Extreme Waves in the North Sea

Deriving extreme wave conditions applying Hierarchical Clustering and Non-Stationary Extreme Value Modelling

Master's Thesis

Michiel Willem Smit

Delft University of Technology

TU Delft

Extreme Waves in the North Sea

Deriving extreme wave conditions applying Hierarchical Clustering and Non-Stationary Extreme Value Modelling

by

Michiel Willem Smit

In fulfillment to obtain the degree of Master of Science at Delft University of Technology. To be defended publicly on June 29th, 2022.

Supervisor TU Delft:	Dr. Robert Lanzafame	Supervisor
Graduation Committee:	Dr. Elisa Ragno	Chair
	Dr. Alessandro Antonini	
Institution:	Delft University of Technology	
Faculty:	Civil Engineering & Geosciences, Delft	
Project Duration:	September, 2021 - June, 2022	

Cover Image: Extreme waves hitting the La Jument Lighthouse in France, 1989. Photo taken by Jean Guichard on December 21st, 1989

TU Delft

Disclaimer

This report is written in partial fulfillment to meet the requirements to obtain the Master's degree at the Faculty of Civil Engineering and Geosciences at Delft University of Technology. It has been prepared with great care under the guidance of staff of Delft University of Technology. However, the reader should realise that this report has been prepared for educational purposes only and is judged on educational criteria. As a result, this report should not be considered as a professional consultancy report. Delft University of Technology cannot accept liability for all content of this report.

Abstract

Coastal and offshore infrastructure must be designed to withstand extreme wave-induced loading conditions. Extreme Value Analysis (EVA) is often employed to infer probabilistic distributions that provide information about extreme design conditions. In traditional practices, EVA is performed under the assumption of stationarity. This means that the probability of extreme events is constant in time. However, hydraulic loading conditions are expected to exhibit temporal variability in severity and frequency as a result of climate change. Therefore, the assumption of stationarity becomes questionable. Non-stationary extreme value analysis (NEVA) for inferring extreme hydraulic loads have become more attractive in recent years. However, the applicability of NEVA models is debatable ansd differs on a case-by-case basis. Large scale oceanic bodies can be characterized by spatially and temporally varying extreme wave characteristics. Clustering analyses have proven to be successful to identify regions exhibiting similar extreme wave characteristics. Creating clusters based on similar extreme wave characteristics can potentially improve extreme value modelling because intra-cluster information can be pooled to derive more accurate extreme value models.

This research presents a practical assessment of the applicability of clustering analysis and non-stationary extreme value modeling of extreme wave statistics at cluster level in the North Sea. The primary objectives of this research are: (1) Study the temporal variability extreme significant wave height (H_{m0}) and extreme wind speeds (U_{10}) in the North Sea domain, (2) Investigate how hierarchical clustering analysis (HAC) can be employed to cluster grid points that exhibit similar extreme wave characteristics, (3) How the obtained clusters and temporal variability can be employed to derive extreme value models describing extreme H_{m0} statistics at cluster level and (4) assess whether NEVA models at cluster level form a practical alternative compared to conventional stationary analysis in the design and risk assessment of hydraulic infrastructure in light of climate change.

Temporal trend analysis of H_{m0} in the North Sea showed that the period between 1990 and 2020 can be characterized by a decreasing trend. Between 1950 and 2020, a decrease in H_{m0} intensity is observed in the Western regions and an increase is observed in the East. This is reason to believe that the variability in extreme wave climate is cyclical rather than monotonic. There is reason to believe that temporal variations of extreme U_{10} are responsible for the temporal variability of extreme H_{m0}. Initial clustering results partition the North Sea domain into 50 clusters based on characteristic values for the significant wave height (H_{m0}), peak period (T_p), and dominant wave directions (θ_1 and θ_2). After splitting clusters based on geo-location and merging clusters based on the intra-cluster statistical properties of the wave parameters, 63 clusters are obtained. The identified clusters and temporal variability are used to define NEVA models describing extreme H_{m0} statistics at cluster level. Intra-cluster H_{m0} observations are detrended before fitting the GEV parameters by means of Bayesian Inference. Informative priors are constructed by pooling the GEV parameter information from the intra-cluster grid points. Potential non-stationarity is accounted for by adding the Theil-Sen parameters (b and b_0) to the location parameter (μ^*), making the location parameter a linear function of time. The model parameters subsequently read: $H_{m0} \sim GEV\left(\mu^* + (b \cdot t + b_0), \sigma^*, \xi^*\right)$. Using the extreme H_{m0} data from the clustering centroid yields the most promising results for describing extreme H_{m0} statistics at cluster level under the condition that the intra-cluster exhibits homogeneous values for b and b_0.

The applied HAC analysis presented in this research is not the optimal strategy. The identified clusters exhibit heterogeneous values for b and b_0 Because non-stationarity of H_{m0} was not accounted for during the HAC analysis. This hinders the performance of the NEVA models at cluster level. Also, whether the derived methodology can be applied for the long-term projection of future extreme wave events in the North Sea is debatable. The non-stationary of extreme H_{m0} is best described by a cyclic pattern. Without a thorough understanding of the underlying causes of the non-stationary in H_{m0} and without future projections of the extreme wave climate, the applicability of NEVA for deriving extreme H_{m0} design conditions in light of climate change cannot be guaranteed.

Preface

In this Report, I present the findings of my research into the derivation of an updated methodology for deriving extreme wave conditions in the North Sea. This Master's thesis is written in partial fulfillment to meet the requirements to obtain the degree of Master of Science in Hydraulic Engineering, with a combined specialization in Coastal Engineering and Flood Risk at Delft University of Technology. Overseeing this research were Dr. Elisa Ragno, Dr. Alessandro Antonini, and Dr. Robert L. Lanzafame.

This thesis is the conclusion of my time as a student in Delft. Apart from studying, I have spent much of my time developing other skills, for example by obtaining my Yachtmaster degree. Over the last few years, I have spent much of my summer holidays sailing the Ionian Seas. The sound of the waves hitting the hull of the ship and the wind blowing through the sails helped me relax during my time away from my studies. However, after being caught in choppy weather more often than I would like to admit, I am well aware of the dangers of the seas. My main motivation for pursuing a master's degree in hydraulic engineering is to apply my skills in the climate adaptation of global flood safety. Combining my interests in the dangers of the sea and climate adaptation in a master's thesis came natural when preparations for this thesis started in the spring of 2021. My graduation thesis provided many lessons, not only in the academic field. When I started my thesis, little did I know about the challenging path I had to follow to complete my thesis. I have not made it a secret that I have found my thesis the most challenging part of my time in Delft. In the end. this thesis taught me a lot about how to tackle large-scale projects, planning, and organization, as well as how to be more decisive when deciding which tasks are necessary and which are not.

Of course, this thesis would not have been possible without the support of several people. I want to start by thanking my supervisors. First and foremost, a big thanks to the chair of my committee, Elisa. Thank you for making time for our weekly meetings which provided many useful insights as to which methodologies to explore in an attempt to solve the problems I encountered during my thesis. I have learned a lot about the field surrounding extreme value modeling, and I am sure this knowledge will be extremely valuable in my career as a civil engineer. To Robert, I want to say thanks for helping me find a research topic and a graduation committee, and in particular helping me shape the green light presentation. An extra pair of eyes taking a critical look at my presentation was massively helpful at the time. And finally, I want to thank Alessandro. Despite your challenging questions, I really appreciated your input during this research. Your knowledge surrounding extreme value analysis and your lighthouse papers have helped a lot while shaping my thesis. Of course, I want to thank to my parents, who offered their unconditional support. Not only over the course of this thesis, but for all the years leading up to this moment. I want to finish by thanking all my friends, my brother and my sister. Whether it be during a coffee break on campus, an ice-cold beer on a Friday afternoon, or a Sunday afternoon hockey match, I am very grateful to have such a large group of friends and family around me that provided continuous and entertaining distractions away from my thesis. Even when you almost quite literally had to pull me away from my desk. I want to thank one friend in particular: my housemate, Tom. In February, you noticed that I was having trouble with my thesis and helped me resolving the issues I was having. These insights have helped me get back on track when my thesis became very difficult.

I am delighted to present my thesis. I can confidently say that I have delivered a great product that is worthy conclusion to my master's education. After my graduation, I will leave for South America for two months. After that, my career as a civil engineer will start and I am looking forward to applying the knowledge and skills I gained over the past eight years in a new and challenging environment.

Michiel Willem Smit
Rotterdam, June 2022

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
AO	Arctic Oscillation
BI	Bayesian Inference
BM	Block Maxima
CB	Confidence Bounds
CDF	Cumulative Distribution Function
CH	Caliniski Harabasz
CI	Confidence Interval
DB	Davies-Bouldin
DI	Dispersion Index
DTL	Declustering Time Lag
ECMWF	European Center for Medium Range Weather Forecasts
EI	Extremal Index
ERA5	ECMWF Reanalysis 5
EVA	Extreme Value Analysis
GEV	Generalized Extreme Value (Distribution)
GOF	Goodness of Fit
GP	Generalized Pareto (Distribution)
HAC	Hierarchical Agglomerative Clustering
HE-MCMC	Hybrid Evolution Markov Chain Monte Carlo
IPCC	Intergovernmental Panel on Climate Change
JONSWAP	Joint North Sea Wave Project
KNMI	Koninklijk Nederlands Meteorologisch Instituut (Royal Netherlands Meteorological Institute)
MBE	Mean Bias Error (Bias)
MCMC	Markov Chain Monte Carlo
MK	Mann-Kendall (test)
MRL	Mean Residual Life
MWD	Mean Wave Direction (as referred to in ERA5 Reanalysis)
MWP	Mean Wave Period (as referred to in ERA5 Reanalysis)
NAO	North-Atlantic Oscillation
NEVA	Non-stationary Extreme Value Analysis
N-GEV	Non-stationary Generalized Extreme Value (Distribution)
N-GP	Non-stationary Generalized Pareto (Distribution)
PDF	Probability Distribution Function
POT	Peak over Treshold
PP1D	Peak Wave Period (as referred to in ERA5 Reanalysis)
PP-plot	Probability-Probability plots
ProNEVA	Process Informed Non-Stationary Extreme Value Analysis
QQ-plot	Quantile-Quantile plots
RL	Return Level
RMSE	Root Mean Square Error
RP	Return Period
RWS	Rijkswaterstaat
SC	Silhouette Coefficient
SOM	Self-organizing maps

Abbreviation	Definition
SWH	Significant Wave Height (as referred to in ERA5 Reanalysis)
TS	Theil Sen (test)
U10	Eastward component of 10m elevation windspeed
V10	Northward component of 10m elevation windspeed
WCSS	Within Cluster Sum of Squares

Symbols

General Wave Symbols

Symbol	Definition	Unit
$E(f,\theta)$	Two dimensional wave energy spectrum	-
E_{tot}	Total Energy stored in wave field	-
H^*_{m0}	Detrended observations for extreme significant wave height	m
H_{m0}	Significant Wave Height	m
L	Wave length	m
L_p	Wave length corresponding to peak wave period	m
m_n	Spectral moment of wave energy spectrum	-
s	Wave steepness	-
$S(f)$	One dimensional wave energy spectrum	-
T	Wave Period	s
$T_{m-1,0}$	Mean wave period as derived from wave energy spectrum	s
T_P	Peak wave Period	s
u_{10}	Eastward component of 10m elevation windspeed	m/s
U_{10}	Absolute magnitude of windspeed	m/s
v_{10}	Northward component of 10m elevation windspeed	m/s
ϕ	Mean wind direction defined as the direction that waves are coming from with respect to true north	^oN
ρ	Water Density	kg/m^3
θ	Mean wave direction defined as the direction that waves are coming from with respect to true north	^oN

Hierarchical Clustering Analysis

Symbol	Definition	Unit
$a(i)$	mean dissimilarity between object i and all other objects included in the same cluster $\mathcal{I}$	-
$b(i)$	smallest mean dissimilarity of object i to all points in any other cluster than $\mathcal{I}$	-
C	Cophenetic Correlation	-
CH	Calinski-Harabasz Score	-
$d(i,j)$	Euclidean distance between grid points i and j	-
$d_{\mathcal{I}}$	Average dissimilarity between all points included in cluster $\mathcal{I}$	-
$D_{\mathcal{I},\mathcal{J}}$	within-to-between cluster distance ratio for the clusters $\mathcal{I}$ and $\mathcal{J}$ used to calculate the Davies Bouldin Ratio	-
$d_{\mathcal{I},\mathcal{J}}$	Euclidean distance between the centroids of clusters $\mathcal{I}$ and $\mathcal{J}$.	-
DB	Davies Bouldin Ratio	-
$\mathcal{I}$	Cluster of interest	-
i,j	grid points	-
$\mathcal{J}$	Cluster that is compared to cluster $\mathcal{I}$	-
$\mathcal{K}$	Number of clusters	-

Symbol	Definition	Unit
$\mathcal{K}_{CH}$	Ideal number of clusters as determined by the Calinski-Harabasz Score	-
$\mathcal{K}_{DB}$	Ideal number of clusters as determined by the Davies Bouldin Ratio	-
$\mathcal{K}_{max}$	Maximum number of clusters to be considered by the Hierarchical Clustering algorithm	-
$\mathcal{K}_{SC}$	Ideal number of clusters as determined by the Silhouette Coefficient	-
$\mathcal{M}$	Mean of all input data	-
$\mathcal{M}_{\mathcal{I}}$	Mean of clustering centroid	-
N	number of grid points considered in clustering analysis	-
$N_{\mathcal{I}}$	Number of grid points in cluster $\mathcal{I}$	-
$s(i)$	Silhouette of grid point i	-
SC	Silhouette coefficient	-
$T(i,j)$	Cophenetic dissimilarity between grid points i and j. Defined as the elevation in the dendrogram where grid points i and j are first connected	-
$T(\mathcal{I},\mathcal{J})$	Cophenetic dissimilarity between clusters $\mathcal{I}$ and $\mathcal{J}$. Determines the elevation in the dendrogram at which clusters $\mathcal{I}$ and $\mathcal{J}$ are connected.	-
X_i^*	Normalized value of parameter X at grid point i	-

Extreme Value Analysis

Symbol	Definition	Unit
b	TS estimator, indicates the slope of the linear trend as calculated by the TS test	m/yr
b_0	TS estimator of the y-intercept at the start of the considered temporal horizon	m
N_{exc}	Number of annual threshold exceedances	exc/yr
T_{dur}	Duration of threshold exceedance, defined as the difference between end and start of threshold exceedance	days
t	Time in years	yr
u	Threshold for Peak over Threshold Analysis	m
y_e	Difference between value of extreme observation and threshold value u	m
Y_{obs}	Set of observations included in Bayesian Inference	m
$y_{Hm0}(t)$	Linear temporal trend for H_{m0}	m/yr
δ	Declustering time lag for POT analysis	days
μ	Location parameter of the GEV distribution	-
mu^*	Location parameter of the GEV distribution corresponding to detrended H_{m0} data	-
Ψ_{GEV}	Cumulative Density Function of the GEV distribution	-
Ψ_{GP}	Cumulative Density Function of the GP distribution	-
σ	Scale parameter of the GEV and GP distribution	-
σ^*	Scale parameter of the GEV distribution corresponding to detrended H_{m0} data	-
θ	Set of GEV distribution parameter, includes μ, σ and ξ	-
θ^*	Set of GEV distribution parameters corresponding to H_{m0}^*, includes mu^*, $sigma^*$ and xi^*	-
ξ	Shape parameter of the GEV and GP distribution	-

Symbol	Definition	Unit
ξ^*	Shape parameter of the GEV distribution corresponding to detrended detrended H_{m0} data	-

List of Figures

List of Tables

1

Introduction

Coastal and offshore infrastructure such as flood defenses (dunes and dikes), breakwaters, offshore wind turbines, oil and gas platforms and lighthouses must be designed to withstand extreme environmental loading conditions. Physical processes such as extreme wave events play an important role in the design and risk assessment of said hydraulic infrastructure. In coastal regions, wave-induced erosion of dunes and overtopping of dikes increase flood hazards (Jonkman et al., 2018). In offshore regions, extreme waves endanger the structural integrity of offshore wind turbines, oil and gas platforms (Jiang et al., 2017), and lighthouses (Antonini et al., 2019; Raby et al., 2019).

The design and risk assessment of hydraulic infrastructure require a comprehensive understanding of the extreme wave climate. This includes the incoming significant wave height, (peak) wave period, and wave direction with a corresponding return period. The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) states that since 1960, anthropogenic activities have had a significant effect on the climate and corresponding weather effects (Masson-Delmotte et al., 2021). As a result, weather and climate extremes are expected to increase in severity and frequency and have changing spatial patterns (Allen et al., 2012). The increased intensity of storms and subsequent extreme wave events will have a direct effect on the design loads of hydraulic infrastructure making the safety of hydraulic infrastructure uncertain, both in coastal and offshore regions.

