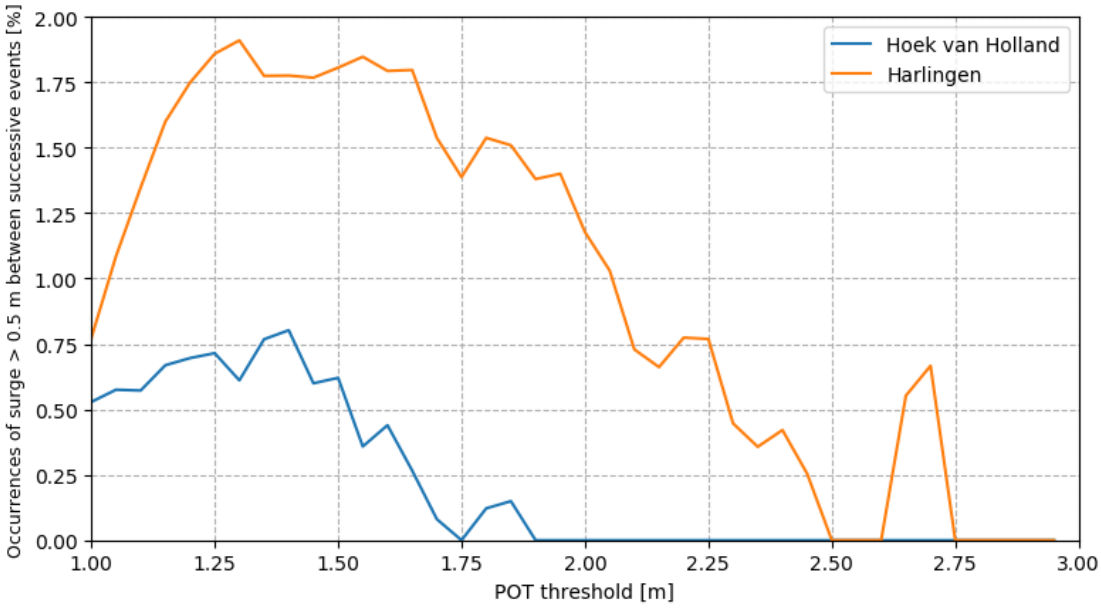

Figure 6-1. Occurrences of the surge height being above 0.5 meter the entire time interval between successive storm surge events expressed in percentage of total number of successive storm surge events. Values result from the POT method with an implemented time interval threshold of 24 hours for POT thresholds in the range of 1.00 to 3.00 meters.

Hoek van Holland where this phenomenon can be observed is at 1.86 meters. Above this value any applied POT threshold results in zero occurrences of the surge height staying above the 0.5 meter water level between successive storm surge events. In Harlingen however, this phenomenon happens more frequently. Nevertheless, for both locations for all POT thresholds between 1 and 3 meters this phenomenon does not occur for more than 1.92% of the total number of successive storm surge events. From this water level perspective check can be concluded that a very small part of successive storm surge event (less than 1.92%) could still belong to the same atmospheric storm. Therefore, the selected storm surge events in this research are classified as virtually independent.

In the second method for the storm surge events, the implemented time interval threshold to account for successive storm surge event independency is varied. Three values are used, namely 12, 24 (the original value) and 36 hours. The percentage of occurrences of time intervals between successive threshold exceedances being below each of these three values of the total occurrences lower than or equal to 72 hours are calculated (Table 6-1). For each location three POT thresholds are applied to cover the largest part of the entire possible ranges of POT thresholds. The type of analysis is also conducted in section 3.1.3, where also figures are provided (Figure 3-3 and Figure 3-5). Two observations can be made from the values in the table. Firstly, the percentages are fairly constant for all time interval values, with small and obvious increases for higher time intervals. These increases become smaller for higher thresholds. Secondly, higher thresholds in general result in higher percentages, suggesting that the extremer storm surge events occur less often in close succession. For example, in Hoek van Holland with a POT threshold of 2.1 meters, a relatively large percentage of threshold exceedances falls within the time interval, which therefore would all be merged to storm surge events. These observations lead to the conclusion that storm surge events with any of the three implemented time interval thresholds tackles the factor of successive independency sufficiently. Hence, the definition of a storm surge event is hardly affected by the applied time interval thresholds as long as it is in the range of 12 to 36 hours.

Table 6-1. Percentage of occurrences of time intervals between successive threshold exceedances being below the time interval threshold of the total occurrences lower than or equal to 72 hours. Percentages are calculated by occurrences lower than the time interval divided by occurrences lower than or equal to 72 hours.

	Hoek van Holland			Harlingen		
POT threshold [m]	1.7	1.9	2.1	2.1	2.35	2.6
Time interval < 12 hours [%]	92.5	94.9	96.1	87.7	89.0	93.5
Time interval < 24 hours [%]	93.4	95.5	97.4	90.5	91.2	97.8
Time interval < 36 hours [%]	94.8	96.0	97.4	93.0	92.6	97.8

In the second method for the storm surge cluster events, the sensitivity analysis is also done by varying the time interval. The definition states that successive storm surge events have to occur between a time interval of 24 and 72 hours (section 3.2.2), resulting in a time window of 48 hours (i.e. 72 minus 24). The lower threshold is adjacent to the time interval threshold used in the definition of a storm surge event. Since, the values 12, 24 and 36 hours are used in the former sensitivity analysis, these values are here implemented as the lower thresholds. Therefore, the following time interval values are used: between 12 and 60 hours, between 24 and 72 hours (the original values) and between 36 and 84 hours. The number of storm surge cluster events within the entire data period (1951-2100) are calculated for different minimal number of storm surge events in a storm surge cluster event (Table 6-2). Important to notice is that the time window length (i.e. time between lower and upper time interval threshold) is not altered in this sensitivity analysis. Two observations can be made from the values in the table. Firstly, the differences for a minimal number of 4 events between the time interval values is largest if compared to the other minimal number of events conditions. Secondly, the number of cluster events in general decreases with a higher time interval. This can be seen by that the lowest table row shows the lowest values for each column. Conclusive, the definition of a storm surge cluster is highly dependent on the time interval thresholds used. In section 5.2.2 was

concluded that the definition is also highly dependent on the minimal number of storm surge events in a storm surge cluster event. Therefore, a definition for the clustering phenomenon of storms from a water level perspective might not be the most suitable one.