In designs and risk assessments, Extreme Value Analysis (EVA) is often employed to infer probabilistic distributions that can be used to obtain design values for the hydraulic variables corresponding to the return period of interest. In traditional practices, the parameters of the extreme value distributions are calculated under the assumption of stationarity (i.e. time-invariant parameters). This is a reasonable assumption if there are no significant changes in the magnitude or frequency of the extreme events over time. Under the assumption of a stationary climate, the concept of return period provides a one-to-one relationship between the magnitude and the frequency of extreme loading conditions. However, in light of climate change, the assumption of stationarity becomes questionable.

In recent years, several studies promoted to move away from stationary models. Non-stationary probabilistic models to infer extreme hydraulic loads are becoming increasingly popular in the design and risk assessment of hydraulic infrastructure (Cheng et al., 2014; Ragno et al., 2019). In a non-stationary model the parameters of the extreme value distribution function change over time or in response to a physical process (Cheng et al., 2014; Ragno et al., 2019). Non-Stationary Extreme Value Analysis (NEVA) potentially forms a useful alternative to study extreme loading conditions in light of climate change.

The applicability of non-stationary models for the design of hydraulic infrastructure is topic for debate with successful and lesser successful attempts. Lowe et al. (2021) applied NEVA to study extreme water levels along the West Coast of Australia. De Leo et al. (2020) andDe Leo et al. (2021) studied the non-stationary of the significant wave height in the Mediterranean Sea and performed a critical assessment whether NEVA models provide a useful alternative for inference of design values com-

pared to conventional stationary modelling. Wang et al. (2021) performed a similar analysis for the South China Sea. De Leo et al. (2021) and Wang et al. (2021) stated that NEVA provides advantages over conventional EVA only when a significant trend is present in the data. Antonini et al. (2019) and Raby et al. (2019) applied both stationary and non-stationary extreme value models to derive extreme wave conditions in the safety assessment of rock-mounted lighthouses. Luke et al. (2017) assessed the applicability of updated stationary models (AST) to compute design conditions for extreme river discharges in watersheds that underwent physical changes. A more comprehensive analysis of the different studies mentioned above is presented in Chapter 2.

To our knowledge, a study regarding non-stationary extreme wave models in the North Sea domain has not yet been performed and forms the basis of this research. Hydraulic infrastructure in the North Sea includes offshore oil & gas platforms and wind turbines that are critical for the energy supply of surrounding counties. Furthermore, flood defense structures like dunes, dikes, and dams can be found along the coasts of the low-lying countries that surround the North Sea domain. NEVA is potentially useful when accounting for increasing extreme wave conditions during design and risk assessment of hydraulic infrastructure. However, because the applicability of NEVA is case-dependent, a critical assessment of the applicability of NEVA in design and risk assessment of hydraulic infrastructure in the North Sea is required.

Extreme wave parameters such as the significant wave height, peak wave period, and dominant wave directions are influenced by, among others, the wind and wave direction, the available fetch for wave growth and dispersion the and the available depth. It is expected that the extreme wave characteristics vary throughout the North Sea domain, as a result of the extent of the North Sea basin and the spatially varying physical properties described above. Furthermore, extreme wave events in different regions may be the result of different storms. To study the extreme wave climate, this research applies a gridded dataset consisting of 1200 time-series for wind and wave parameters throughout the North Sea domain. A complete analysis of the extreme wave statistics at all grid points is too computationally expensive.

Clustering analysis can be applied to partition a gridded dataset into a select number of clusters. The intra-cluster grid points exhibit similar characteristics. Clustering analysis concerning the extreme wave climate in the North Sea potentially provides useful information regarding regions with similar extreme wave characteristics. This provides useful information about extreme wave loading conditions for the design and risk assessment of hydraulic infrastructure. Grid points with similar extreme wave characteristics can pool data to produce more accurate extreme value models describing extreme wave loading conditions. Moreover, clustering analysis can help to reduce the data volume required to study the North Sea basin's extreme wave climate. Potentially, one representative time-series per cluster needs to be studied to get a comprehensive understanding of the extreme wave statistics throughout the North Sea domain.

Several studies have employed clustering analysis to partition ocean basins into clusters that exhibit similar extreme wave characteristics. Sartini et al. (2018) and Weiss et al. (2014) applied regional frequency analysis (RFA) on the Eastern regions of the North Atlantic Ocean to identify regions with similar storm footprints. Within the identified regions, the extreme wave climate can subsequently be analyzed using univariate techniques (Sartini et al., 2018). Goharnejad et al. (2022) employed Self Organizing Maps (SOM) to partition the Northern Atlantic Ocean into clusters that exhibit similar extreme statistics of significant wave height, period, and direction. They did so using complete time-series of the aforementioned parameters at the grid points. Subsequently, the intra-cluster extreme wave characteristics can be described using time-series of the clustering centroid. A more comprehensive analysis of the different studies mentioned above is presented in Chapter 2.

In this research, hierarchical clustering analysis will be performed to define characteristic regions throughout the North Sea domain that exhibit similar extreme wave characteristics. Subsequently, extreme value analysis is coupled to the clustering output to study the applicability of representative extreme value models that describe the extreme wave statistics at cluster level. Non-stationary components are added to the extreme value models to study the evolution of extreme wave loading conditions

in light of a changing climate.

1.1. Research Objectives

This research studies the applicability of clustering analysis and subsequent non-stationary extreme value modelling at cluster level. This is done to improve the methodology in which extreme wave conditions are derived for the design and risk assessment of hydraulic infrastructure in light of a changing climate. The main research objective can be covered by the following research question:

How can clustering analysis and non-stationary extreme value models at cluster level be used to characterize the statistics of extreme H_{m0} events in the North Sea and when do non-stationary extreme value models form a practical alternative compared to traditional stationary models for the design and risk assessment of hydraulic infrastructure?

To answer this question, the research question is split up into three sub-research questions:

Sub-research question 1: What are the observed changes over time of the extreme H_{m0} events in the North Sea basin and what are the underlying causes?

Temporal trend analysis will be used to better understand the non-stationary behavior of the extreme wave climate in the North Sea area. Whether NEVA provides a useful alternative to conventional stationary modelling depends on whether a statistically significant trend is present in the data. Also, because this research concerns temporal variability in light of climate change, the underlying causes of non-stationarity in the extreme wave climate must be addressed. Extreme value models are developed at the cluster level. Therefore, the non-stationarity of extreme H_{m0} must be investigated at cluster level. To answer this research question, the following tasks will be performed:

1. Analyze the presence and magnitude of statistically significant trends of the extreme H_{m0} events in the North Sea on a basin scale through Mann-Kendall (MK) and Theil-Sen (TS) tests using several extreme value selection methods and temporal horizons;
2. Analyze the presence and magnitude of statistically significant trends in extreme wind speeds (U_{10} to determine whether they provide an explanation for the observed temporal variability of extreme H_{m0};
3. Analyze the presence and magnitude of statistically significant trends of extreme H_{m0} at cluster level by accounting for the intra-cluster dominant wave directions. Subsequently, compare the results at cluster-level with the results at basin scale.

Sub-research question 2: How can Hierarchical Agglomerative Clustering (HAC) be employed to identify regions in the North Sea that exhibit similar extreme wave characteristics?

To study the extreme wave statistics in different regions of the North Sea, it is vital to have information about the extreme wave characteristics throughout the North Sea basin. Hierarchical Agglomerative Cluster (HAC) analysis will be performed to partition grid-points with similar extreme wave characteristics into clusters. To this end, the following research objectives are identified:

1. Identify important wave parameters and representative values that can be used as input for the HAC analysis;
2. Explore parameter combinations and weight configurations and study their effect on the HAC analysis by. Evaluate several internal evaluation metrics to find the best clustering output;;

3. Validate the clustering output for the different configurations by performing a careful comparison between the clustering output and the HAC input values;
4. Further development of cluster map by means of splitting and merging clusters if there is enough physical and statistical evidence of the extreme wave parameters.

Sub-research question 3: How can the derived clustering output and temporal variability be applied to derive representative non-stationary statistical models that describe extreme H_{m0} at cluster level?

The main goal of the cluster analysis is to identify regions that exhibit similar extreme wave characteristics. Extreme value models are derived that are describe the extreme wave climate at cluster level. Also, a method must be found to include potential non-stationarity. To this end, the following tasks will be performed:

1. Detrend extreme H_{m0} data and form representative time-series for extreme H_{m0} at cluster level using different aggregation and selection strategies;
2. Perform stationary extreme value analysis on the representative data by means of Bayesian Inference (BI) techniques implemented in the ProNEVA toolbox (Ragno et al., 2019);
3. Evaluate how the temporal trends for extreme H_{m0} can be included in the extreme value models to account for potential non-stationarity of extreme H_{m0};
4. Assess the performance of the derived non-stationary extreme value model using several assessment methods, including QQ-plots, RL-plots and goodness-of-fit (GOF) tests;
5. Determine which representative time-series yields the best results with respect to its ability to describe the extreme wave climate on a cluster level.

Sub-research question 4: Do representative extreme value models at cluster level form a useful alternative compared to stationary modelling at grid point-level in the design and risk assessment of hydraulic infrastructure in light of a changing climate?

Whether non-stationary extreme value modelling is applicable is debatable and often differs on a case-by-case basis. The applicability of NEVA models at cluster level to derive extreme wave conditions in the North Sea in light of climate change must be assessed. To this end, a critical comparison between NEVA models at cluster level and conventional stationary modelling is required. Furthermore, before using NEVA in light of climate change, a thorough review of the consideration that must be taken into account is essential. To answer this research question, the following tasks are performed:

1. Provide a practical example where typical extreme H_{m0} return derived using both NEVA models at cluster level and conventional stationary models at grid-point level are compared;
2. Explore different considerations to be accounted for before using non-stationary modelling over conventional stationary modelling.
3. Assess if and when non-stationary extreme value models at cluster level form a useful alternative compared to conventional stationary models at grid-point level.

1.2. Research Limitations

Despite the magnitude of this research, some things are left outside the scope. This section presents a list of limitations relevant to this research:

- This research only considers the significant wave height (H_{m0}) to study non-stationarity in the extreme wave climate and to assess the applicability of non-stationary extreme value models in the design and risk assessment of hydraulic infrastructure. Other parameters typically used in the design of Hydraulic infrastructure, such as the peak wave period and dominant wave direction are not accounted for in the temporal trend analysis and extreme value analysis;
- This research considers historical data only. Future climate projections are not studied to make projections about the extreme wave climate in the North Sea. The research aims to arrive at conclusions surrounding the applicability of non-stationary extreme value models based on historical observations only;
- To study the underlying causes of non-stationarity in extreme H_{m0}, non-stationarity in extreme wind speeds (U_{10}) are evaluated. It is assumed that all observed changes in time for both H_{m0} and U_{10} can be attributed to climate change. A further attribution study of climatic drivers to better understand temporal variability of H_{m0} and U_{10} in light of climate change is not performed;
- Local water levels are not considered. The North Sea is a relatively shallow basin, meaning that the extreme waves are not always in deep water conditions. Bottom effects play a significant role in wave development in the North Sea because higher water depths can accommodate higher waves. Climate change induced sea level rise can therefore have a significant effect on the temporal development of the extreme wave climate in the North Sea. However, the effect of sea level is not considered in this research.

1.3. Document Structure

This report starts with a literature review (Chapter 2) presenting all necessary background information relevant to this research. This includes a description of the North Sea storm climate generating extreme wave events and several studies that performed non-stationary extreme value analysis to derive hydraulic boundary conditions in light of a changing climate and several studies that applied clustering analysis of large scale oceanic bodies with respect to extreme waves. Chapter 3 presents the research methodology applied in this research. This includes all information required to understand extreme value selection, temporal trend testing, hierarchical clustering, and (non-stationary) extreme value analysis and how they were applied in this research.

Chapter 4 provides an exploration of the North Sea domain. This includes a description of the applied ERA 5 reanalysis dataset that provides time-series for waves and wind parameters in the research domain. Furthermore, temporal trend analysis is performed on extreme H_{m0} and U_{10} observations to get a better understanding of the temporal non-stationarity of these parameters. Next, in Chapter 5, the results of the clustering analysis and extreme value analysis are presented in the North Sea domain are presented. This includes the evaluation of the Hierarchical Agglomerative Clustering analysis, a temporal trend analysis to better understand the non-stationarity of H_{m0} and U_{10} at cluster level and the derivation of non-stationary extreme value models at cluster level. The applicability of non-stationary extreme value models at cluster level in light of climate change is evaluated in Chapter 6. This includes a practical example where the non-stationary extreme value models at cluster level are compared with conventional stationary models in their ability to derive typical design values for hydraulic infrastructure. Also, the limitations surrounding the use of non-stationary models in light of climate change are discussed.

Finally, the results are reviewed and put into context by examining the research's assumptions and limitations in the Discussion. Finally, the research questions are answered in the Conclusion.

2

Literature Review

This chapter presents a literature review about the topics relevant to this research. An overview the characteristic storm climate in the North Sea resulting in extreme sea states is given in Section 2.1. This study explores the applicability non-stationary extreme value analysis (NEVA) to derive design conditions for hydraulic infrastructure in light of a changing climate. Section 2.2 explores previous studies at explored the applicability of NEVA to derive hydraulic design conditions. hierarchical Clustering Analysis is applied in this research to identify spatial clusters that exhibit similar extreme wave characteristics. Section 2.3 presents several studies that performed clustering analysis to cluster large scale oceanic bodies based on wave characteristics.

2.1. North Sea Storm Activity

The North Sea is located in the Northern mid-latitudes (between 50 and 60 degrees North of the equator) and therefore under the influence of the prevailing West Wind Belt, i.e. the dominant west-to-east wind motion in the northern mid-latitudes. Dominant weather phenomena resulting in storms in this area are the so-called extra-tropical cyclones and the less frequent polar vortexes (Weisse et al., 2012). Extra tropical cyclones and polar vortexes develop at the interface between colder and warmer air in the mid-latitudes and Arctic regions respectively. Both phenomena are characterized by eastward migrating atmospheric low pressure disturbances, often in combination with large pressure gradients resulting in high wind speeds (Weisse et al., 2012). Extra tropical cyclones tend to occur within and propagate along regionally confined areas, so-called storm tracks. The North Sea is under the influence of the North-Atlantic storm track.

In short, extreme wave events in the North Sea domain are caused by cyclonic low pressure areas resulting in anti-clockwise wind fields travelling in Eastward direction (Weisse et al., 2012). Extra tropical cyclones may be in order of 1000 km across. The location of the low-pressure center determines the prevailing wind directions (See also Figure 2.1) and subsequently the wave development. How the location of the low-pressure area influences the development of the extreme wave development is illustrated in Figure 2.1. Here, two historic storm events are illustrated, each with an opposite wind and wave direction in the North Sea. The illustrated storms correspond to the 1953 flood disaster and storm Deirdre in December 2018, respectively. Both storms are examples of extratropical windstorms. Looking at the 1953 storm, it can be seen that the low pressure center of the extratropical cyclone lies above Denmark (Weisse et al., 2012), resulting in Northerly winds and corresponding waves that travel and grow in Southward direction. Contrarily, for the 2018 storm, it can be seen that the low pressure center lies above the Irish Sea (UK Meteorological office, 2020). This results in extreme southerly winds in the North Sea domain, resulting in waves travelling and gaining energy in Northward Direction.

Weisse et al. (2012) and Huthnance (1989) state that the storm activity in The North Sea is not constant and is subjected to consistent changes in both frequency and intensity on inter-annual and decadal time scales. In recent history, observations show that storm activity was relatively low in the 1960's. A strong increase in storm activity towards the mid 1990's can be observed after that. After the 1990's, another

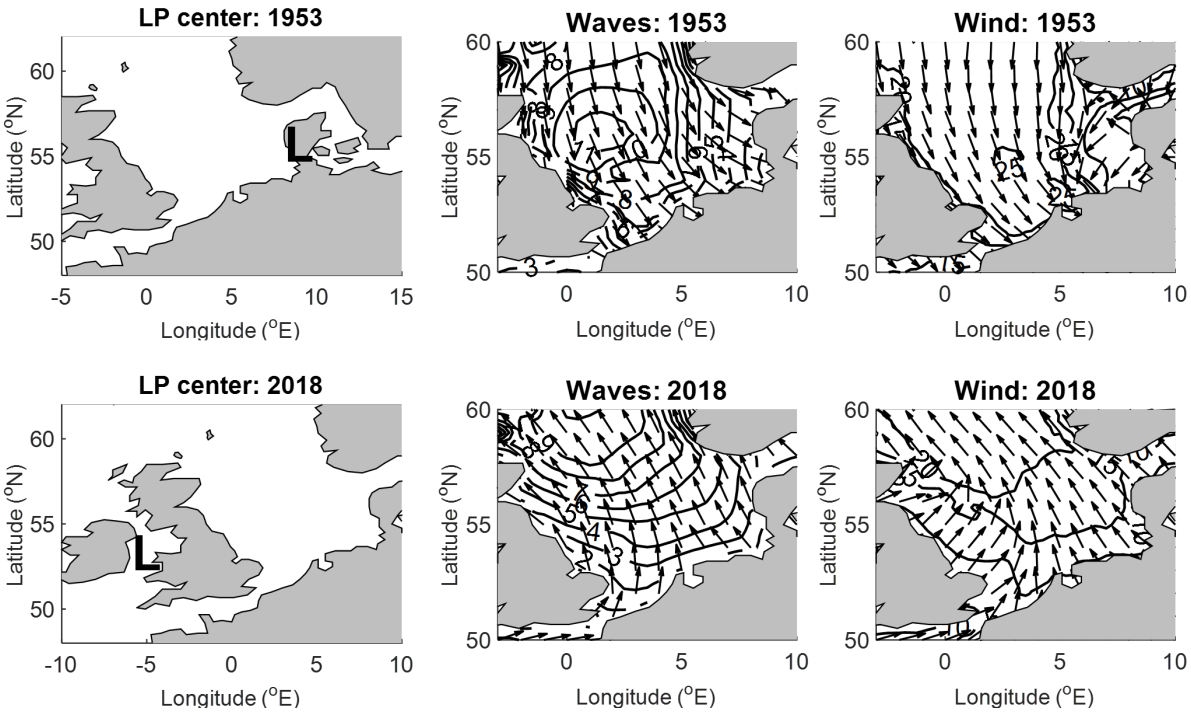


Figure 2.1: Illustration showing the extreme wave climate during two extratropical cyclones in the North Sea. The left figures show the location of the low pressure center during the storm peak (defined by the highest observed wind speeds during the storm). Middle figures show contours of the significant wave height and the mean wave direction during the peak of the storm. Right figures show the contours of the wind speed and the wind direction during peak of the storm.

decrease in storm activity can be inferred (Weisse et al., 2012). Other studies also observed significant spatial movement of North Sea storm activity. Most notably, McCabe et al. (2001) showed that, the following spatial development can be observed based on reanalysis data. Storm frequency in the mid-latitudes (between 30 and 60 degrees N) decreased in the second half of the twentieth century, while an increase in the Northern latitudes (60 degrees N and above) can be observed. Storm intensity also increased in the Northern latitudes, while they remained nearly constant for the mid latitudes.

Studies have shown that the variability in storm activity in North-Western Europe can partially be explained by the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO) (Weisse et al., 2012). The NAO is an atmospheric circulation mode driven by changes in the atmospheric pressure gradient between the Azores and the North Atlantic region. Through fluctuations of the Icelandic low pressures and the Azores high pressures, it controls the strength and direction of westerly winds and the location of storm tracks across the North Atlantic (NOAA, 2022). The Arctic Oscillation (AO) is characterized by pressure anomalies between the arctic regions (low pressure areas), and areas South of 55 degrees latitude. Centres of action can be found near Greenland and the Azores (Baldwin, 2001). The Arctic Oscillation (AO) is closely related to the NAO, and it is often said that the NAO is a more regional expression of the AO. Thee AO has a larger spatial scale whilst the NAO is constricted to the North Atlantic Ocean (Baldwin, 2001). Large pressure difference for the NAO and AO (expressed as a positive NAO and AO index respectively) leads to increased westerly wind speeds and storm tracks that are located in the Northern regions of the mid latitudes. This typically results more frequent and more intense storms in the North Sea region. In contrast, if the pressure difference is low (indicated by a negative index), westerlies are suppressed and the storms track southwards toward the Mediterranean Sea (NOAA, 2022).

Appendix A explains that extreme wave fields might consist of locally generated wind waves, swell waves or a combination of the two. For the design and risk assessment of hydraulic infrastructure, it is critical to understand which wave type is dominant during extreme wave events. 7 location throughout the North Sea are studied to determine the dominant wave type during extreme wave events. (See

Appendix D for more information). Information about the significant wave height (H_{m0}) and the peak wave period T_p is plotted on wave steepness plots. The most extreme wave events are of special interest because these are normative during design and risk assessment of hydraulic infrastructure. The steepness plots show that the extreme waves have a steepness values between $s = 0.03$ and $s = 0.05$, which typically correspond to locally generated wind waves. Based on this analysis, it is assumed that the extreme wave climate in all regions of the North Sea is dominated by locally generated wind waves.

2.2. Non-Stationary Extreme Value Analysis

Most risk assessments of hydraulic infrastructure are performed under the assumption of stationarity. However, as a result of climate change and subsequent evolution of the intensity and frequency of extreme events (Allen et al., 2012), the assumption of stationarity becomes debatable. In recent years, non-stationary probabilistic models to infer extreme hydraulic loads are becoming increasingly popular in the design and risk assessment of hydraulic infrastructure (Cheng et al., 2014; Ragno et al., 2019). However, the applicability of non-stationary models for the design and risk assessment of hydraulic infrastructure is debatable and recent studies present successful and lesser successful attempts. This section presents several studies that have evaluated the applicability of non-stationary models to derive extreme hydraulic loading conditions.

Lowe et al. (2021) used a non-stationary extreme value analysis to study the extreme sea-level variability and to study how individual sea level contributors interact to trigger extreme sea-level events at several locations along the West coast of Australia. To better understand past events as well as make projections for future extreme water level events, the behavior of the individual contributors must be better understood. The individual contributors were subjected to a thorough analysis to study their significance towards extreme water level events in light of past events to better understand their non-stationary behavior. Besides temporal covariates, several physical processes were identified that affected extreme water level distributions (e.g. El Niño and La Niña events). The significance of the covariates differed on a spatial scale, emphasizing the need to study the necessity of a non-stationary study for the different hydrodynamic processes on a case by case basis (Lowe et al., 2021).

De Leo et al. (2020) studied long term non-stationarity of extreme significant wave height (H_{m0}) events in the domain of the Mediterranean Sea. Furthermore, they studied the applicability of several methods to detect the presence of statistically significant trends, including the Mann-Kendall (MK) test and the Theil-Sen (TS) test. De Leo et al. (2020) found that there is a strong anti-correlation between the TS estimator (b, i.e. slope of linear trend) and the p-value of the MK-test (p_{MK}), thereby concluding that the TS test is a suitable method to assess non-stationarity of the extreme wave climate. Subsequently, the MK and TS tests are used to study the non-stationarity of extreme H_{m0} events in the Mediterranean Sea using several extreme value selection methods. The TS test was only applied when the MK test detected the presence of a statistically significant trend. It was found that the direction and magnitude of the temporal trends strongly varied depending on the spatial location. Wang et al. (2021) performed a similar study and found similar results for the South China Sea.

De Leo et al. (2021) provides a practical assessment of the reliability of non-stationary extreme value analysis under the assumption that non-stationary behavior of extreme events can be described by a linear trend. Furthermore, they show how and when non-stationary models for extreme waves provide a potentially better alternative to conventional stationary models. De Leo et al. (2021) used linear temporal covariates to study the non-stationary behavior for the significant wave height (H_S) and peak period (T_P). Results show that the non-stationary analysis provides advantages over the stationary analysis only when a significant trend is present in the data. Moreover, De Leo et al. (2021) emphasized the need for a careful analysis of the reference data before fitting the extreme value distributions since the presence and magnitude of a trend may vary depending on the considered temporal horizon. In short, when applying Non-Stationary EVA, the hypothesis of a (linear) trend and the length of the reference data used for the inference of the non-stationary Extreme Value Distribution should be carefully considered (De Leo et al., 2021).

Raby et al. (2019) and Antonini et al. (2019) studied the extreme wave climate in the safety assessment of rock-mounted lighthouses. Antonini et al. (2019) employed Bayesian Inference to find stationary extreme value distributions for significant wave height and peak period. Raby et al. (2019) used Bayesian Inference to find the non-stationary behavior of the extreme wave climate. The analysis aims to capture a long-term trend in the wave climate by making the location parameter of the Generalized Pareto distribution time-dependent by adding a linear component. The idea of this non-stationary method is to hold the design probability of occurrence constant in time, but allow the corresponding return value of the design parameters to vary in time (Raby et al., 2019). Thereby, Raby et al. (2019) attempts to determine design values for future extreme wave climates based on historic data only. Raby et al. (2019) assessed the structural integrity of the lighthouse in both the present and in the year 2067, under the assumption that the detected upward, linear trend prevails in the future.

Luke et al. (2017) assessed the applicability of NEVA for quantifying the magnitude and frequency of extreme river discharges in The United States. They applied split sample testing to 1250 annual river maximum discharge time-series in the United States and compared the predictive capabilities of extreme discharge events between stationary and non-stationary extreme value analysis. Split sample testing involves the definition of the fitting period and evaluation period of a time-series. The fitting period involves the first half of the time-series and was used to infer extreme value distributions. The evaluation period involves the latter half of the time-series and was used to assess the predictive capabilities of the extreme value models. Luke et al. (2017) challenged the use of NEVA because flood risk management relies on predictions of out-of-sample extreme events. Luke et al. (2017) concluded that for the prediction of extreme river discharges, stationary models are favorable over non-stationary models, to avoid over- and under-extrapolation of extreme events in the future. However, when information about watershed alterations is available, an updated stationary model which accounts for the detected changes in the watershed should be adopted (Luke et al., 2017). In updated stationary modelling, the non-stationary parameter values at the end of the fitting period are used to predict extreme discharge events in the evaluation period under the assumption of stationarity.

Although non-stationary extreme value analyses are becoming increasingly popular in the design of civil infrastructure, some have challenged the use of non-stationary models. Matalas (1997) argued that observed trends in hydrological records cannot be firmly established because of the parameter's natural variability and limited length of observations. This means that the observed trend might only be part of slow oscillations. Limited observations could affect the results of diagnostic tests used to justify the use of non-stationary approaches. This could lead to potentially higher uncertainties in the results of the extreme value analysis. Moreover, a non-stationary approach may involve a future prediction of a covariate which in itself may be uncertain. This adds to the overall uncertainty surrounding future return values of extreme loads (Ragno et al., 2019). Montanari and Koutsoyiannis (2014) stated that efforts should focus on including relevant physical processes in stochastic models and suggested stochastic-process-based models. These can serve as a bridge to fill the gap between "physically-based models without statistics and statistical models without physics." In the case of extreme wave climate, extreme waves are the result of extreme wind speeds. However, it is still essential to make the distinction between locally generated wind waves and traveling swell waves that find their origin in a completely different location.

2.3. Clustering Analysis of Large Scale Oceanic Bodies

The hydrodynamic conditions during extreme wave events potentially vary between different regions in the North Sea. Extreme wave characteristics are influenced by, among others, the dominant wind and wave direction, the available fetch and local water depth (see Appendix A). Creating clusters with similar extreme wave characteristic helps to gain a better understanding of the extreme wave statistics in different regions of the North Sea. Several studies have performed clustering analysis of large scale oceanic bodies. In this section, emphasis is placed on clustering analyses performed based on extreme wave characteristics.

Sartini et al. (2018) and Weiss et al. (2014) applied Regional Frequency Analysis (RFA) on Eastern parts of the North Atlantic ocean to identify regions with similar storm footprints. RFA is applied to

calculate the probability that neighboring grid points are impacted by the same extreme wave events. The probability of occurrence between grid points is subsequently used to calculate the dissimilarity between grid points which enables the clustering of areas based on the storm footprint. Within the identified clusters, the extreme wave statistics can be analyzed using univariate techniques (Sartini et al., 2018).

Goharnejad et al. (2022) employed Self Organizing Maps (SOM) to partition the Northern Atlantic Ocean into clusters that exhibit similar extreme statistics of significant wave height, period, and direction. They did so using complete time-series of the aforementioned parameters at the grid points. The coordinates of the grid point is applied as additional input to separate clusters based on geo-location. The primary reason to do so is that grid points in different geographical locations are typically not affected by the same storms and should therefore be assigned to different clusters. Subsequently, the extreme wave climate can be described at cluster level using a representative time-series, namely the clustering centroid. Using characteristic time-series decreases the time required to do a full extreme value analysis while still being able to analyze the extreme wave statistics throughout the research domain.

Instead of applying clustering analysis of wave time-series at several locations to analyze their spatial relationship, Camus et al. (2011) applied SOM to cluster different sea states at a single location. The SOM analysis was performed using complete time-series of the significant wave height, peak period, mean wave direction and wind variables including the wind speed and direction. Camus et al. (2011) show that SOM identifies the relevant multivariate sea-state types at the research location, and enables the analysis of the dependency between wave and wind parameters by means of visual inspection. The information about the different sea-states at grid point level can be used to set up a clustering analysis to identify spatial clusters that share characteristic sea-states. A temporal analysis of the different sea-states is performed to see if there are changes for the dominant sea-states, which is relevant information for the design and risk assessment of hydraulic infrastructure. Changing sea-states potentially result in changing normative design conditions for the hydraulic infrastructure in question.

3

Methodology

This chapter discusses the methodology applied in this research. The chapter discusses the techniques applied in this research in chronological order. Section 3.1 explains extreme value sampling techniques. Section 3.2 explains the methods for deterction and quantification of statistically significant temporal trends in extreme observations. A hierarchical clustering analysis is applied to identify clusters base on similar extreme wave characteristics. The most important principles behind hierarchical clustering are explained in section 3.3 as well as how hierarchical clustering is applied in this research. The clustering output and detected temporal variability are used to perform extreme value analysis at cluster level. This includes the derivation of non-stationary extreme value models describing the significant wave height (H_{m0}) at cluster level. The most important principles behind extreme value modelling under stationary and non-stationary conditions are explained in Section 3.4. The derivation of the extreme value models at cluster level is explained in Section 3.5.

3.1. Approaches for Sampling Extremes

In extreme value analysis, we are particularly interested in the extreme events within the long term statistics of hydraulic loading parameters. The long term statistics describe how one extreme event differs from the next (Van Den Bos & Verhagen, 2018). This poses the challenge of selecting the extreme events. The goal is to select the extreme events corresponding to the tail of its extreme value distribution (Coles, 2001). Requirements are that the selected storm observations in the dataset are both independent and homogeneous, i.e. the storm events originate from the same meteorological events and only vary in magnitude, (Van Den Bos & Verhagen, 2018). This requires careful analysis of the available dataset such that a homogeneous subset of extreme observations is obtained. Typically, selection of extreme events is performed using Block Maxima (BM) and Peak over Threshold (POT), both of which have been widely applied in studying the behaviour of extreme loading conditions in the design and risk assessment of civil infrastructure.

3.1.1. Block Maxima

In Block Maxima (BM), the time-series is divided into blocks of equal duration, and from each block, the maximum (or minimum) values are obtained (see Figure 3.1). The block duration depends on the application and the available data (Coles, 2001). The selected observations are fitted to one of three distributions form the Generalized Extreme Value (GEV) distributions: Gumbel, Freichet or Weibull (Cheng et al., 2014; Katz et al., 2002). Although being straightforward, BM has its disadvantages. Most notably is the potential loss of useful information (Lang et al., 1999). Within a block, multiple relevant extreme events may be present. However, from each block, only the most extreme event is selected. The contrary may also be true. Within each block, one value is sampled, regardless of whether that value can be characterized as an extreme event.

3.1.2. Peak Over Threshold

In extreme value analysis for engineering purposes, Peak Over Threshold (POT) has become a well-established method, especially for relatively short time-periods when the number of extreme events is

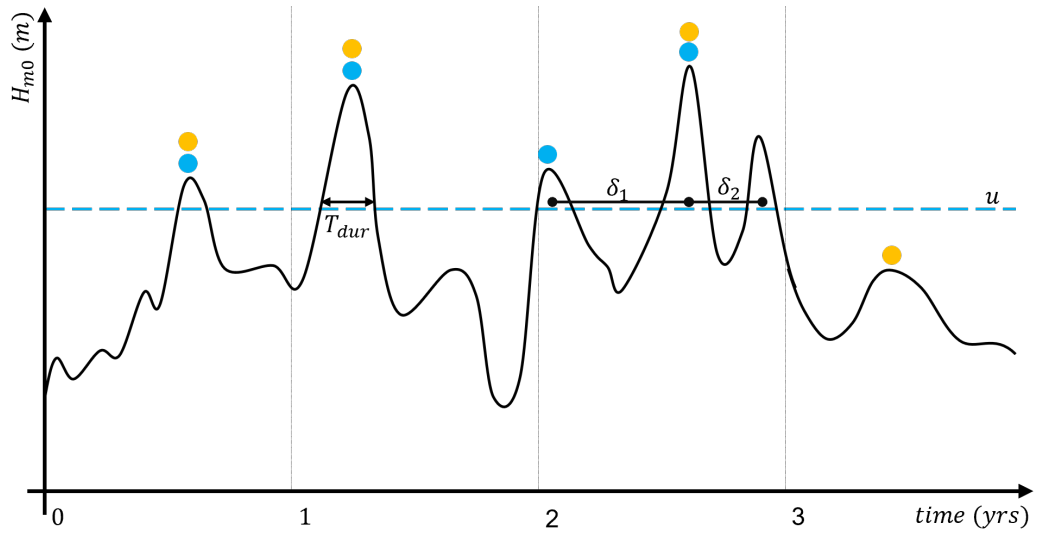


Figure 3.1: Illustration of extreme value selection for Block Maxima (BM) and Peak over Threshold methods (POT). Orange dots indicate values selected for block maxima, while cyan dots indicate selected peaks exceeding threshold u. Note how the third peak in block 3 is not selected, because the inter cluster time δ_2 is smaller than the appropriate declustering time lag δ. Also note how the highest value in block 4 is selected for BM, but not for POT, as the peak does not exceed u.

relatively limited. In a POT analysis, extreme values that exceed a threshold (u) are sampled (Coles, 2001). The POT approach provides the ability to control the number of extreme events to be included in the EVA depending on the selected value for u. Therefore, POT is able to capture more information than the BM approach. There are some challenges associated to the Peak Over Threshold method. An appropriate threshold (u) and declustering time lag (δ) must be selected such that the observed extreme events follow a Poisson distribution (Antonini et al., 2019). This relies on three properties of the peaks over the selected threshold:

- Peaks should occur randomly in time according to the Poisson process;
- The exceedances should have an approximate GP distribution;
- Exceedances should be sufficiently far apart to be considered independent.