Table 6-2. Number of storm surge cluster events within the entire data period (1951-2100) for different minimal number of storm surge events in a storm surge cluster event.

	Hoek van Holland			Harlingen		
POT threshold [m]	0.8	1.0	1.3	1.0	1.2	1.6
Minimal number of storm surge events	4	3	2	4	3	2
12 hours ≤ time interval < 60 hours	696	600	525	602	647	516
24 hours ≤ time interval < 72 hours	532	541	496	443	550	490
36 hours ≤ time interval < 84 hours	323	371	430	257	389	385

6.1.2. Model accuracy GPD fit sensitivity

To provide insight in the accuracy of the model combination RACMO and WAQUA the water level exceedance frequencies of the WAQUA data were compared with those based on measurement data (chapter 4). Herein, the GPD was fitted to the selected peak water levels above a water level threshold corresponding to an exceedance frequency of 0.5 per year (Figure 4-8). This threshold is applied, because this threshold was also used for the measurement data based exceedance frequencies in Dillingh (2013). In this section this threshold parameter (μ) of the GPD is altered at the start of the calculation procedure (i.e. before the shape and scale parameters are calculated), to provide insight in the sensitivity of this parameter. Additionally, a lower exceedance frequency boundary is implemented that determines which part of the highest water levels are left out of the GPD fit procedure. For both of these conditions three values are applied in this sensitivity analysis. For exceedance frequency q_μ (which corresponds to water level threshold parameter μ) the following values are applied: 10^0, $5 * 10^{-1}$ $(= 0.5)$, and 10^{-1} per year. For the lower exceedance frequency boundary, the following values are applied: no exceedance frequency lower boundary, 10^{-3} and $5 * 10^{-3}$ per year. Hence, there are in total nine combinations of these conditions, and therefore nine groups of selected peak water levels to which the GPD is fitted.

The results are presented in Figure 6-2 and Figure 6-3 for Hoek van Holland and Harlingen respectively. In these figures the applied lower exceedance frequency (ef) boundary can be interpreted as a vertical line that determines which water levels shown in Figure 4-8 as plus signs are included in the GPD fitting procedure. For example, the exceedance frequency lower boundary of 10^{-3} leaves out the two highest water levels for both locations. The black solid lines in Figure 6-2 and Figure 6-3 are the GPD fits used to assess the model accuracy in chapter 4 (i.e. the original lines). From the results can be observed that when more of the highest water levels are left out of the GPD fitting procedure, the more the line tends to diverge towards lower water levels for the lowest exceedance frequencies. This can be observed for both locations. Furthermore, for Hoek van Holland, a lower q_μ and thus a higher μ (water level threshold), also results in this “diverging” phenomenon mentioned. However, for Harlingen this does not apply. For Harlingen, at the lower exceedance frequencies, the highest water levels are observed when $q_\mu = 0.5$. In general can be concluded that leaving water levels out of the GPD fitting procedure has a prominent effect on the obtained GPD fit and it seems that rather more selected peak water levels than less leads to a more sensible resulting GPD fit.