In a Poisson process, the observed events are defined as points in time (Antonini et al., 2019). In the study of the hydraulic boundary conditions, these events are storm events that can be defined as follows. A storm starts when the variable (i.e. water level, significant wave height) exceeds a defined threshold (u) and has been less than u for at least n consecutive days. The storm event finishes when the variable drops below u and remains below u for at least n consecutive days (Antonini et al., 2019).

The choice for a threshold u in the POT analysis results in a trade-off between bias and variance (Roscoe et al., 2010). If the selected threshold is too low, the number of observations used to compute the model is very high. This likely leads to violation of the asymptotic basis of the model, leading to bias. A very high threshold may result in too few observations to compute the model with. This likely means that the model meets the mathematical requirements (i.e. independent and homogeneous) but also results in a larger uncertainty (Roscoe et al., 2010). It is possible to select appropriate thresholds (u) and declustering times (δ) using the diagnostic tests mean residual life (MRL), dispersion index (DI) and extremal index (EI) and the stability of the Generalized Pareto (GP) distribution parameters (Antonini et al., 2019; Coles, 2001; Davison & Smith, 1990). The process of selecting appropriate POT threshold values with the use of diagnostic tests is further explained in Appendix E.

Apart from analysing the intensity of extreme events (i.e. magnitude of the peak values), POT analysis is also useful for the analysis of the frequency and duration of extreme events. The frequency of extreme events in this research is defined as the number of annual threshold exceedances, and is denoted by N_{exc}. The duration of an extreme events is defined as the time between the start and end of the

threshold exceedance and is denoted by T_{dur}. The definitions of N_{exc} and T_{dur} are also illustrated in Figure 3.1.

3.2. Temporal Trend Analysis

Whether a non-stationary approach in Extreme Value Analysis is applicable, depends on whether a statistically significant trend can be detected in the data. Several methods for the detection and measurements of trends exist. In this study, the non-parametric Mann-Kendall hypothesis test (Kendall, 1955; Mann, 1945) is applied to detect the presence of a statistically significant temporal trend. The Theil-Sen (TS) test (Sen, 1968; Theil, 1992) is applied to quantify the detected temporal trend. Both will be explained in the following sections.

3.2.1. Mann-Kendall Test

The Mann-Kendall (MK) test is an hypothesis test aimed at evaluating whether a significant monotonic trend can be detected in the dataset. The hypotheses in the MK-test are stated as follows:

- H_0 (Null Hypothesis): There is no monotonic trend in the data;
- H_A (Alternative): There is a monotonic trend in the data.

For an arbitrary time-series containing n observations, the test statistic S_{MK} for the MK-test can be found using the sign function over the differences between subsequent observations:

$$S_{MK} = \sum_{i<k} \text{sign}\,(x_k - x_i) \quad \text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \tag{3.1}$$

The significance of the trend can be obtained by converting the test statistic S_{MK} to the standardized test statistic Z_{MK} as follows:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sigma(S)} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sigma(S)} & \text{if } S < 0 \end{cases} \quad \sigma(S) = \sqrt{\frac{n(n-1)(2n+5)}{18}} \tag{3.2}$$

In equation 3.2, Z_{MK} denotes the standardized test statistic of S, $\sigma(S)$ is the standard deviation, and n is the sample size. The value of Z_{MK} is then compared to the percentile of the standard normal distribution, leading to the p-value p_{MK} of the statistic:

$$p_{MK} = 2\Phi\left(-\left|Z_{MK}\right|\right) \tag{3.3}$$

In Equation 3.3, Φ stands for the cumulative distribution function of the standard normal distribution. If p_{MK} is larger than the significance level α (typically 0.05), the null hypothesis (H_0) cannot be rejected, and there is no significant trend in the data. In its common use, the MK-test does not provide any information about the direction or magnitude of temporal trends. It only detects the presence of a trend. De Leo et al. (2020) states that Z_{MK} and α can be used to determine the direction of the trend (i.e. upward or downward oriented) using the following equations:

$$\begin{cases} Z_{MK} > \Phi^{-1}(1-\alpha/2) \rightarrow \text{ upward trend} \\ Z_{MK} < -\Phi^{-1}(1-\alpha/2) \rightarrow \text{ negative trend} \\ Z_{MK} < \left|\Phi^{-1}(1-\alpha/2)\right| \rightarrow \text{ no trend.} \end{cases} \tag{3.4}$$

3.2.2. Theil-Sen Test

The Theil-Sen (TS) test (Sen, 1968; Theil, 1992) is a method of linear regression and can be used to quantify the trend under the assumption that the non-stationarity can be described by a linear trend. The TS test is insensitive to outliers, and is therefore preferred over other linear regression methods (De Leo et al., 2020). Consider a series of data points x_i ($i = 1...n$) where n is the total number of data points in the dataset. The TS-slope can be computed as:

$$b = \text{Median}\left(\frac{x_j - x_l}{j - l}\right) \quad \forall \quad l < j, \quad l, j = 1 \ldots n \tag{3.5}$$

In equation 3.5, x_j and x_l are the jth and lth data of the series respectively, such that every combination between two data points in the series is included. Note that the values for j and l can be replaced by for example date numbers (e.g. days or years) to find the rate of change over time (De Leo et al., 2020). This can be done as long as j and l are consistent with the data points x_j and x_l. Apart from the slope of the trend, we are also interested in calculating the intercept (b_0) at the start of the considered temporal horizon. The value for b_0 can be calculated as follows:

$$b_0 = median(X_i - b \cdot t_i) \tag{3.6}$$

In equation 3.6, X_i is the magnitude of observed variable i, and t_i is the elapsed time since the start of the considered temporal horizon. Finally, b is the slope of the trend as determined using Equation 3.5.

3.3. Hierarchical Clustering

The main goal of a Hierarchical Agglomerative Clustering (HAC) analysis is to partition a gridded dataset into a select number of clusters such that the established clusters exhibit similar characteristics. In this research, hierarchical clustering analysis will be performed with the aim to define characteristic areas with similar extreme wave characteristics in the North Sea domain. The extreme wave characteristics vary throughout the North Sea. The local extreme wave climate in influenced by numerous factors, including the wind and wave direction, available fetch for wave growth and dispersion and the available depth which potentially limits wave growth.

Creating clusters with similar characteristic extreme wave conditions helps to gain a better understanding of the presence of regions with similar extreme wave characteristics, with the aim of improving extreme value analysis (EVA) intended for the design and risk assessment of hydraulic infrastructure. The second objective is to reduce the data volume required to process to get a better understanding about the North Sea wave climate. Potentially, one representative time-series per cluster must be studied instead of the entire population of spatial grid points.

As the name suggests, hierarchical clustering is a clustering method that aims to build a hierarchy of clusters, often visualized using dendrograms (Everitt et al., 2011; Maimom & Rockach, 2010). A dendrogram is a multilevel hierarchy in which objects on one level are joined as clusters on the next level. This hierarchical structure enables the selection of an appropriate cut-off point that determines the number of clusters that yields the optimal clustering division of the extreme wave climate in the North Sea. Hierarchical clustering is an unsupervised learning algorithm, which means that no prior information about the system are presented to the algorithm, leaving the algorithm on its own to find structure in the input values (Sokal & Michener, 1958). Within hierarchical clustering, two strategies exist (Maimom & Rockach, 2010):

- Agglomerative clustering (bottom-up): Each grid point starts as its own cluster. The grid points (and subsequently the newly found clusters) are merged as one moves higher up the dendrogram until all objects are in the same cluster;
- Divisive clustering (top-down): All grid points start as one cluster, and splits within the cluster are formed as one moves down the hierarchy until each grid point can be described as an individual cluster.

This research employs Hierarchical Agglomerative Clustering (HAC). HAC was selected because it is straightforward to understand and implement. Furthermore, the output in the form of a dendrogram provides a good understanding of the established cluster boundaries. This section explains how HAC analysis was applied in this research.

3.3.1. Input Data for Hierarchical Clustering Analysis

Prior to running the HAC algorithm, the input must be generated. This includes the careful selection of wave parameters and representative input values that form the basis of the clustering output. Furthermore, pre-processing and standardization of the input values and assigning weights to the input parameter that must be prioritized in the clustering analysis form the initiation of the HAC analysis.

3.3.1.1. Parameter selection

In this research, we perform a clustering analysis to identify regions that exhibit similar extreme wave characteristics. This requires the careful selection of wave parameters that describe the extreme wave climate at the grid points. In HAC analysis, all parameters are represented by a single, characteristic value. This limits the amount of information about the parameter that is presented to the HAC algorithm. Furthermore, the clustering output of the HAC analysis depends on the specific pre-selected input values. That means, if grid-points share similar values for the input data, they are assigned to the same cluster. These two limitations mean that a careful consideration of the wave parameters as well as the input values representing the wave parameter is required. Section 4.4.1 explains the motivation and derivation of the wave parameters that serve as input for the HAC analysis in this research.

3.3.1.2. Pre-Processing of Input Values

Clustering analysis involves the evaluation of input values at grid points over a very large surface area. It is possible that the input values vary smoothly across the research domain, resulting in continuous input values. Clustering spatially continuous data is potentially problematic because of the small dissimilarities between spatially neighboring grid points. Due to these small dissimilarities, the HAC algorithm continuously links spatially neighboring grid points into one big cluster because it cannot determine a clear cut off point. This is often referred to as the Chaining Problem (Everitt et al., 2011). To overcome the Chaining Problem, pre-processing of the input values is required. This typically includes rounding the input values to increase the dissimilarity between grid points. It should be noted that pre-processing of the input data is very subjective and directly influences the clustering output.

3.3.1.3. Standardization of Input Values

Before running the HAC algorithm, the input values must be normalized. Otherwise, the difference in magnitudes between the wave parameters may become problematic whilst calculating the pairwise dissimilarities. Without normalization, greater importance is given to parameters with greater magnitude (Mohamad & Usman, 2013; Suarez-Alvarez et al., 2012). Within HAC, several methods for standardization exists, e.g. the min-max linear transformation approach and the z-score standardization (Suarez-Alvarez et al., 2012). The min-max normalization can be calculated as follows:

$$X_i^* = \frac{X_i - min(X)}{max(X) - min(X)} \tag{3.7}$$

Min-Max normalization is a method of linear scaling and will transform the data so that all values are in the domain $[0, 1]$. Another method is z-score standardization and can be calculated as follows:

$$X_i^* = \frac{X_i - \mu_X}{\sigma_X} \tag{3.8}$$

in equation 3.8, μ_X is the mean of all observations for variable X and σ_X is the standard deviation for X. Z-score standardization transforms the values for X in such a way that the mean of the input values X^* becomes 0 and the standard deviation 1. A critical assessment of the best normalization technique lies outside the scope of this thesis. However, Suarez-Alvarez et al. (2012) states that z-score standardization is best suited when calculating the dissimilarities using Euclidean distance as it best preserves the relative difference between the individual observations. Min-Max normalization usually performs worse than z-score standardization, as it depends on the two most extreme values, which tend to be outliers (Suarez-Alvarez et al., 2012).

3.3.1.4. Weight Configurations

Because the input values are standardized, it is possible to assign weights to the different parameters. Weight may be assigned to emphasize particular parameters in the clustering that are thought more significant or to add additional steering to improve the quality of the clustering output (Fern Tay, 2021). Weights must be assigned in such a way that the sum of the weights assigned to the input parameters is equal to 1:

$$\sum_{k=1}^{M} W_k = 1 \tag{3.9}$$

In this equation, W_k is the weight assigned to wave parameter k, and M represents the total number of dimensions, i.e. the number of wave parameters included in the HAC analysis

3.3.2. Pairwise Dissimilarity

The first step in the Hierarchical Agglomerative Clustering (HAC) algorithm is to determine the dissimilarity between every pair of grid points in the dataset. For a dataset containing N grid points, there exist $N * (N-1)/2$ pairs for which the dissimilarity can be calculated. In this research, Euclidean distance is employed to calculate the dissimilarity between grid points (Everitt et al., 2011). The Euclidean distance between two grid points (i and j) can be calculated as follows:

$$d(i,j) = \|i-j\|_2 = \sqrt{\sum_{k=1}^{M} (i_k - j_k)^2} \tag{3.10}$$

In this equation, M represents the number of dimensions, i.e. the number of different parameters involved in the clustering process. This research includes several circular parameters, for example the mean wave direction (θ). The Euclidean distance between two circular parameters can be calculated using the shortest distance around the circle, and the equation for Euclidean distance becomes (Camus et al., 2011):

$$d(i,j) = \|i-j\|_2 = \sqrt{\sum_{k=1}^{M} \left(min\left((i_k - j_k), 2\pi - (i_k - j_k)\right)\right)^2} \tag{3.11}$$

3.3.3. Linkage Mechanisms

The next step in the HAC algorithm is to determine how the pairwise dissimilarity between grid points should be applied to link grid points to form clusters. This requires the selection of the appropriate linkage mechanism. The linkage mechanism how the dissimilarities between a pair of objects is calculated. This is also referred to as the cophenetic dissimilarity (T). At each step during the linkage process, the two objects separated by the lowest value for T are combined to form a new cluster at the next level. Subsequently, the linkage function links the newly formed clusters to create bigger clusters until all grid points are linked together in a hierarchical tree. The calculation of T is straightforward when only individual grid points are considered, but becomes ambiguous when clusters consisting of multiple grid points must be linked together. Different linking mechanisms apply different definitions of dissimilarity between established clusters. This section explains the linkage mechanisms that are employed in this thesis.

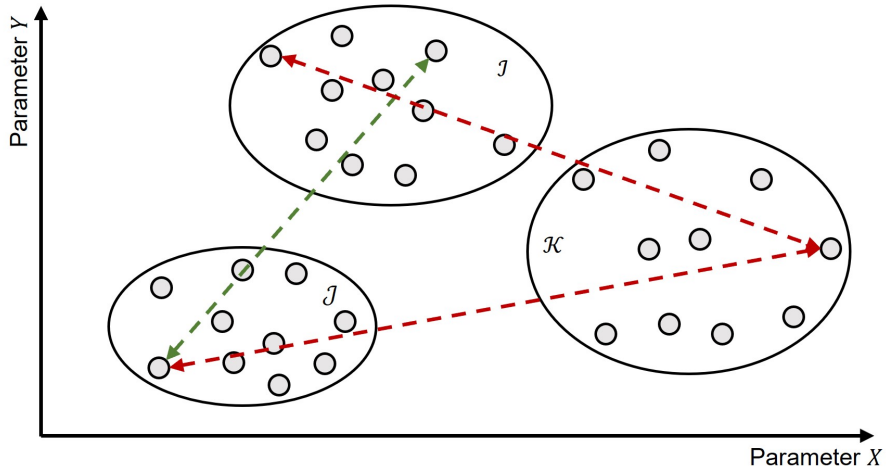


Figure 3.2: Illustration describing Complete Linkage. Arrows show the considered grid points in the clusters to calculate the dissimilarities between a pair of clusters ($T(\mathcal{I},\mathcal{J})$, $T(\mathcal{I},\mathcal{K})$ and $T(\mathcal{J},\mathcal{K})$). Green Arrows indicate the clusters linked together based on Complete Linkage.

3.3.3.1. Complete (Maximum) Linkage Function

In complete linkage clustering, the dissimilarity between clusters is defined by the dissimilarities between those two grid points (one in each cluster) that are farthest away from each other. This is il-

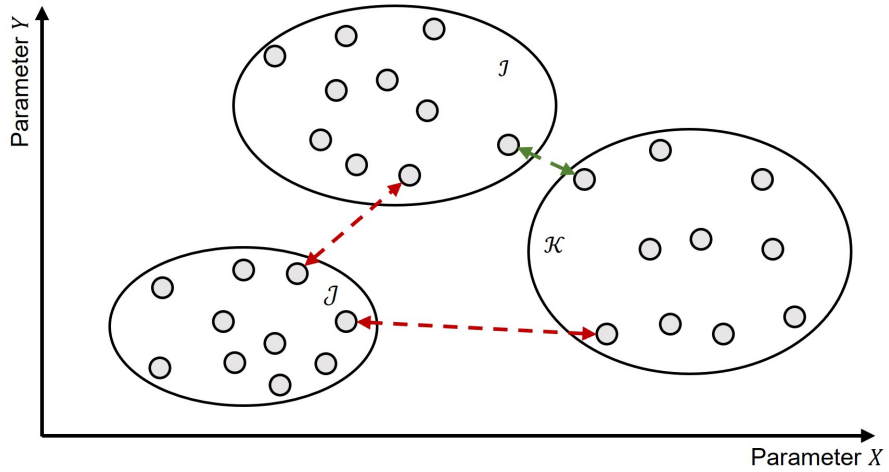


Figure 3.3: Illustration describing Single Linkage. Arrows show the considered grid points in the clusters to calculate the dissimilarities between a pair of clusters ($T(\mathcal{I},\mathcal{J})$, $T(\mathcal{I},\mathcal{K})$ and $T(\mathcal{J},\mathcal{K})$). Green Arrows indicate the clusters linked together based on Single Linkage.

lustrated in Figure 3.2. Mathematically, the dissimilarities between two clusters as calculated by the Complete linkage function can be described using equation 3.12 (Sokal & Michener, 1958).

$$T(\mathcal{I},\mathcal{J}) = \max_{i\in\mathcal{I},j\in\mathcal{J}} d(i,j) \tag{3.12}$$

In equation 3.12, $d(i,j)$ is the euclidean distance between elements $i \in \mathcal{I}$ and $j \in \mathcal{J}$. $\mathcal{I}$ and $\mathcal{J}$ are two clusters that are potentially linked together at a given step during the linkage process. $T(\mathcal{I},\mathcal{J})$ is the determined dissimilarity between clusters $\mathcal{I}$ and $\mathcal{J}$ as determined by the complete linkage function. The value T can be calculated for every pair of clusters (see Figure 3.2). The cluster combination that yields the lowest value for T are linked to form a new cluster. Complete is less vulnerable to chaining than Single linkage, but potentially suffers from crowding (Everitt et al., 2011). Because the calculated dissimilarity is based on the worst-case dissimilarity between two clusters (i.e. grid points that are farthest apart from each other), a point can be closer to points in other clusters than to points in its own cluster

3.3.3.2. Single (Minimum) Linkage Function

In single-linkage clustering, the dissimilarity between two clusters is determined by the pair of grid points (one in each cluster) that are closest to each other. This is illustrated in Figure 3.3. Mathematically, the dissimilarity between two clusters as calculated by the Single linkage function can be described using the following expression (Sokal & Michener, 1958):

$$T(\mathcal{I},\mathcal{J}) = \min_{i\in\mathcal{I},j\in\mathcal{J}} d(i,j) \tag{3.13}$$

Similarly to complete linkage, the lowest value of T for all pairs of clusters determines which two clusters should be merged. Comparing Figures 3.2 and 3.3 shows that different linkage mechanisms can result in different clusters to linked. Single linkage is particularly vulnerable to chaining. In order to merge two clusters, only need one pair of grid-points to be close, irrespective of all others (Everitt et al., 2011; Murtagh & Contreras, 2012).

3.3.3.3. Average Linkage Function

In the average linkage criterion, the distance between the two clusters is determined based on the average distance between any member from one cluster to any member of the other cluster (Sokal & Michener, 1958). This process is illustrated in Figure 3.4. This can mathematically be expressed as Equation 3.14. In equation 3.14, $|\mathcal{I}|$ and $|\mathcal{J}|$ are the sizes of the clusters, i.e. the number of objects in each of the clusters.

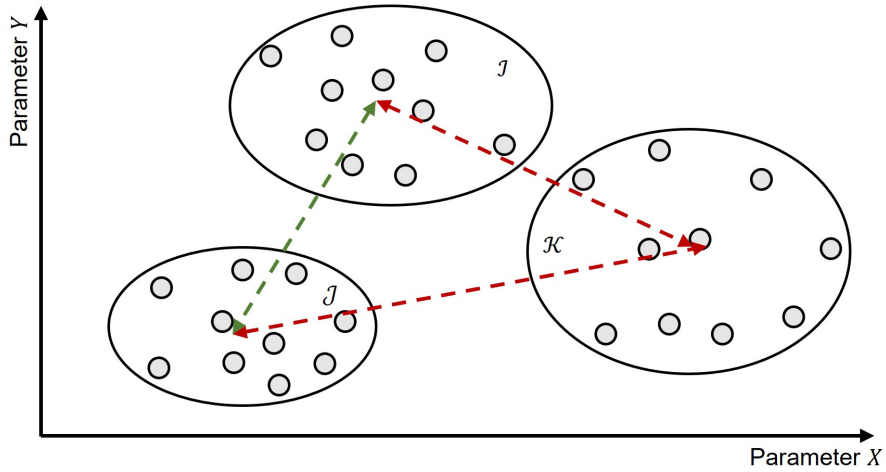


Figure 3.4: Illustration describing Average Linkage. Arrows show the considered grid points in the clusters to calculate the dissimilarities between a pair of clusters ($T(\mathcal{I},\mathcal{J})$, $T(\mathcal{I},\mathcal{K})$ and $T(\mathcal{J},\mathcal{K})$). Green Arrows indicate the clusters linked together based on Average Linkage.

$$T(\mathcal{I},\mathcal{J}) = \frac{1}{|\mathcal{I}|\cdot|\mathcal{J}|}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} d(i,j) \tag{3.14}$$

3.3.3.4. Ward Linkage Function

Ward's linkage criterion (Ward, 1963) is different from the three aforementioned linkage mechanisms as it is based on minimizing the within cluster sum of squares (WCSS). Ward linkage is illustrated in Figure 3.5. WCSS is defined as the sum of squares of the dissimilarities between all grid-points in the cluster and the centroid of the resulting cluster (Ward, 1963) after merging. To implement this method, at each linkage step, the algorithm finds the pair of clusters that leads to minimum increase in WCSS after merging (Ward, 1963). This is illustrated in Figure 3.5. Ward linkage can mathematically be expressed as Equation 3.15. In equation 3.15, $\overline{\mathcal{I}}$ and $\overline{\mathcal{J}}$ are the centroids of clusters $\mathcal{I}$ and $\mathcal{J}$ and $\|\overline{\mathcal{I}}-\overline{\mathcal{J}}\|_2$ is the Euclidean distance between the centroids.

$$T(\mathcal{I},\mathcal{J}) = \sqrt{\frac{2|\mathcal{I}||\mathcal{J}|}{|\mathcal{I}|+|\mathcal{J}|}}\|\overline{\mathcal{I}}-\overline{\mathcal{J}}\|_2 \tag{3.15}$$

3.3.3.5. Dendrograms

The hierarchical linkage is best understood when viewed graphically using dendrograms. In a dendrogram, the horizontal axis represents the objects included in HAC. The links between objects are represented by upside-down U-shaped lines. The height of the U-shapes represents the dissimilarity between the objects (grid points or clusters) as calculated by the selected linkage mechanism (T). This is also referred to as the cophenetic dissimilarity (Maimom & Rockach, 2010).

3.3.3.6. Cophenetic Correlation

It must be verified that the cophenetic dissimilarities in the dendrogram accurately reflect the dissimilarities in the original dataset. One measure to study the performance of the linkage process is to compare the cophenetic dissimilarities (T) between grid points with the original pairwise dissimilarities (d), using the cophenetic correlation coefficient (C, see Equation 3.16). If the clustering is valid, the cophenetic dissimilarities in the dendrogram should correlate strongly with grid point dissimilarities, and C yields values close to 1.

$$C = \frac{\sum_{i<j}\left(d(i,j)-\overline{d}\right)\left(T(i,j)-\overline{T}\right)}{\sqrt{\sum_{i<j}\left(d(i,j)-\overline{d}\right)^2\sum_{i<j}\left(T(i,j)-\overline{T}\right)^2}} \tag{3.16}$$

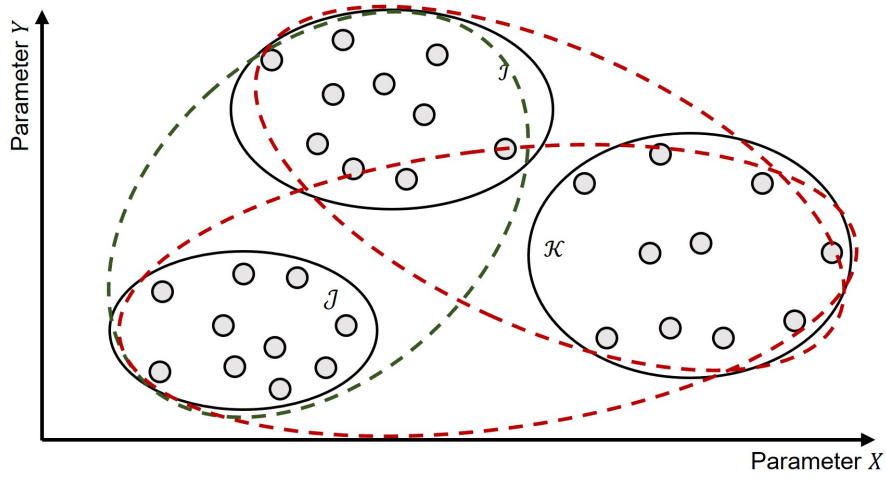


Figure 3.5: Illustration describing Ward Linkage. The dashed areas indicate the pairs of clusters considered in this clustering step. The green area indicates the two clusters linked together based on Ward Linkage. The green area has the smallest surface area, and therefore the smallest value for WCSS.

In equation 3.16, $d_{i,j}$ is the pairwise Euclidean distance between grid points, and $\overline{d}$ is the mean Euclidean distance for all Euclidean distances. Let $T(i,j)$ be the cophenetic dissimilarity between objects i and j and $\overline{T}$ be the average cophenetic distance of all cophenetic correlations. The closer the value of the cophenetic correlation coefficient (C) is to 1, the more accurately the clustering solution reflects your data.

3.3.4. Internal Evaluation Metrics

The final step in HAC is to select the appropriate cut-off points that determines the number of clusters ($\mathcal{K}$) that adequately describes the extreme wave climate in the North Sea. To determine the ideal number of clusters, two types of clustering validation criteria exist. External validation criteria evaluate the clustering results with respect to a pre-specified structure (e.g. knowledge about North Sea wave climate presented to algorithm prior to clustering), which is not available for unsupervised learning algorithms. Internal criteria on the other hand evaluate the clustering results with respect to the information intrinsic to the input data alone, and is therefore the preferred method in this research. The Internal validation methods that are used in this research are the silhouette coefficient (SC), Calinski-Harabasz (CH) Criterion and the Davies-Bouldin (DB) score.

3.3.4.1. Silhouette coefficient

The silhouette value is a measure of how similar a grid point i is to its own cluster (hereafter denoted by $\mathcal{I}$) compared to other clusters (hereafter denoted by $\mathcal{J}$, (Rousseeuw, 1987)). The silhouette ($s(i)$) for every grid point in the dataset can be calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ if } |N_{\mathcal{I}}| > 1 \tag{3.17}$$

In equation 3.17, $N_{\mathcal{I}}$ is the number of grid points in cluster $\mathcal{I}$. If $N_{\mathcal{I}}$ is equal to 1, $s(i)$ is equal to zero. Furthermore, $a(i)$ is the mean dissimilarity between object i and all other objects included in cluster $\mathcal{I}$, and is often referred to as cohesion (Kaufman & Rousseeuw, 1990). Let $b(i)$ be the smallest mean dissimilarity of object i to all points in any other cluster than $\mathcal{I}$ (Rousseeuw, 1987). $b(i)$ is often referred to as separation. The cluster with the smallest mean dissimilarity is said to be the neighboring cluster, as it would be the "second best" cluster for i to be assigned to (Kaufman & Rousseeuw, 1990). The equations for $a(i)$ and $b(i)$ are as follows:

$$a(i) = \frac{1}{|N_{\mathcal{I}}| - 1} \sum_{j \in C_I, i \neq j} d(i,j) \tag{3.18}$$

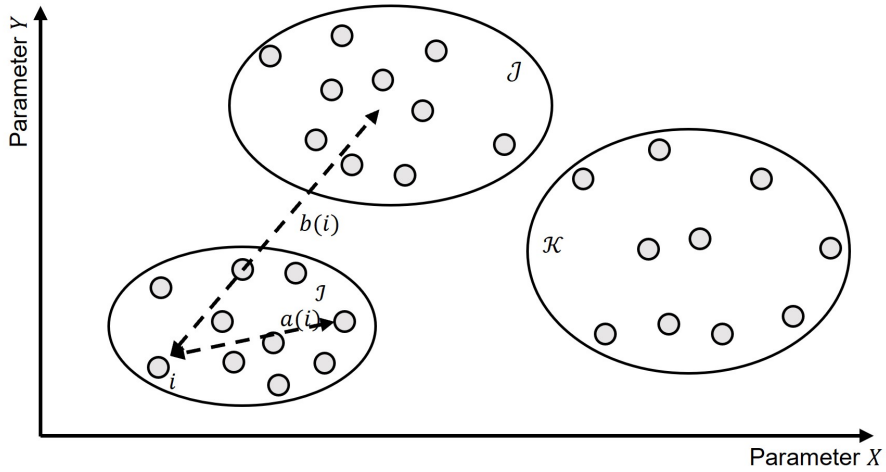


Figure 3.6: Illustration showing how the silhouette of a grid point (i) is determined. Parameter $a(i)$ is the mean dissimilarity between object i and all other objects in cluster $\mathcal{I}$. Parameter $b(i)$ is the smallest mean dissimilarity of object i to all points in any other cluster than $\mathcal{I}$.

$$b(i) = \min_{\mathcal{J} \neq \mathcal{I}} \frac{1}{|N_{\mathcal{J}}|} \sum_{j \in C_{\mathcal{J}}} d(i,j) \tag{3.19}$$

The silhouette value ranges between -1 and 1. A high value for $s(i)$ indicates that the object i is well matched to its own cluster and poorly matched to neighboring clusters. For $s(i)$ to be close to 1, it should hold that $b(i) >> a(i)$. If all objects in the clustering analysis have a high value for $s(i)$, the clustering output is appropriate. The mean of all $s(i)$ values (denoted by $\overline{S}$) is a measure of how tightly grouped the grid points are within the clusters they are assigned to. This is an indication of how well the grid points have been clustered (Rousseeuw, 1987). To this end, Kaufman and Rousseeuw (1990) introduced the term silhouette coefficient (SC) to determine the optimal number of clusters for the objects in the dataset, and can be written as follows:

$$SC = \overline{S}(\mathcal{K}) \tag{3.20}$$

In Equation 3.20, $\overline{S}(\mathcal{K})$ is the mean value of all $s(i)$ for the selected number of clusters $\mathcal{K}$. SC can be calculated for any number of clusters $\mathcal{K}$. The value for $\mathcal{K}$ that gives the highest value for SC, is the ideal number of clusters.