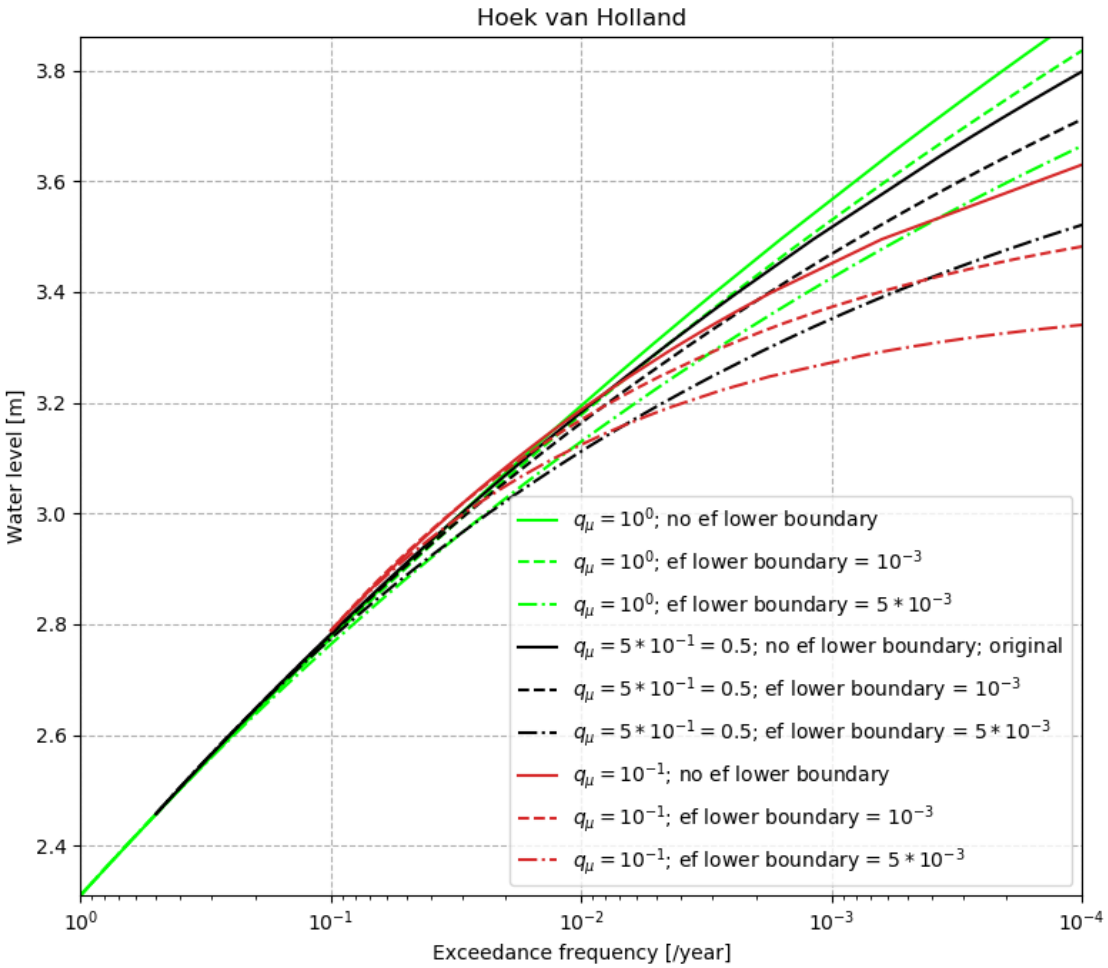

Figure 6-2. Nine different GPD fits for Hoek van Holland. The selected peak water levels to which the GPDs are fitted is altered by applying exceedance frequency boundaries on the left side of the graph (q_μ) and right side of the graph (exceedance frequency (ef) lower boundary). The solid black line is the original used GPD fit (in chapter 4).

6.1.3. Trends assessment sensitivity

Due to the result of the sensitivity analysis in section 6.1.1, the trends were also visually analysed for the different time interval thresholds. This resulted in negligible differences of fitted linear trend results between the original implemented time interval thresholds and the varied time interval thresholds. Therefore, plots of these results are not provided. Nevertheless, the largest identified changes in the means are mentioned. Mean durations increased with a higher time interval threshold for just a few 30 year periods, but with no more than 2 hours. This was only the case for the lower of two POT thresholds of each location. The mean total water level intensities increased with a higher time interval threshold for just a few 30 year periods, but with no more than 3 centimeters. This observation and value are the same for the mean surge height intensities. For the number of storm surge events trends remain similar, with negligible changes in numbers of storm surge events. Storm surge cluster events show much larger differences in trends. Observed is an increase in variation between the minimum number of events definitions and also between 30 year periods. The variation increases with an increase in the implemented time interval thresholds, which is a logical result following from the observations done in the definition sensitivity analysis in section 6.1.1, since sample sizes are also decreasing.

In the trend assessment of chapter 5, linear trends are fitted to overlapping 30 year periods. For the frequency criterion the successive 30 year periods overlap with 29 years, whereas for the intensity and duration criteria a 15 year overlap is used. This overlapping method (also known are running mean or a smoothing algorithm) in general results in lower p-values. Hence, trend

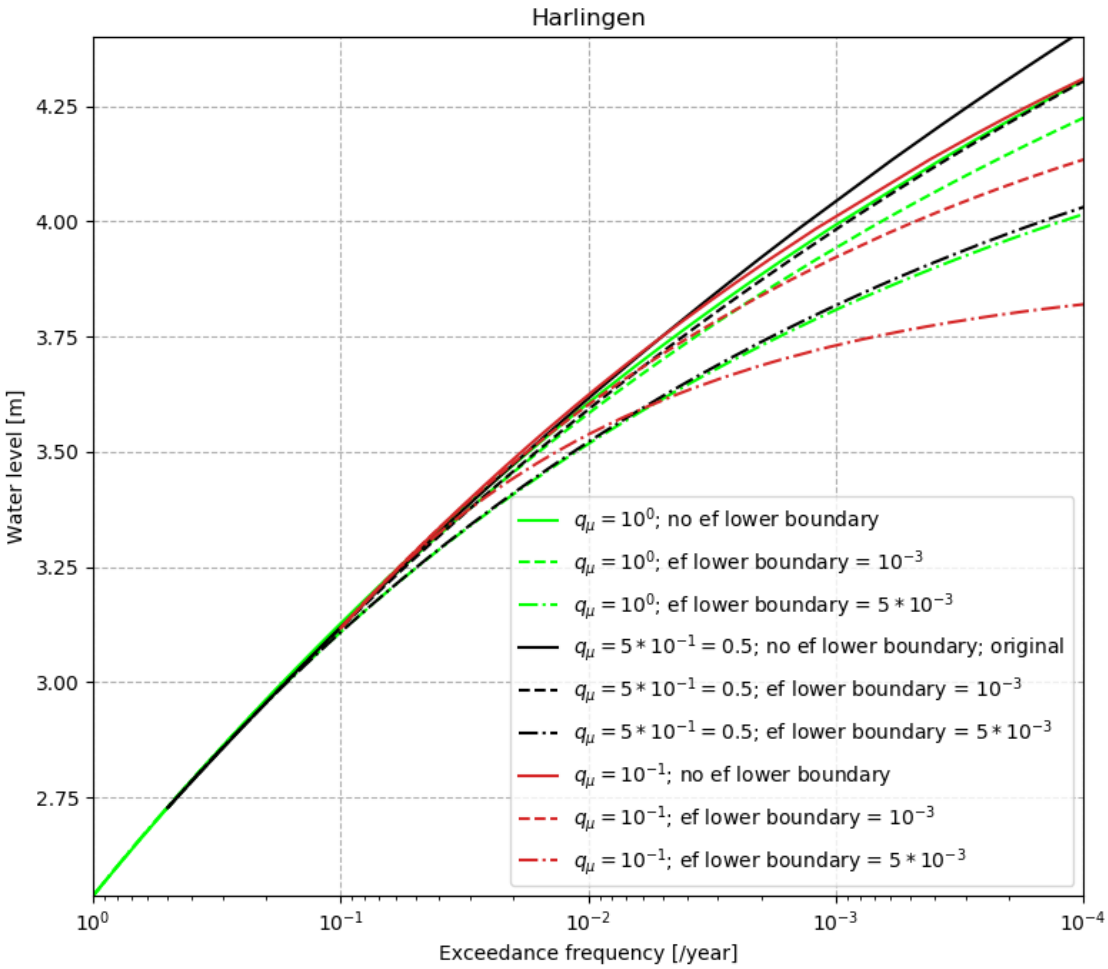

Figure 6-3. Nine different GPD fits for Harlingen. The selected peak water levels to which the GPDs are fitted is altered by applying exceedance frequency boundaries on the left side of the graph (q_μ) and right side of the graph (exceedance frequency (ef) lower boundary). The solid black line is the original used GPD fit (in chapter 4).

lines are more often significant in comparison to when non-overlapping 30 periods would have been used. In this section the sensitivity of this overlapping periods choice is tested by fitting the trend lines to only the non-overlapping 30 year periods, that is 1951-1980, 1981-2010, 2011-2040, 2041-2070 and 2071-2100 for all trend assessment criteria.