3.3.4.2. Calinski-Harabasz Criterion

The Calinski-Harabasz (CH) criterion (Caliński & Harabasz, 1974) aims to minimize the within cluster variance and to maximize the between cluster variance. The equation for the CH criterion is as follows:

$$CH(\mathcal{K}) = \frac{SS_B}{SS_W} \cdot \frac{N - \mathcal{K}}{\mathcal{K} - 1} \tag{3.21}$$

In Equation 3.21, N is the total number of grid points in the cluster analysis, and $\mathcal{K}$ is the number of clusters. Furthermore, SS_B is a measure of the between-cluster variance, and SS_W is a measure of the within cluster variance (Caliński & Harabasz, 1974). Mathematically, SS_B and SS_W can be defined as:

$$SS_B = \sum_{\mathcal{I}=1}^{\mathcal{K}} N_{\mathcal{I}} \left\| \overline{\mathcal{M}}_{\mathcal{I}} - \overline{\mathcal{M}} \right\|_2 \tag{3.22}$$

$$SS_W = \sum_{\mathcal{I}=1}^{\mathcal{K}} \sum_{i \in \mathcal{C}_{\mathcal{I}}} \left\| i - \overline{\mathcal{M}}_{\mathcal{I}} \right\|_2 \tag{3.23}$$

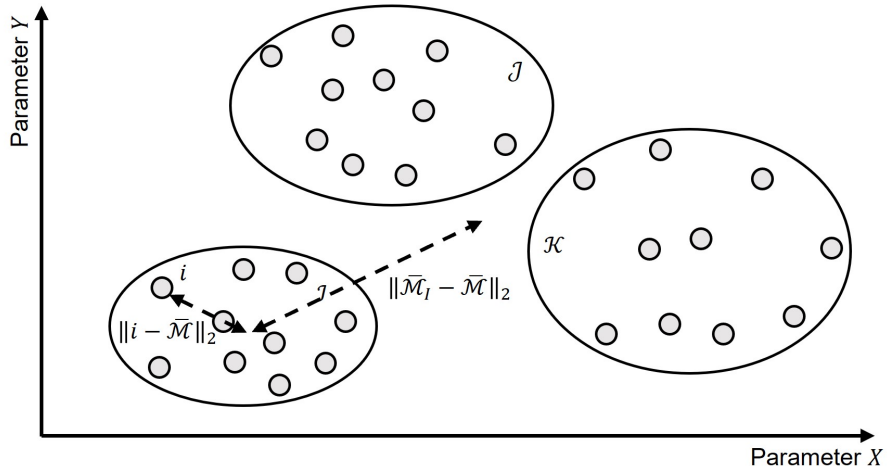


Figure 3.7: Illustration showing how the Calinski-Harabasz Score is calculated. $\left\|\overline{\mathcal{M}}_{\mathcal{I}} - \overline{\mathcal{M}}\right\|_2$ is the Euclidean distance between the centroid of cluster $\mathcal{I}$ and the overall mean of the input data $\overline{\mathcal{M}}$. Furthermore, $\left\|i - \overline{\mathcal{M}}_{\mathcal{I}}\right\|_2$ be the Euclidean distance between an observation i in cluster $\mathcal{I}$ and the centroid of the cluster $\overline{\mathcal{M}}_{\mathcal{I}}$.

In equations 3.22 and 3.23, $\mathcal{K}$ is the number of clusters, $N_{\mathcal{I}}$ is the number of grid points in cluster $\mathcal{I}$, and $\left\|\overline{\mathcal{M}}_{\mathcal{I}} - \overline{\mathcal{M}}\right\|_2$ is the Euclidean distance between the centroid of cluster $\mathcal{I}$ (denoted by $\overline{\mathcal{M}}_{\mathcal{I}}$ and the overall mean of the input data $\overline{\mathcal{M}}$. Finally, let $\left\|i - \overline{\mathcal{M}}_{\mathcal{I}}\right\|_2$ be the Euclidean distance between an observation i in cluster $\mathcal{I}$ and the centroid of the cluster $\overline{\mathcal{M}}_{\mathcal{I}}$. According to Caliński and Harabasz (1974), well defined clusters have a small value for SS_W and large value for SS_B. This means that better clustering solutions yield higher values for $CH(\mathcal{K})$.

3.3.4.3. Davies-Bouldin Index

The final evaluation criterion applied in this research is the Davies-Bouldin (DB) index (Davies & Bouldin, 1979). Similar to SC, DB compares within-cluster and between-cluster dissimilarities. The Davies-Bouldin index is mathematically defined as:

$$DB(\mathcal{K}) = \frac{1}{\mathcal{K}} \sum_{\mathcal{I}=1}^{\mathcal{K}} \max_{\mathcal{I} \neq \mathcal{J}} \{D_{\mathcal{I},\mathcal{J}}\} \tag{3.24}$$

In equation 3.24, $D_{\mathcal{I},\mathcal{J}}$ is the within-to-between cluster distance ratio for the clusters $\mathcal{I}$ and $\mathcal{J}$. $D_{\mathcal{I},\mathcal{J}}$ can be defined as:

$$D_{\mathcal{I},\mathcal{J}} = \frac{(\bar{d}_{\mathcal{I}} + \bar{d}_{\mathcal{J}})}{\bar{d}_{\mathcal{I},\mathcal{J}}} \tag{3.25}$$

In equation, 3.25, $\bar{d}_{\mathcal{I}}$ is the average dissimilarity between all points included in cluster $\mathcal{I}$ and the centroid of cluster $\mathcal{I}$. Similarly, $\bar{d}_{\mathcal{J}}$ is the average dissimilarity between all points included in cluster $\mathcal{J}$ and the centroid of cluster $\mathcal{J}$. $\bar{d}_{\mathcal{I},\mathcal{J}}$ is the Euclidean distance between the centroids of clusters $\mathcal{I}$ and $\mathcal{J}$. According to (Davies & Bouldin, 1979), for a clustering solution to perform well, the values for $\bar{d}_{\mathcal{I}}$ and $\bar{d}_{\mathcal{J}}$ must be as low as possible, meaning that the within-cluster grid points are tightly grouped. Also, the value for $\bar{d}_{\mathcal{I},\mathcal{J}}$ must be as high as possible, meaning that dissimilarities between the clusters are high. Davies and Bouldin (1979) then selects $max\,(D_{\mathcal{I},\mathcal{J}})$, because it gives relatively high values for $\bar{d}_{\mathcal{I}}$ and $\bar{d}_{\mathcal{J}}$ and low values for $\bar{d}_{i,j}$, referring to the two clusters that give the poorest clustering results. This means that the optimal number of clusters is the value for $\mathcal{K}$ that gives the lowest value for $DB(\mathcal{K})$.

3.3.5. Initial Number of Clusters

Before running the Hierarchical Agglomerative Clustering (HAC) algorithm, a maximum number of clusters ($\mathcal{K}_{max}$) that must be considered by the HAC-algorithm should be selected. The HAC algorithm assesses the clustering output for all numbers of clusters ($\mathcal{K}$) between 1 and $\mathcal{K}_{max}$ to find the value for

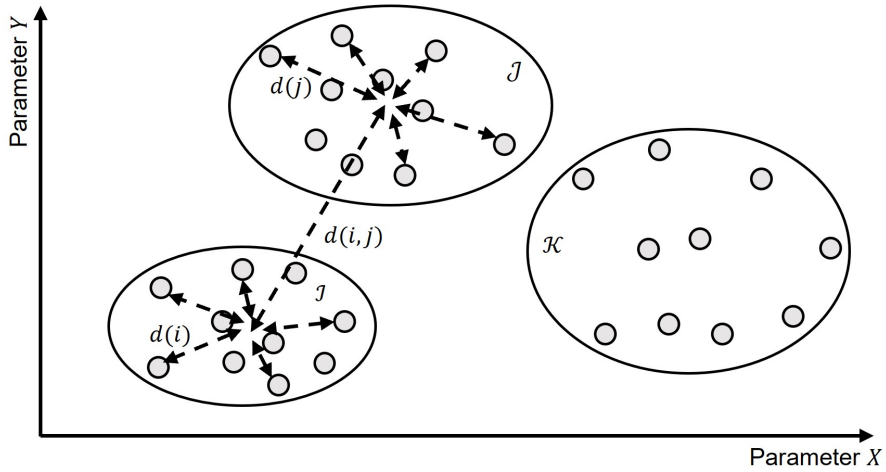


Figure 3.8: Illustration showing how the Davies-Bouldin ratio is calculated. $\bar{d}_{\mathcal{I}}$ is the average dissimilarity between all points included in cluster $\mathcal{I}$ and the centroid of cluster $\mathcal{I}$. Similarly, $\bar{d}_{\mathcal{J}}$ is the average dissimilarity between all points included in cluster $\mathcal{J}$ and the centroid of cluster $\mathcal{J}$. $\bar{d}_{\mathcal{I},\mathcal{J}}$ is the Euclidean distance between the centroids of clusters $\mathcal{I}$ and $\mathcal{J}$.

$\mathcal{K}$ that yields the best clustering output.

One of the goals of the clustering analysis is potential data volume reduction. To maximize data volume reduction, the value of $\mathcal{K}$ should not be too high. However, in HAC analysis, there is always a trade-off between $\mathcal{K}$ and the quality of the clustering output. If the value of $\mathcal{K}$ is too low, grid-points with different extreme wave characteristics will be linked together. Contrarily, if the value of $\mathcal{K}$ is too high, the clustering output will be a good representation of the extreme wave characteristics, but the data volume reduction will not be as efficient. To limit the number of clusters identified by the HAC-algorithm, this research imposed a maximum number of clusters to be analyzed,: $K_{max} = 50$.

3.3.6. Assessing Clustering Output

The HAC analyses in this research are performed for different wave parameter combinations and weight configurations. For each configuration, different linkage mechanisms are evaluated that subsequently give different values for the number of clusters ($\mathcal{K}$) that yields the best clustering output. For each configuration, the best clustering output must be identified. This section explains the steps required to determine the best clustering output for each configuration. This involves the selection of the the best linkage mechanism, determining the ideal value for $\mathcal{K}$ for each linkage mechanism and a critical assessment of the clustering output to compare how well the clustering output represents the input values. Once the best clustering output for each configuration is determined, the best overall clustering output to be used in the research must be determined.

3.3.6.1. Selecting Linkage Mechanism

The linkage mechanisms Single, Complete, Ward and Average are used to form clusters based on the dissimilarity information. The first step is to compare the performance of the different linkage mechanisms using the cophenetic correlation C (See Section 3.3.3.6). However, the best scoring dendrogram does not always result in the best clustering solution. Based on the value for C, the two best scoring dendrograms are selected for further analysis.

3.3.6.2. Selecting Number of Clusters

For the two linkage mechanisms selected in Section 3.3.6.1, the internal evaluation criteria - Silhouette Coefficient (SC), Calinski Harabasz criterion (CH) and the Davies-Bouldin value (DB) are assessed to find the value for $\mathcal{K}$ that yields the best clustering output. For each of the internal evaluation criteria (SC, CH and DB), an optimal number of clusters is given (denoted by $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$). The values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ are analyzed to see how much agreement there is for an ideal number of cluster between the different internal evaluation metrics.

The ideal number of clusters may differ between the different internal evaluation metrics. This is because HAC is an unsupervised learning approach, where no prior information is presented to the clustering algorithm. This leaves the HAC algorithm on its own to establish a structure in the data. When the values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ are close, there is a lot of agreement regarding an optimal value for $\mathcal{K}$. If there is a large range between the values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$, a careful analysis of the clustering output for different values for $\mathcal{K}$ is required. This means that there is always a level of subjectiveness involved in HAC analysis. It is critical that for each value of $\mathcal{K}$, the clustering output is carefully compared against input values of the wave parameters to ensure that the found clusters are a good representation of the extreme wave characteristics in different regions of the North Sea. The optimal value for $\mathcal{K}$ is the value that yields the best representation of the HAC input values.

It should be noted that the internal evaluation metrics (SC, CH and DB) are only used to assess the performance of the HAC analysis for different values of $\mathcal{K}$ for a fixed parameter combination and weight configuration (Yanchi Liu et al., 2013). It is not possible to analyze C, SC, CH and DB to compare the clustering performance between different configurations. This is because the (1) assigned weight configuration, (2) the resulting dissimilarities, (3) linkage mechanisms and (4) the internal evaluation metrics form a closed system as points (2), (3) and (4) all depend on the assigned weight configuration(Yanchi Liu et al., 2013). The assigned weight configuration influences the calculated pairwise dissimilarities between grid points. Subsequently, the cophenetic distances (T) are determined by the linkage mechanisms. Furthermore, SC, CH and DB use the pairwise dissimilarities to assess the clustering output for different values of $\mathcal{K}$, thus are also dependent on the assigned weight configuration.

3.3.6.3. Selecting Best Clustering Output

The next step is to select the best clustering output. This is done in two steps. First, the best clustering output for each weight configuration must be determined by selecting the linkage mechanisms and $\mathcal{K}$ that yields the best clustering output. Both clustering solutions under considerations have the same weight configuration. This means that the performance of the two clustering solutions can be assessed by a comparison of the relative scores for SC, CH, and DB. Furthermore, the clustering output is directly influenced by the input values representing the wave parameters. Therefore, a careful analysis of the clustering output against the input values is required to see linkage mechanism and $\mathcal{K}$ give the best representation of the input values. The best clustering output is determined based on how well the cluster boundaries represent the differences between the HAC input values. In this research, 5 different weight configurations for each wave parameter combination are evaluated. This means that 5 different clustering outputs are obtained.

After the best clustering output for each weight configuration has been determined, the best clustering solution for each wave parameter combination must be determined. Weight can be assigned to the parameters to place more emphasis on parameters that are deemed more important, or to add additional steering to optimize the clustering output. However, too much weight can cause the clustering algorithm to miss some essential details while determining the cluster boundaries. A critical analysis of the clustering output is required to see if the resulting clusters are a good representation of the background data. In the end, for each wave parameter combination, 1 clustering solution is selected for further development (see Section 3.3.6.4).

3.3.6.4. Further Development of Clustering Output

The final clustering output is subjected to further development to improve the cluster map. Because Hierarchical Clustering is an unsupervised learning approach, the clustering output is always a direct representation of the input data. If the input values representing H_{m0}, T_p, θ_1 and θ_2 are similar for different grid points, they are assigned to the same cluster. This means that some ambiguity may arise in the clustering output that must be assessed by the modeler (Kaufman & Rousseeuw, 1990). This is involves a critical (but very subjective) assessment of the clustering output by analyzing statistical properties of the different clusters. Because the clustering output always is a direct result of the input values, grid-points in different geographical locations may be assigned to the same cluster, while there is no evidence that the statistical properties between the geographical locations are the same. Therefore, clusters in different geographical locations are preferably split. Similarly, grid points in close proximity may share similar statistical properties but are assigned to different clusters because the input

values between the grid points are different as a result of the pre-processing of the input values prior to clustering (see Section 3.3.1.2. Preferably, these clusters are merged together. The post-processing of the clustering output in this research is further explained in Section 5.1.2.

3.4. Extreme Value Analysis

The design and risk assessment of hydraulic infrastructure requires a comprehensive understanding of the extreme wave climate. Extreme value theory provides tools for modelling the stochastic behavior at unusually large or small magnitudes (Coles, 2001). Typically, this involves the estimation of events not observed within the available dataset because the available data has a limited temporal horizon with respect to the normative return period (Coles, 2001). Extreme events of interest can then be inferred by fitting an extreme value distribution to a dataset of observed events. Subsequently, unobserved extreme events can be inferred by means of extrapolation of the fitted extreme value distribution (Coles, 2001).

Most current risk assessment models rely on the assumption of stationarity when performing extreme value analysis (EVA). This is a reasonable assumption if there are no significant changes in the magnitude and frequency of the extreme events over time. However, climate change is likely to influence the frequency and magnitude of extreme events. This potentially requires the application of EVA under the non-stationary assumption. In a non-stationary model the parameters of the underlying probability distribution function change over time or in response to a physical covariate (Cheng et al., 2014; Ragno et al., 2019).

This section provides an overview of the theory of Extreme Value Analysis, both under the stationary and non-stationary approach. Furthermore, Bayesian Inference (BI) to obtain extreme value distribution parameters will be explained. Finally, Section 3.4.3 explains methods to verify the performance of extreme value distributions.

3.4.1. Extreme Value Distributions

The behavior of the extreme events can be captured by extreme value distributions. The distribution parameters of these extreme events are largely determined by the tail behavior of the extreme events (Coles, 2001). The goal is to find the extreme value distribution that describes the sampled extreme events best (Coles, 2001). In this section, the Generalized Extreme Value (GEV) distribution, Generalized Pareto (GP) distribution will be explained under both stationary and non-stationary conditions.

3.4.1.1. Generalized Extreme Value Distribution

The generalized extreme value (GEV) distribution is used to model the extreme values that have been obtained using Block Maxima. The GEV distribution is a family of continuous distributions that combines three types of extreme value distributions. In its general form, the cumulative density function (CDF) of the GEV distribution is expressed as follows (Coles, 2001; Ragno et al., 2018):

$$\Psi_{GEV}(x) = \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right\} \tag{3.26}$$

where:

$$\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right) > 0 \tag{3.27}$$

The Generalized Extreme Value distribution is flexible for modelling different behaviors of extremes using the three distribution parameters (θ, Cheng et al. (2014)):

- μ is the location parameter and specifies the center of the distribution;
- σ is the scale parameter and determines the deviations surrounding the location parameter;
- ξ is the shape parameter and determines the tail behavior of the GEV distribution. Three different cases can be identified. The limiting case of $\xi \to 0$ gives the Gumbel distribution, $\xi < 0$ gives the Weibull distribution and $\xi > 0$ gives the Fréchet distribution.

Stationarity is defined as the time-invariance of extreme events (Cheng et al., 2014; Ragno et al., 2019). The extreme events are independent and identically distributed. As a result, the GEV parameters ($\theta = (\mu, \sigma, \xi)$) have fixed values and the resulting values for $\Psi_{GEV}(x)$ are constant in time. Under non-stationary modelling, extreme events are still independent but are no longer identically distributed. θ becomes time or process dependent. That means the properties of the extreme value distribution vary with the given covariate (Cheng et al., 2014; Ragno et al., 2019). In this research only temporal covariates (t) are studied. Under non-stationary conditions, the distribution parameters become:

$$\theta(t) = \mu(t), \sigma(t), \xi(t) \tag{3.28}$$

The CDF of the non-stationary GEV distribution is defined as follows (Ragno et al., 2019):

$$\Psi_{GEV}(x \mid t) = \exp\left\{-\left(1 + \xi(t)\left(\frac{x - \mu(t)}{\sigma(t)}\right)\right)^{-\frac{1}{\xi(t)}}\right\} \tag{3.29}$$

Where:

$$\left(1 + \xi(t)\left(\frac{x - \mu(t)}{\sigma(t}\right)\right) > 0 \tag{3.30}$$

Because we are looking at non-stationary processes, the GEV parameters become dependent on temporal covariates. For example, if we consider μ to have a linear trend in time, it can be expressed as follows (Cheng et al., 2014; Coles, 2001; Ragno et al., 2019):

$$\mu(t) = \mu_1 \cdot t + \mu_0 \tag{3.31}$$

The set of distribution parameters of a GEV distribution with a non-stationary location parameter according to equation 3.31 can be expressed by $\theta = (\mu_1, \mu_0, \sigma, \xi)$. This means that there are additional distribution parameters that must be estimated compared to stationary modelling. The total number of parameters to be estimated depends on the parameters that are assumed to be non-stationary, and the assumed temporal trend of said parameters (e.g. linear or polynomial, (Cheng et al., 2014; Ragno et al., 2019)).

3.4.1.2. Generalized Pareto Distribution

The Generalized Pareto (GP) Distribution is used for modelling extreme values that have been obtained using Peak over Threshold (POT) analysis. Consider a set X of independent extreme observations exceeding a threshold (u). The CDF of the exceedances ($y_e = x - u$), can be expressed as follows (Antonini et al., 2019; Pickands, 1975; Ragno et al., 2019):

$$\Psi_{GP}(y_e) = 1 - \left(1 + \xi \cdot \left(\frac{y_e}{\sigma}\right)\right)^{-\frac{1}{\xi}} \tag{3.32}$$

In equation 3.32, y_e is the exceedance of x above u ($y_e | x > u = x–u$), σ is the scale parameter and ξ is the shape parameter. Similar to the GEV distribution, ξ determines the tail behavior of the GP distribution (Coles, 2001). If $\xi < 0$, the GP distribution of the exceedances has an upper bound of $u–\sigma/\xi$. In the case that $\xi > 0$, the GP distribution has no upper limit. And finally, for the limiting case that $\xi \to 0$, the GP distribution can be rewritten as:

$$\Psi_{GP}(y_e) = 1 - exp\left(\frac{1}{\sigma}\right) \tag{3.33}$$

This corresponds to an exponential distribution with parameter $1/\sigma$. In the non-stationary GP distribution, both the threshold value (y_e) as well as the distribution parameters (σ and ξ) can be expressed as a function of a temporal covariate. The CDF of the non-stationary GP distribution is then defined as (Ragno et al., 2019):

$$\Psi_{GP}(y_e \mid t) = 1 - \left(1 + \xi(t) \cdot \left(\frac{y_e(t)}{\sigma(t)}\right)\right)^{-\frac{1}{\xi(t)}} \tag{3.34}$$

3.4.2. Bayesian Inference

In this research, The parameters of the extreme value distributions will be inferred using Bayesian Inference (BI). The Matlab toolbox ProNEVA (Ragno et al., 2019) applies Bayesian Inference techniques to infer the extreme value distributions under both stationary and non-stationary conditions. Bayesian Inference combines prior knowledge about the extreme value distribution parameters and a set of additional observations (Cheng et al., 2014) to update its belief about the extreme value distribution. Bayesian inference, as the name suggests, makes use of Bayes' theorem, which in its general form can be expressed as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{3.35}$$

Simply explained, Bayes' theorem is a mathematical equation used for updating probabilities when new observations or information becomes available. Within this framework, the interest lies in estimating the distribution parameters (θ) of the Extreme Value distributions given a set of observations (Y_{obs}). It is possible to rewrite Bayes' Theorem as follows:

$$p\left(\theta \mid Y_{obs}\right) = \frac{p\left(\theta\right) \cdot p\left(Y_{obs} \mid \theta\right)}{p\left(Y_{obs}\right)} \tag{3.36}$$

In equation 3.36, θ denote the parameters of the extreme value distribution, and Y_{obs} is a set of observations. Within equation 3.36, we distinguish the following terms (for more information, see Figure C.1):

- $p\left(\theta\right)$ is the prior. It is the probability of the parameters θ and represents prior knowledge about the parameters of the extreme value distribution;
- $p\left(Y_{obs} \mid \theta\right)$ is the likelihood, and can conceptually be described as the probability of occurrence of the observations (Y_{obs}) given the distribution parameters θ;
- $p\left(Y_{obs}\right)$ is the marginal likelihood of the observations, often also referred to as "evidence". It is the overall probability of the newly obtained information. Within this framework, this will be neglected because it is assumed that it is a constant value (Ragno et al., 2019);
- $p\left(\theta \mid Y_{obs}\right)$ is the posterior. The posterior represents the updated probability and confidence interval of the distribution parameters θ given an updated set of observations or information Y_{obs}.

Given the fact that the marginal likelihood $p\left(Y_{obs}\right)$ is neglected, equation 3.36 can be rewritten as:

$$p\left(\theta \mid Y_{obs}\right) \propto p(\theta) \cdot p\left(Y_{obs} \mid \theta\right) \tag{3.37}$$

Assuming independence between observations, Bayes theorem for the estimation of the non-stationary GEV distribution with a linear location parameter (See Equation 3.31) can be written as follows (Cheng et al., 2014; Ragno et al., 2019):

$$p\left(\mu_1, \mu_0, \sigma, \xi \mid Y_{obs}\right) \propto p\left(\mu_1\right) p\left(\mu_0\right) p(\sigma) p(\xi) \cdot \prod_{i=1}^{N} p\left(y_i \mid \mu_1, \mu_0, \sigma, \xi\right) \tag{3.38}$$

In Equation 3.38, $\theta = (\mu_1, \mu_0, \sigma, \xi)$ are the distribution parameters of the non-stationary GEV distribution. The stationary case can be treated as a special case of equation 3.38, but without the dependence on a covariate, and the equation becomes (Cheng et al., 2014):

$$p\left(\mu, \sigma, \xi \mid Y_{obs}\right) \propto p(\mu) p\left(\sigma\right) p(\xi) \cdot \prod_{i=1}^{N} p\left(y_i \mid \mu, \sigma, \xi\right) \tag{3.39}$$

The resulting posterior distributions $p\left(\mu_1, \mu_0, \sigma, \xi \mid Y_{obs}\right)$ and $p\left(\mu, \sigma, \xi \mid Y_{obs}\right)$ provide information about the distribution parameters under non-stationary and stationary conditions respectively.

3.4.2.1. Bayesian and Frequentist Approach for Parameter Estimation

Within the framework of EVA, it is important to make a clear distinction between the Frequentist approach and the Bayesian approach in the estimation of distribution parameters. Within the Frequentist approach, the distribution parameter of the extreme value distribution has a fixed value. No probabilities can be assigned to describe its uncertainty (Bickel & Lehmann, 2012). The value of the mentioned distribution parameter is also known as the maximum likelihood estimate (MLE). The MLE of the distribution parameter is equal to the sample parameter.

The Bayesian approach agrees that the distribution parameter (e.g. the location parameter) of the extreme value distribution has a fixed but unknown value. The uncertainty surrounding its true value can be represented using a probabilistic distribution (Cheng et al., 2014). An assumption is made about the possible values of the distribution parameters by defining a prior probability distribution, and sample data is used to update the extreme value distribution (likelihood). In the Bayesian setting, Bayes's theorem is applied to the newly obtained data to make probability distribution over the unknown parameter narrower (the posterior). In the end, this results in a probabilistic distribution of the GEV distribution parameters.

3.4.3. Model Checking

The performance of extreme value distribution must be assessed to determine their applicability to accurately project extreme return levels. The lack of data with respect to the normative return periods and the requirement of extrapolation is a fundamental difficulty when assessing the performance of the extreme value distributions (Coles, 2001). However, assessment of the characteristic GEV distributions can be made with reference to the observed data. To this end, Quantile-Quantile plots (QQ-plots) and goodness of fit (GOF) tests including the Bias and Root Mean Square Error (RMSE) are used to assess the performance of the GEV model. Furthermore return level plots are used to study the expected return level estimates of extreme H_{m0} for long return periods.

3.4.3.1. Quantile-Quantile (QQ) plots

Quantile-Quantile (QQ) plots compare the performance of the fitted extreme Value distribution against the extreme observations (Coles, 2001). This allows us to study how well the theoretical extreme value distribution describes the extreme observations. Consider two sets of extreme observations. The first set includes N extreme observations derived using Annual Maxima, denoted by z_i. The second set includes K observations derived using peak over threshold, denoted by y_j. The extreme data can be ordered so that $z_1 \leq z_2 \leq \cdots \leq z_N$ and $y_1 \leq y_2 \leq \cdots \leq y_K$. Subsequently, the empirical cumulative distribution functions (CDF) $\tilde{G}$ and $\tilde{H}$ of the extreme observations is given as follows (Coles, 2001):

$$\tilde{G}\left(z_{(i)}\right) = i/(N+1) \tag{3.40}$$

$$\tilde{G}\left(y_{(i)}\right) = j/(K+1) \tag{3.41}$$

For the QQ-plot, the inverse of the fitted GEV or GP model ($\hat{G}^{-1}$ and $\hat{H}^{-1}$ respectively) is used to compute the theoretical extreme values corresponding to the empirical CDFs $\tilde{G}$ and $\tilde{H}$. The inverse CDFs of the GEV and GP distribution can mathematically represented by the following equations (Coles, 2001).:

$$\hat{G}^{-1}\left(\frac{i}{N+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\log\left(\frac{i}{N+1}\right)\right\}^{-\hat{\xi}}\right] \tag{3.42}$$

$$\hat{H}^{-1}(\frac{j}{K+1}) = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[y^{-\hat{\xi}} - 1\right] \tag{3.43}$$

The QQ-plot can then be constructed by plotting the theoretical extreme values as derived by $\hat{G}^{-1}$ and $\hat{H}^{-1}$ against the extreme observations z_i and y_j as given in Equation 3.42. If the fitted extreme value models are a reasonable estimate of the extreme wave climate, the QQ-plots show points close to the diagonal reference line (Coles, 2001). Departures from linearity in the quantile plot could represent model failure, in the sense that $\hat{G}^{-1}$ $\hat{H}^{-1}$ may either under- or overestimate the extreme values for H_{m0}.

$$\left\{\left(\hat{G}^{-1}(i/(N+1)), z_{(i)}\right), i = 1, \ldots, m\right\} \tag{3.44}$$

$$\left\{\left(\hat{H}^{-1}(j/(K+1)), y_{(j)}\right), j = 1, \ldots, K\right\} \tag{3.45}$$

3.4.3.2. Return Level Plots

Return level plots can be used to study the expected return level estimates of extreme H_{m0} for long return periods. For extreme events selected using Annual Maxima the return period ($\hat{R}$) can be calculated from $\hat{G}$ using the following equation:

$$\hat{R_{GEV}} = \frac{1}{1 - \hat{G}} \tag{3.46}$$

For the extreme value models derived using POT analysis (i.e. GP distributions), this is a little less straightforward. POT analysis results in an unequal distribution of the number of extreme events selected for each year. An approximation can be made by including the average annual threshold exceedances (N_{exc}) as follows:

$$\hat{R_{GP}} = \frac{1}{N_{exc}(1 - \hat{H})} \tag{3.47}$$

In equations 3.46 and 3.47, $\hat{R_{GEV}}$ and $\hat{R_{GP}}$ are the return period of extreme for the theoretical GEV and GP distributions respectively.

3.4.3.3. Goodness of Fit Equations

The performance of the modelled extreme value distributions can be analysed using several goodness of fit tests, including bias and Root Mean Square Error (RMSE). Bias is the discrepancy between the theoretical extreme values and the observation data. Bias is used to determine whether Extreme value model tends to over- or underestimate the data compared to the observations. The Root Mean Square Error (RMSE) is employed to reflect the standard deviation of the differences between the extreme value distribution and the extreme observations. Bias and RMSE can be calculated using the following equations (here only given for the GEV distribution):

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{G}^{-1}\left(\frac{i}{N+1}\right) - z_i\right) \tag{3.48}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{G}^{-1}\left(\frac{i}{N+1}\right) - z_i\right)^2} \tag{3.49}$$

In equations 3.48 and 3.49, N is the number of extreme observations included to fit the extreme value distribution.

3.5. Extreme Value Analysis on Cluster Level

This section explains the derivation of the representative extreme value models describing the extreme significant wave height (H_{m0}) at cluster level. Before performing the (non-stationary) extreme value analysis at cluster level, characteristic time-series describing extreme H_{m0} at cluster level must be obtained. This includes the detrending of the extreme H_{m0} observations and the generation of 5 different representative time-series. Subsequently, Bayesian Inference (BI) is employed to infer the parameters of the Generalized Extreme Value (GEV) distribution corresponding to the detrended H_{m0} observations. Subsequently, potential non-stationarity of extreme H_{m0} is accounted for by adding the temporal trend information to the location parameter. This makes the location parameter of the GEV distribution a function of time.

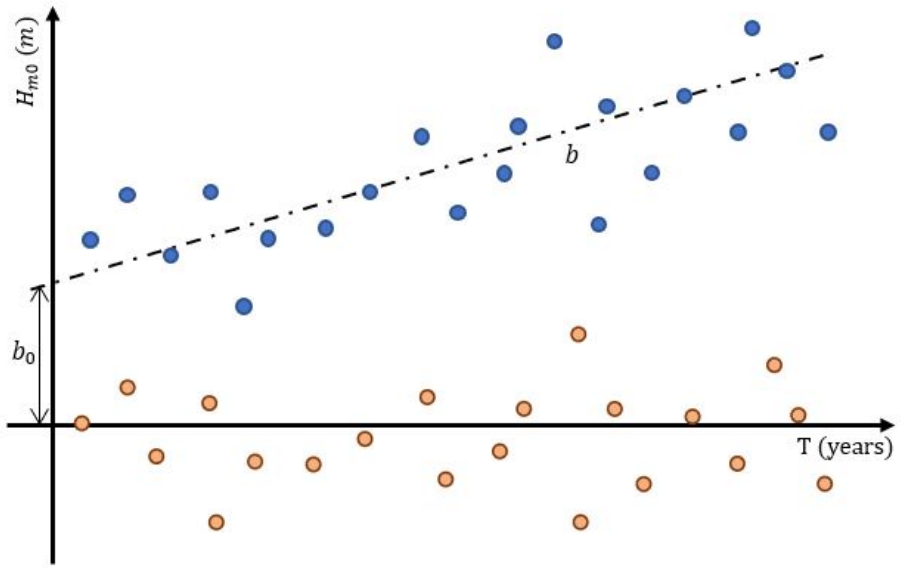


Figure 3.9: Visualization of detrending AM H_{m0} observations. The Theil Sen parameters b and b_0 are used to fit a linear regression line which is subtracted from the AM H_{m0} observations to arrive at H^*_{m0}.

3.5.1. Pre-Processing of Data at Grid Points

Within the considered cluster, the time-series at the grid point level have been filtered according to the dominant wave direction. This ensures that the selected extreme events are both homogeneous and independent. The performed hierarchical clustering analysis ensures that all intra-cluster grid point have similar values for the dominant wave direction (θ_1). Subsequently, the dominant range surrounding θ_1 is identified from which the extreme H_{m0} observations are selected. The relevant range of the dominant wave directions (i.e. upper and lower bounds of θ_1) are given in Table 5.4.

Annual Maxima (AM) was applied to select the extreme H_{m0} events. This was done for several reasons. Extreme H_{m0} values selected using AM are better suited for the quantification of long-term temporal trends than those selected using POT because POT requires the careful selection of a threshold. For more information, see Discussion. Furthermore, the POT approach results in different numbers of exceedances for every year, thereby posing further challenges for the temporal trend analysis. Annual maxima (AM) ensures that one extreme sample per year is considered at each grid point (Laface et al., 2016; Liang et al., 2019).

Before constructing representative time-series, the AM H_{m0} data must be detrended at each of the grid points. Despite the fact that hierarchical clustering analysis is performed to identify clusters with similar extreme H_{m0} characteristics, the intra-cluster range for extreme H_{m0}. This is a direct result of the pre-processing of the HAC input values for H_{m0} prior to clustering. To ensure that the intra-cluster AM H_{m0} observations are homogeneous, detrending is required. Detrending the annual maxima removes any long-term trend and forces the mean for intra cluster AM H^*_{m0} observations to be approximately equal.