For the frequency criterion only a significant p-value was found for Hoek van Holland for the POT threshold of 1.7 meters, namely a p-value of 0.017. This is the only significant trend obtained for Hoek van Holland in this sensitivity analysis. In the trend assessment of chapter 5, this POT threshold resulted also in a significant trend. For Harlingen however, no significant trend resulted from the non-overlapping periods trend fitting of the frequency criterion, whereas one significant trend was found in the original trend assessment. For the frequency of storm surge cluster events of both locations no significant trends resulted for the non-overlapping periods. This contradicts greatly with the original trend assessment, where five out of six obtained trends were found to be significant. However, the overall observation for storm surge cluster events was that there was too much variety between the results of the used definitions that not an unequivocal trend could be determined. This overall observation is thus in agreement with the results found in this sensitivity analysis. The intensity criterion resulted only in significant trends for Harlingen, namely a p-value of 0.016 for the 95th percentile of surge height intensity for the 2.6 meter POT threshold and a p-value of 0.006 for the 25th percentile of total water level intensity for the 2.1 meter POT threshold. This first result is in agreement with the original trend assessment, where also a significant trend was obtained for the 95th percentile. For the second result, in the original trend assessment were three out of six obtained trends significant, whereas

for the non-overlapping periods only one out of six obtained trends is significant. Finally, the duration criterion resulted in one significant trend with a p-value (0.047) that is just below the significance level of 0.05. This trend was found for the 5th percentile for the 2.1 meter POT threshold for Harlingen. In the original trend assessment, the duration criterion for Harlingen resulted in six out of twelve obtained trends that were significant. By using the non-overlapping 30 year periods, it shows that in general less significant trends are obtained.

A few of the most significant trends found in the trend assessment of chapter 5 are also significant when using non-overlapping periods. Two noteworthy differences are found. Firstly, the duration criterion for Harlingen shows clearly less significance and the frequency of storm surge cluster events shows completely different results. However, for the latter was already determined that no clear trends came out. These results show that one should be careful with making statistical statements when only examining overlapping periods.

6.1.4. Model accuracy

The model accuracy analysis resulted in significant differences between the measurement data based and WAQUA model water level exceedance frequencies. Possible explanations of this result could originate from the physical working of the combination of the RACMO and WAQUA model. Climate change in RACMO is only simulated by the greenhouse gas forcing of RCP8.5. The IPCC (2013) reports that in general models still produce a storm track that is too zonal and underestimate cyclone intensity, although that storm track biases in the North Atlantic have improved slightly (Zappa et al., 2013). In the research of Woth et al. (2006) such a result was found where the surge intensity was generally too weak leading to an underestimation of the storm surge 99.5th percentile. This finding was consistent with Flather's and Smith's (1998, referred to in Woth et al., 2006). Van Meijgaard et al. (2012) conducted research to the performance of the RACMO model at high resolution and found that although high resolution is providing much more spatial refinement, it does in general not add information to the statistical distribution of high wind speeds associated to large-scale baroclinic instabilities that could not already be obtained from the coarser gridded models, as opposed to extreme precipitation. This is because the high wind speeds are associated to a large-scale phenomenon, whereas extreme precipitation events are frequently triggered by medium to small-scale atmospheric disturbances (van Meijgaard et al., 2012). On the other hand the atmospheric structures favouring the most extreme wind speeds are generated at scales that are finer than current RCM's can resolve (van Meijgaard et al., 2012). If this is of significant influence on storm surge along the Dutch coast is not clear, since an assessment of the physical working of the models is beyond the scope of this research, but it could be the explanation for the much lower water levels for the lower exceedance frequencies in comparison to measurement data based values. This statement also holds for the WAQUA model, which does not take the influence of surge on the tides into account. All this leaves the question whether climate models in general are accurately able to predict extremes unanswered.

Another explanation of the large differences could be the applied calculation method. The method makes use of the Generalised Pareto Distribution (GPD) and the Maximum Likelihood (ML) estimation of the GPD parameters. The sensitivity analysis in section 6.1.2 shows that the GPD fit can differ significantly when leaving points (water levels) out. For the GPD fit also holds true that the level of confidence decreases with a decrease in exceedance frequency. The statistical confidence can be improved by extending the simulation record length, although it must be recognised that the quality of model simulations of extreme events with such long return periods is difficult to assess (van den Hurk et al., 2015).

6.1.5. Storm surge trends

Several studies argue that more severe storms will hit the Netherlands in the future due to climate change (Haarsma et al., 2013; Pinto et al., 2013; Ulbrich et al., 2009). However, storm surge occurrence along coastlines is greatly dependent on wind directions. For the Dutch coast northerly winds are most dangerous because they have the longest fetch (Sterl et al., 2009). The factor of wind direction could be an explanation for the differences obtained between

Harlingen and Hoek van Holland. The orientation of the North Sea coastline near Harlingen is facing more towards the north as shown in Figure 2-4. Sterl et al. (2009) used a global climate model in combination with WAQUA/DCSM98 to calculate changes in 10,000 return values of surge heights during the 21st century at Hoek van Holland. They did not find any significant change. The WASA Group (1998) results point to a moderate future increase of surges in the North Sea. However, this increase is within the range of previously observed natural variations, leading to the conclusion it cannot be directly attributed to climate change. A similar result was found by Alexandersson et al. (1998, 2000). Weisse et al. (2012) concluded that storm activity in the North Sea area underwent considerable variations on time scales of decades and longer, but no clear long-term trends could be identified. Time periods of increased storm surge levels prevailed in the late nineteenth and twentieth centuries without any evidence for significant long-term trends (Dangendorf et al., 2014). Hence, multiple studies in the same field of research show comparable results as found in this research, that is, mainly no significant trends or small significant trends. A comparison with the results of storm surge cluster events is not made, since no comparable studies were found.

A possible explanation of the lack of significant trends could be due to the choice of the locations Hoek van Holland and Harlingen. Other locations, in the surrounding of the Dutch coast, are more prone to increasing trends according to Woth et al. (2006), who used the winds from the a regional climate model to force a surge model. This resulted in small increases of the 99.5th percentiles of surge heights. The increase is found to be significant in the German Bight and along the Danish coast, but not along the Dutch coast. Harlingen is much closer to these regions, which could be the explanation why increasing trends are only found for Harlingen and not for Hoek van Holland. The IPCC (2013) states that long term changes of storm surge activity are in general very region/location dependent. Another possibility is the influence of the Wadden Sea basin. Basins like this are sometimes prone to tidal resonance. Also, the water depth in the Wadden Sea is much smaller than near Hoek van Holland. Nevertheless, the latter two factors do not explain the differences in trends between the two locations, but could have an amplifying effect on the water levels at Harlingen.

6.1.6. Research limitations

There are some limitations of this research identified. Firstly, Horsburgh & Wilson (2007) state that surge affects tide water levels in a similar way as tide affects surge height. However, the tide water level time series provided by the WAQUA model are exactly the same for each of the 16 members per location. This implies that the WAQUA model does not take this process into account. Secondly, as mentioned before, literature points out there are two variables commonly used to analyse storm surge activity, which are the storm surge residual and the skew storm surge (Caires, 2011; Dillingh et al., 1993; Tijssen & Diermanse, 2009).The storm surge residual, in this research referred to as surge height, is used to perform the trend analysis. However, one could argue that the use of skew storm surge provides more accurate results. This is because skew surge (illustrated in Figure 6-4) better separates the astronomical from the meteorological impact on the water level, as, in general, the maximum of the wind-driven part of the water level (surge) does not coincide with the maximum of the astronomically-driven part (tide) (Sterl et al., 2009). Caires (2011) emphasises that in order to carry out an extreme value analysis the data need to be homogeneous (i.e. coming from the same population of climatological and hydraulic circumstances). Conclusive, in this research no correction for the influence of the tide on the surge heights is applied, meaning the used variable is not completely homogeneous. Fortunately, the effect on the trend results is believed to be relatively small, although no quantitative analysis on this phenomenon is conducted. Further limitations of this research would be the exclusion of river run-off effects (Hoek van Holland is situated at the river mouth of the Meuse) and potential large-scale external surges. The latter in this case are surges generated in seas around the North Sea basin which propagate into the North Sea, pushing additional water masses into the basin (Woth et al., 2006). Nevertheless, there is no proof of significant magnitude of external surges and the RACMO and WAQUA model domain extend to reasonable sizes to assume the influence is negligible.

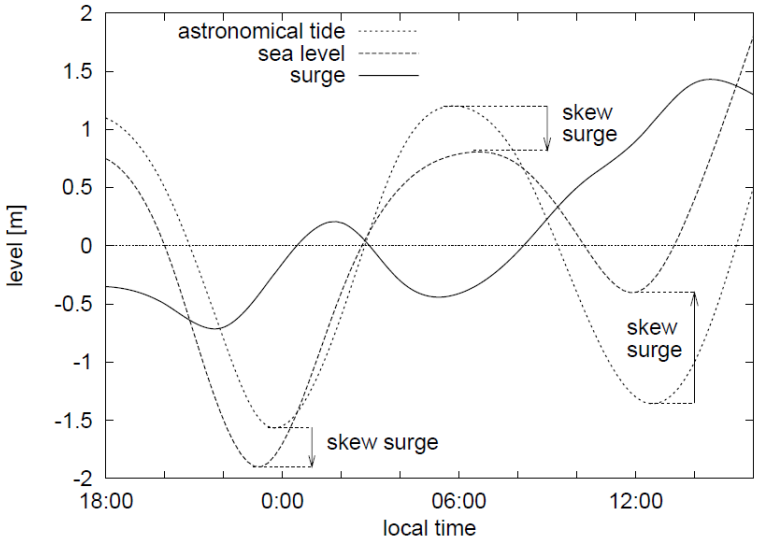

Figure 6-4. The skew surge is the difference between the maximum (minimum) water level and the closest high (low) astronomical tide, which does not need to take place at the same moment (Reproduced from van den Brink et al., 2003, by Sterl et al., 2009).

6.1.7. Impact on the Dutch coast

To provide an indication of what the impact of the results are on the flood safety of the Dutch coast multiple factors need to be considered.

First of all, a stand-alone increase in the numbers of storm surge events criterion would be less relevant for the flood safety of the Dutch coast, but an increase in duration and intensity of the storm surge events corresponding to the lower exceedance frequencies is of greater importance (Woth et al., 2006). Such a result is obtained in this research for the location of Harlingen, where the surge height intensity of the extremes and the overall duration showed significant increasing trends. However, it is a difficult procedure to determine one normative storm surge event, since a storm surge event with a low intensity and long duration could cause similar flood defence failure probabilities as a storm surge event with a high intensity and short duration. Storm surge cluster events are an example of this where flood defence constructions have to cope with the occurrences of high water levels for several days. Hence, there would be a range of normative storm surge events with combinations of intensities and durations corresponding to exceedance frequencies.

Secondly, mean-sea-level rise has a potential influence on storm surge. In a warming climate the sea level will rise due to the thermal expansion of the sea water and the melting of land ice (Sterl et al., 2009). To give an indication, the IPCC (2013) predicts a global mean-sea-level rise for 2081-2100 between 0.26 and 0.82 meters for all representative concentration pathways (RCP's) relative to 1986-2005. De Vries et al. (2014) reported similar values but for the North Sea specifically, namely between 0.25 and 0.80 meters by the year 2085. One could argue that a higher mean-sea-level results in a damped effect on the wind-driven surge, because surge heights are significantly greater at low water levels than at high water levels (Horsburgh & Wilson, 2007). However, Sterl et al. (2009) found that a higher mean-sea-level due to global warming does not impact the height of storm surges (i.e. the surge height values do not change). They performed some test experiments in which the water depth was uniformly increased by 2 meters in the whole domain of WAQUA/DCSM98. This resulted in negligible differences in surge height. This leads to the conclusion that surge heights do not change with a rising mean-sea-level (e.g. Woth et al., 2006). Additionally, uniformly increasing the water depth impacts the tide and surge in the same way with no net effect (Horsburgh & Wilson, 2007).

Thirdly, a final contributor to an impact on the Dutch coast assessment is land subsidence, which is also stressed by the (IPCC, 2013). Koster et al. (2018) addressed the magnitude of

land subsidence in The Netherlands and found potential lowering of the Dutch coastal land surface situated between Amsterdam and Rotterdam varying between 0 and 0.8 meters in 30 years due to human-induced groundwater level lowering. Lower lying hinterland could result in larger numbers of casualties and more damage when flooded, which in turn could lead to an amendment of Dutch flood protection standards. Despite these three factors, an indication of the impact on the Dutch coastline due to only storm surge based on this research can be provided. For Hoek van Holland hardly any changes are expected, since there were almost no significant trends obtained. Only the frequency of storm surge events decreased at this location, for which no physical explanations are found. This indication is considered reasonably representative for the more southern Dutch North Sea coastline, consisting of South-Holland, Zeeland (apart from the estuaries) and part of North-Holland. For Harlingen, the clearest and largest trend was found for the duration of storm surge events, which increases with 2 to 6 hours according to the RACMO and WAQUA combination. This indication is considered reasonably representative for the Wadden Sea coastline of Friesland and the Afsluitdijk.

6.1.8. Future research

Many preceding storm surge studies obtained a similar conclusion as this research (e.g., Alexandersson et al., 1998, 2000; Sterl et al., 2009; The WASA Group, 1998). Expected was that the large data size of WAQUA (2400 years in total) enabled to calculate storm surge events corresponding to relatively low exceedance frequencies with a low statistical uncertainty. However, data size might not be the factor which cancels out uncertainty completely, since natural variability of storm surges is the main bottleneck. Future research towards the uncertainties due to natural variability in a future storm surge climate is therefore a priority. In addition, the models used in this research need to be further analysed. This is concluded from the relatively large differences between measurement data based and WAQUA model water level exceedance frequencies. Therefore, the question arises whether RACMO and WAQUA in combination are able to simulate the more extreme storm surge events. For the calculation of storm surge events with an exceedance frequency of 1/10,000 per year is, based on this research, still recommended to use extrapolation methods, since computation power nowadays is still limited to run this many years. This might change in the coming decades. Finally, in this research an evident difference between surge heights in Hoek van Holland and Harlingen is found. The largest surge heights resulted for Harlingen were significantly larger than for Hoek van Holland. Additionally, the results for Harlingen in general consisted of larger variations in surge heights than for Hoek van Holland. The physical explanations of these findings could be due to resonance and/or water depth, but this has to be further investigated.

6.2. Conclusions

In this research new insights are provided in storm surge trends for the period 1951-2100 at the Dutch coast based on the model combination of RACMO (a Regional Atmospheric Climate Model) and WAQUA (a storm surge model). Two locations along the Dutch coast with very different topographical characteristics are analysed, namely Hoek van Holland and Harlingen. For the purpose of interpreting the resulting trends, the model accuracy was quantified to provide insight in the prediction accuracy of the models. For the same reason physical processes in the WAQUA output data were identified.

The definition of single storm surge events is, apart from the chosen POT threshold, hardly dependent on the implemented time interval threshold making selected events virtually independent. However, storm surge cluster events are highly dependent on the implemented time interval thresholds. Therefore is concluded that analysing trends of storm surge cluster events with a POT method applied on surge height is not fully reliable and perhaps other methods should be sought. The used definition of single storm surge events is considered sufficiently reliable to be used in future research endeavours.

Found was that the model combination resulted in systematically lower exceedance frequencies for the entire range of assessed exceedance frequencies (i.e. 1/0.2 till 1/4000 per year). The water level differences between the model and measurement data based exceedance frequencies becomes larger for smaller exceedance frequencies. This especially holds true for Hoek van Holland. Therefore, it is questionable if this model combination is able to simulate the most extreme storm surge events. Physical processes, such as the various tide cycles were present in the data and seemed a fair representation of reality. Surge-tide interaction was partly identified. There was a clear influence of the tide on the surge heights, but the surge height did not have any effect on the tide water levels. Conclusive, the model combination shows reasonable approximation of reality in physical processes, but also results in systematically lower water levels, especially for the lower exceedance frequencies. Nevertheless, the models show potential to extend current knowledge of future climate extremes implications.

Most of the obtained trends were found to be insignificant. For Hoek van Holland no significant increasing trends were obtained. Only a significant decrease from 1951-1980 to 2071-2100 of the frequency of storm surge events of about 0-3 events per 30 years. For Harlingen mainly an increase in the duration above the 0.5 meter surge height of storm surge events is expected of about 2-6 hours from 1951-1980 to 2071-2100. However, the obtained trends are much smaller than the decadal and inter-decadal variability of storm surge activity, leading to the conclusion that the trends cannot be directly attributed to climate change.

References

Alexandersson, H., Schmith, T., Iden, K., & Tuomenvirta, H. (1998). Long-term variations of the storm climate over NW Europe. *The Global Atmosphere and Ocean System*, *6*(2), 97–120. Retrieved from https://www.researchgate.net/publication/246724763_Long-term_variations_of_the_storm_climate_over_NW_Europe

Alexandersson, Hans, Tuomenvirta, H., Schmith, T., & Iden, K. (2000). Trends of storms in NW Europe derived from an updated pressure data set. *Climate Research*, *14*(1), 71–73. https://doi.org/10.3354/cr014071

Caires, S. (2011). Extreme value analysis: still water level. *JCOMM Technical Report*, (58), 1–33.

Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, *43*(4), 478–484. https://doi.org/10.1198/00401700152672573

Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., … Whetton, P. (2007). Regional Climate Projections. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*.

Claassen, J. N. (2018). *The Occurrence of Multiple High Water Level Days in Rapid Succession due to Storm Clusters : a meteorological impact model perspective for the Dutch coast*. Amsterdam University College.

Cropper, T., Hanna, E., Valente, M. A., & Jónsson, T. (2015). A daily Azores–Iceland North Atlantic Oscillation index back to 1850. *Geoscience Data Journal*, *2*(1), 12–24. https://doi.org/10.1002/gdj3.23

Dangendorf, S., Müller-Navarra, S., Jensen, J., Schenk, F., Wahl, T., & Weisse, R. (2014). North sea storminess from a novel storm surge record since AD 1843. *Journal of Climate*, *27*(10), 3582–3595. https://doi.org/10.1175/JCLI-D-13-00427.1

Dangendorf, S., Wahl, T., Hein, H., Jensen, J., Mai, S., & Mudersbach, C. (2012). Mean sea level variability and influence of the North Atlantic oscillation on long-term trends in the German Bight. *Water (Switzerland)*, *4*(1), 170–195. https://doi.org/10.3390/w4010170

de Vries, H., Katsman, C., & Drijfhout, S. (2014). Constructing scenarios of regional sea level change using global temperature pathways. *Environmental Research Letters*, *9*(11). https://doi.org/10.1088/1748-9326/9/11/115007

Dillingh, D. (2013). *Kenmerkende waarden kustwateren en grote rivieren*.

Dillingh, D., de Haan, L., Helmers, R., Können, G. P., & van Malde, J. (1993). *De basispeilen langs de Nederlandse kust; statistisch onderzoek*. *Rapport DGW-93.023 Deel 1*. Den Haag.

European EC-Earth consortium. (2019). About EC-Earth. Retrieved June 14, 2019, from http://www.ec-earth.org/about/

Feser, F., Barcikowska, M., Krueger, O., Schenk, F., Weisse, R., & Xia, L. (2015). Storminess over the North Atlantic and northwestern Europe - A review. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 350–382. https://doi.org/10.1002/qj.2364

Geerse, C., & Wojciechowska, K. (2014). *Betrouwbaardheidsintervallen voor kwantielen van de overschrijdingsfrequentie: Toepassing op kuststations en IJsselmeergebied*.

Haarsma, R. J., Hazeleger, W., Severijns, C., de Vries, H., Sterl, A., Bintanja, R., … van Den Brink, H. W. (2013). More hurricanes to hit western Europe due to global warming. *Geophysical Research Letters*, *40*(9), 1783–1788. https://doi.org/10.1002/grl.50360

Horsburgh, K. J., & Wilson, C. (2007). Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research: Oceans*, *112*(8). https://doi.org/10.1029/2006JC004033

IPCC. (2013). *Climate Change 2013: The Physical Science Basis*. *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.Summary

Kaye, C. A., & Stuckey, G. W. (1973). Nodal Tidal Cycle of 18.6 Yr .: Its importance in sea-level curves of the east coast of the United States and its value in explaining long-term sea-level changes. *Geology*, *1*(3), 141–145.

KNMI. (n.d.). Project ICOWEX. Impacted by coincident weather extremes: How clustered and compound extreme weather events imprint on our society now and in the future. Retrieved from https://www.knmi.nl/kennis-en-datacentrum/project/icowex

Kok, M., Jongejan, R., Nieuwjaar, M., & Tanczos, I. (2017). *Fundamentals of Flood Protection*. Retrieved from https://www.enwinfo.nl/images/pdf/Grondslagen/GrondslagenEN_lowresspread.pdf

Koster, K., Stafleu, J., & Stouthamer, E. (2018). Differential subsidence in the urbanised coastal-deltaic plain of the Netherlands. *Geologie En Mijnbouw/Netherlands Journal of Geosciences*, *97*(4), 215–227. https://doi.org/10.1017/njg.2018.11

Lenderink, G., van den Hurk, B., van Meijgaard, E., van Ulden, A., & Cuijpers, H. (2003). *Simulation of present-day climate in RACMO2: first results and model developments*.

Mailier, P. J., Stephenson, D. B., & Ferro, C. A. T. (2006). Serial Clustering of Extratropical Cyclones. *Monthly Weather Review*, *134*, 2224–2240.

Mastenbroek, C., Burgers, G., & Janssen, P. A. E. M. (1993). The Dynamical Coupling of a Wave Model and a Storm Surge Model through the Atmospheric Boundary Layer. *Journal of Physical Oceanography*, *23*, 1856–1866. Retrieved from https://journals.ametsoc.org/doi/pdf/10.1175/1520-0485%281993%29023%3C1856%3ATDCOAW%3E2.0.CO%3B2

Philippart, M. E., Dillingh, D., Pwa, S. T., Bavelaar, A. E. R., van der Male, C., Soerdjbali, M., … Storm, W. (1995). *De basispeilen langs de Nederlandse kust: de ruimtelijke verdeling en overschrijdingslijnen*. Den Haag.

Pinto, J. G., Bellenbaum, N., Karremann, M. K., & Della-Marta, P. M. (2013). Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions. *Journal of Geophysical Research Atmospheres*, *118*(22), 12476–12485. https://doi.org/10.1002/2013JD020564

Pinto, J. G., Gómara, I., Masato, G., Dacre, H. F., Woollings, T., & Caballero, R. (2014). Large-scale dynamics associated with clustering of extratropical cyclones affecting western Europe. *Journal of Geophysical Research*, *119*(24), 13,704-13,719. https://doi.org/10.1002/2014JD022305

Priestley, M. D. K., Pinto, J. G., Dacre, H. F., & Shaffrey, L. C. (2017). The role of cyclone clustering during the stormy winter of 2013/2014. *Weather*, *72*(7), 187–192. https://doi.org/10.1002/wea.3025

Ridder, N., de Vries, H., Drijfhout, S., van den Brink, H., van Meijgaard, E., & de Vries, H. (2018). Extreme storm surge modelling in the North Sea: The role of the sea state, forcing frequency and spatial forcing resolution. *Ocean Dynamics*, *68*(2), 255–272. https://doi.org/10.1007/s10236-018-1133-0

Rijkswaterstaat. (2013). *Kenmerkende waarden - Getijgebied 2011.0*. Retrieved from http://publicaties.minienm.nl/documenten/kenmerkende-waarden-getijgebied-2011

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., & Ludwig, C. (2015). The trajectory of the Anthropocene: The Great Acceleration. *Anthropocene Review*, *2*(1), 81–98. https://doi.org/10.1177/2053019614564785

Sterl, A., van den Brink, H., de Vries, H., Haarsma, R., & van Meijgaard, E. (2009). An ensemble study of extreme North Sea storm surges in a changing climate. *Ocean Science Discussions*, *6*(2), 1031–1059. https://doi.org/10.5194/osd-6-1031-2009

Svasek Hydraulics. (n.d.). WAQUA - TRIWAQ. Retrieved May 22, 2019, from http://www.svasek.nl/model-research/waqua/

The Open University. (1999). *Waves, Tides and Shallow-Water Processes* (2nd ed.). Oxford: Elsevier Ltd. https://doi.org/https://doi.org/10.1016/B978-0-08-036372-1.X5000-4

The Scipy community. (2014). scipy.stats.linregress. Retrieved April 26, 2020, from https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.linregress.html

The WASA Group. (1998). Changing Waves and Storms in the Northeast Atlantic ? *Bulletin of the American Meteorological Society*, *79*(5), 741–760.

Tijssen, A., & Diermanse, F. (2009). *Storm surge duration and storm duration at Hoek van Holland*.

Ulbrich, U., Leckebusch, G. C., & Pinto, J. G. (2009). Extra-tropical cyclones in the present and future climate: A review. *Theoretical and Applied Climatology*, *96*(1–2), 117–131. https://doi.org/10.1007/s00704-008-0083-8

van den Adel, N. (2013). Storm in een glas water. Retrieved June 23, 2019, from https://niekvandenadel.nl/storm-in-een-glas-water/

van den Brink, H. W., & Können, G. P. (2011). Estimating 10000-year return values from short time series. *International Journal of Climatology*, *31*(1), 115–126. https://doi.org/10.1002/joc.2047

van den Brink, H. W., Können, G. P., & Opsteegh, J. D. (2003). The Reliability of Extreme Surge Levels , Estimated from Observational Records of Order Hundred Years The reliability of extreme surge levels , estimated from observational records of order hundred years H . W . van den Brink , G . P . K ¨ The research is, (June).

van den Hurk, B., van Meijgaard, E., de Valk, P., van Heeringen, K. J., & Gooijer, J. (2015). Analysis of a compounding surge and precipitation event in the Netherlands. *Environmental Research Letters*, *10*(3). https://doi.org/10.1088/1748-9326/10/3/035001

van der Grinten, R. (2011). *The impact of gustiness on sea surface height in storm surge situations*. University Utrecht.

van Meijgaard, E., Ulft, L. H. Van, Bosveld, F. C., Lenderink, G., & Siebesma, a P. (2008). The KNMI regional atmospheric climate model RACMO version 2.1. *Technical Report; TR - 302*, 43.

van Meijgaard, E., van Ulft, L. H., Lenderink, G., de Roode, S. R., Wipfler, L., Boers, R., & Timmermans, R. M. A. (2012). *Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe*.

Visser, H., & Petersen, A. C. (2012). Inferences on weather extremes and weather-related disasters: A review of statistical methods. *Climate of the Past*, *8*(1), 265–286. https://doi.org/10.5194/cp-8-265-2012

Weisse, R., von Storch, H., Niemeyer, H. D., & Knaack, H. (2012). Changing North Sea storm surge climate: An increasing hazard? *Ocean and Coastal Management*, *68*, 58–68. https://doi.org/10.1016/j.ocecoaman.2011.09.005

Woth, K., Weisse, R., & Storch, H. Von. (2006). Climate change and North Sea storm surge extremes : an ensemble study of storm surge extremes expected in a changed climate projected by four different regional climate models. *Ocean Dy*, (56), 3–15. https://doi.org/10.1007/s10236-005-0024-3

Yan, Z., Tsimplis, M. N., & Woolf, D. (2004). Analysis of the relationship between the North Atlantic oscillation and sea-level changes in northwest Europe. *International Journal of Climatology*, *24*(6), 743–758. https://doi.org/10.1002/joc.1035

Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., & Stephenson, D. B. (2013). A Multimodel Assessment of Future Projections of North Atlantic and European Extratropical Cyclones in the CMIP5 Climate Models. *Journal of Climate*, *26*(16), 5846–5862. https://doi.org/10.1175/JCLI-D-12-00573.1

Zhao, X., Zhang, Z., Cheng, W., & Zhang, P. (2019). A new parameter estimator for the generalized pareto distribution under the peaks over threshold framework. *Mathematics*, *7*(5). https://doi.org/10.3390/math7050406