Detrending AM H_{m0} is performed under the assumption that potential non-stationary of extreme H_{m0} can be described by a linear trend. The temporal trends are calculated at grid-point level by fitting the Theil-Sen (TS) estimator b and intercept b_0 (see Section 3.2.2) to the AM H_{m0} data. Subsequently, the temporal trend for AM H_{m0} (y_{Hm0}) can be described by the following linear relation:

$$y_{Hm0}(t) = b_0 + b \cdot t; \tag{3.50}$$

In this equation, t is the time in years. This chapter only considers the full length of the wave data, i.e. spanning between 1950 and 2020. The AM H_{m0} values are detrended by subtracting the trend y_{Hm0}, resulting in H^*_{m0}.

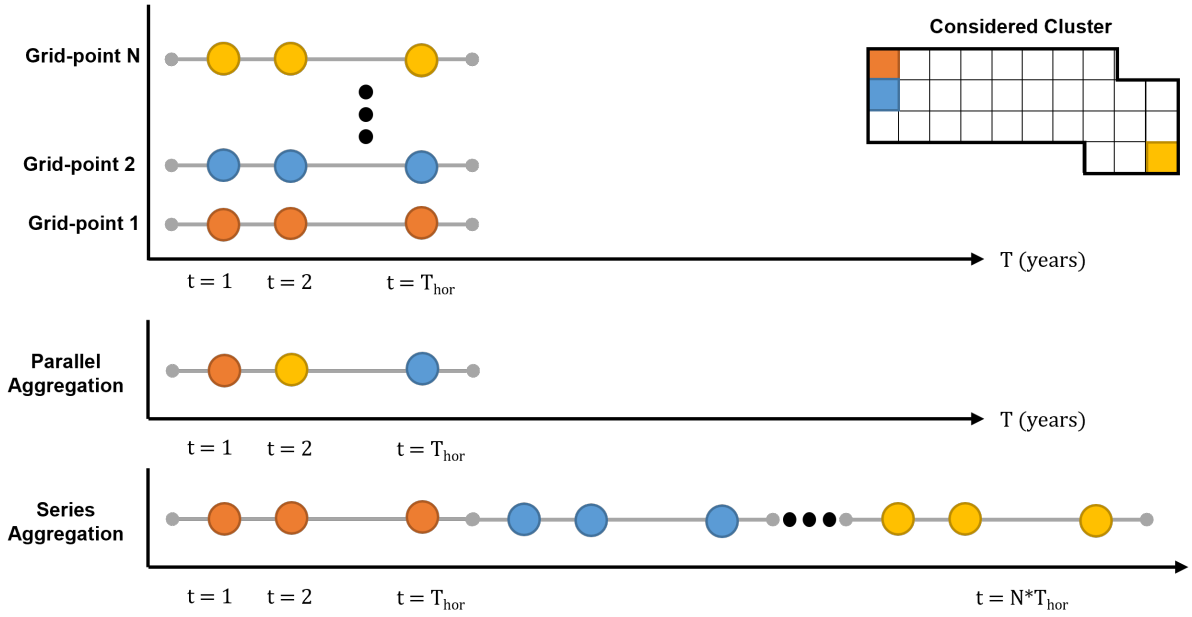


Figure 3.10: Illustration visualizing Series and Parallel Aggregation for a hypothetical cluster containing N_{gp} grid points. For series aggregation, the N_{gp} sets of annual maxima are placed behind each other. For parallel aggregation, 1 observation is selected for each time-stamp to construct the representative time-series.

3.5.2. Constructing Representative Time Series

The H^*_{m0} observations at the grid-points are evaluated to select characteristic datasets that describe the extreme H_{m0} at the cluster level. Five methods to construct representative time-series for the clusters will be evaluated in this research. Three aggregation methods will be evaluated to see how intra-cluster extreme wave data can be pooled to generate characteristic datasets. Two methods involve grid points selection to study whether the extreme wave data from a single grid-point is sufficient to describe the wave climate at cluster level. The following sections elaborate on the different methods.

3.5.2.1. Series Aggregation (Method 1)

The first method to generate representative time-series is series aggregation. In series aggregation, the H^*_{m0} series for all intra-cluster grid points are placed ”behind each other”, thereby creating a single time-series consisting of $N_{gp} * T_{hor}$ observations (See figure 3.10), where N_{gp} is the number grid points in the considered cluster, and T_{hor} is the considered temporal horizon expressed in years. The benefit of this aggregation method is that it contains all detrended values for AM H^*_{m0}, thereby maximizing the available knowledge about extreme waves in the cluster.

3.5.2.2. Parallel Aggregation (Methods 2 and 3)

The second and third method to compute Representative H^*_{m0} time-series of extreme is parallel aggregation. In parallel aggregation, a single value for H^*_{m0} from each year is selected to construct the dataset (See figure 3.10). For aggregation method 2, the maximum annual H^*_{m0} value is selected. And for aggregation method 3, the median H^*_{m0} value between all intra-cluster grid points is selected. For both methods 2 and 3, the assumption is made that the selected AM H^*_{m0} observations at the grid-points occurred during the same storms and therefore the relative magnitude of the extreme wave events after detrending remains constant between the different grid points.

3.5.2.3. Grid-point of clustering Centroid (Method 4)

The fourth method involves the selection of extreme wave data from a single grid-point. This method assumes that the time-series of H_{m0} at the grid point corresponding to or closest to the clustering centroid is representative for modelling extreme H_{m0} at cluster level. This method is adopted from Goharnejad et al. (2022), who obtained promising results using this method to assess the extreme wave climate at cluster level in the North Atlantic Ocean. See Figure 3.11 for an illustration how this representative time-series is selected.

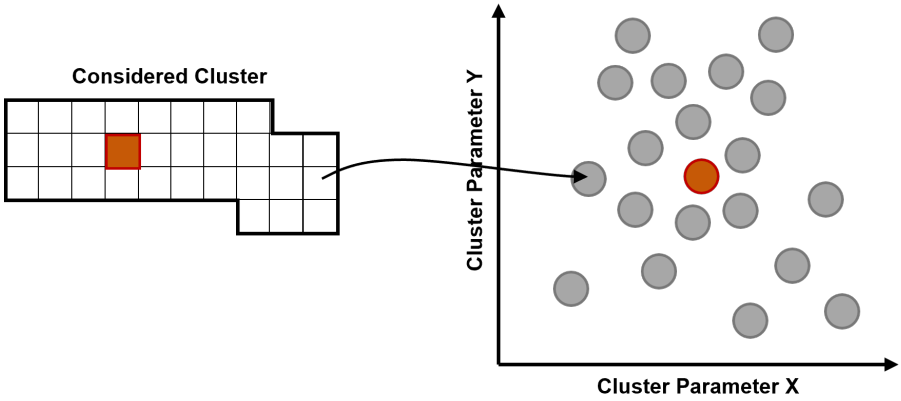


Figure 3.11: Illustration visualizing the selection of the grid point that yields the clustering centroid. Consider a cluster from a HAC analysis based on two cluster parameters, X and Y. The cluster variables can be standardized and plotted against each other. The selected grid point is the grid point that yields the centroid (given in red).

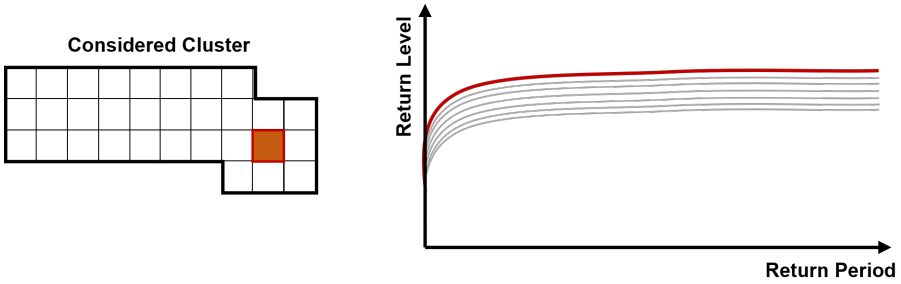


Figure 3.12: Illustration visualizing the selection of the grid point that yields the maximum return levels for the detrended values of AM H_{m0}. The return level plots can be plotted for all N_{gp} grid points. The selected grid point in this analysis is the grid point that yields the highest return levels.

3.5.2.4. Grid-point with Highest Return Levels (Method 5)

The fifth and final method is to select the grid-point that yields the highest return values for the GEV distributions of H_{m0} before detrending to describe extreme H_{m0} at cluster level. The resulting return levels for the model derived using this representative time-series are expected to remain relatively high compared to the individual intra-cluster grid points after deriving the extreme value model. This allows this extreme value model to yield safe estimates regarding extreme H_{m0} conditions when used for the design and risk assessment of hydraulic infrastructure.

3.5.3. Extreme Value Analysis using Bayesian Inference

The representative time-series of AM H^*_{m0} are subsequently subjected to Extreme Value Analysis (EVA) to determine the Generalized Extreme Value (GEV) distribution parameters ($\theta = \{\mu, \sigma, \xi\}$). Because H^*_{m0} concerns detrended data, the GEV parameters represent the detrended data, and will be denoted by $\theta^* = \{\mu^*, \sigma^*, \xi^*\}$. The parameters for θ^* will be inferred using Bayesian Inference (BI). See Section 3.4.2 for more information. BI is preferred over Maximum Likelihood Estimation (MLE) because BI allows the construction of informative priors based on extreme value information from the grid points that share similar extreme wave characteristics. This provides more accurate extreme value models than using MLE (Rickets, 2021). Furthermore, BI is selected because the posterior distribution allow for straightforward uncertainty quantification of the GEV parameters. This section presents the BI process applied in this research including the appropriate informative priors ($P(\theta)$) and analysis of the posterior distributions. The aggregated H^*_{m0} datasets are used as the observations in the BI process.

3.5.3.1. Priors for GEV parameters

For the extreme value analysis of H^*_{m0}, appropriate prior distributions must be selected. This research adopts its approach to determine informative priors from Antonini et al. (2019) and Raby et al. (2019).

The aim is to combine the information about θ^* at the intra-cluster grid points to construct informative normal priors. This makes it possible to maximize the available prior information about the extreme value distribution, thereby enhancing the performance of the extreme value model at cluster level. The general assumption is that within the cluster, the properties of θ^* at each grid point are similar compared to neighboring grid points. As a result, the values for θ^* vary smoothly throughout the cluster.

At each of the intra-cluster grid-points, the values for H^*_{m0} have been examined. The GEV distribution has been fitted to the H^*_{m0} observations at each grid point by means of maximum likelihood estimation (MLE), resulting in N_{gp} values for μ^*_{gp}, σ^*_{gp} and ξ^*_{gp} within the considered cluster. Note that the subscript gp refers to grid points, and not to the Generalized Pareto distribution. The values for μ^*_{gp}, σ^*_{gp} and ξ^*_{gp} are subsequently used to determine the mean and standard deviations of the normal prior distributions. In Antonini et al. (2019) and Raby et al. (2019), the distance between the grid points and an extraction point of interest is considered to assign weight to θ^*_{gp} before fitting the normal prior distribution. In this study, all values for μ^*_{gp}, σ^*_{gp} and ξ^*_{gp} at the grid points are assigned equal weight. The reason for this is that information from multiple grid points is combined in the aggregation process, and therefore appropriate weights cannot be assigned.

3.5.3.2. ProNEVA for Bayesian Inference

The Bayesian Inference procedure will be performed using MATLAB toolbox ProNEVA developed by Ragno et al. (2019). ProNEVA applies a Hybrid Evolution Markov Chain Monte Carlo (HE-MCMC) algorithm (Ragno et al., 2019; Sadegh et al., 2017). The (HE)-MCMC approach for obtaining the posterior distribution of parameters has become increasingly popular and is used in several studies of extremes (Cheng et al., 2014). More information about MCMC and how ProNEVA applies MCMC to infer distribution parameters is presented in Appendix C. A detailed sensitivity analysis studying the effects of the HE-MCMC settings on the resulting GEV distribution parameters lies outside the scope of this thesis. Nevertheless, it is important to carefully select the settings of the HE-MCMC algorithm to obtain optimal results. For each evaluation of the prior and observation distributions, 5 chains are selected to run in parallel. The total number of iterations for each evaluation is equal $10\,000$. The considered burn-in period is equal to $1\,000$ iterations, meaning that the first $1\,000$ iterations of each run are neglected and the subsequent analysis is based on the remaining $9\,000$ iterations for each chain, resulting in a total of $45\,000$ samples forming the posterior distributions ($P\left(\theta^* \mid H^*_{m0}\right)$) of the GEV distribution parameters.

3.5.3.3. Analyzing the Posterior Distributions

The posterior ($P(\theta^* \mid H^*_{m0})$) distributions are subsequently analyzed to determine the GEV distribution parameters (for $\theta^* = \{\mu^*, \sigma^*, \xi^*\}$. In contrast to Maximum Likelihood Estimation (MLE), the GEV distribution parameters inferred using Bayesian Inference must be derived from the posterior distributions (see Section 3.4.2.1). The median values the posterior distributions are selected to represent the best estimate of the GEV distribution parameters. Furthermore, the 5 and 95 percentile values of the posterior distributions are used to the derive the GEV distribution parameters that yield the 90% confidence interval.

3.5.4. Including Non-Stationarity in GEV Distribution

In this research, we account for potential non-stationarity of H_{m0} by adding a linear component that represents the temporal trend to μ^*. This essentially results in a horizontal shift of the cumulative distribution function (CDF) over time (See figure 3.13). The expression of CDF of the GEV distribution for H^*_{m0} is given as follows:

$$H^*_{m0} \sim GEV\left(\mu^*, \sigma^*, \xi^*\right) \tag{3.51}$$

$$\Psi^*_{GEV}(H^*_{m0}) = \exp\left\{-\left(1+\xi^*\left(\frac{(H^*_{m0}-\mu^*}{\sigma^*}\right)\right)^{-\frac{1}{\xi^*}}\right\} \tag{3.52}$$

In equation 3.52, μ^*, σ^* and ξ^* are the location, scale and shape parameters respectively, derived using Bayesian Inference. The CDF of the non-stationary GEV model is then computed by adding a linear component representing the temporal trend of extreme H_{m0} to μ^*. This causes the GEV distribution to shift horizontally with time (See figure 3.13). From Equation 3.50, we know that:

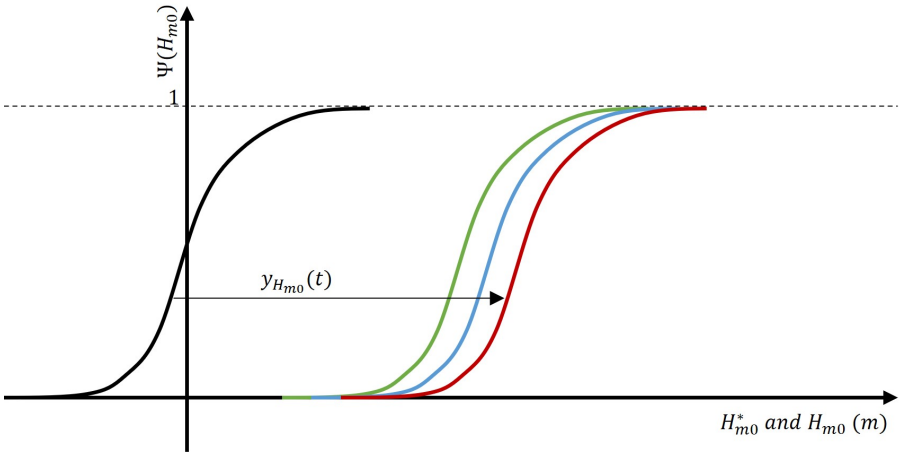


Figure 3.13: Illustration showing the reinstatement of the trend back in to the GEV distributions. Black line illustrates $\Psi^*_{GEV}(H^*_{m0})$. The green, blue and red line illustrate $\Psi_{GEV}(H_{m}0)$ for the years, 1950, 1990 and 2020 respectively.

$$H_{m0} = H^*_{m0} + y_{Hm0}(t) = H^*_{m0} + (b \cdot t + b_0) \tag{3.53}$$

For simplicity reasons, this research only considers non-stationarity of the location parameter. Non-stationarity for the scale parameter (σ^*) is not considered because it is assumed that extreme deviations for H_{m0} do not change with time. The shape parameter (ξ^*) are also considered to be fixed in time in this research. There is no reason to believe that the extreme distributions will switch from bounded to un-bounded distributions (or vice versa). Equation 3.53 can be substituted into Equation 3.52, yielding the following equation for the GEV distribution for H_{m0}:

$$H_{m0} \sim GEV\left(\mu^* + (b \cdot t + b_0), \sigma^*, \xi^*\right) \tag{3.54}$$

$$\Psi_{GEV}(H_{m0} \mid t) = \exp\left\{-\left(1 + \xi^*\left(\frac{(H_{m0} - (\mu^* + b \cdot t + b_0))}{\sigma^*}\right)\right)^{-\frac{1}{\xi^*}}\right\} \tag{3.55}$$

Finally, the inverse CDFs for H^*_{m0} and H_{m0} can be written as follows (adopted from (Coles, 2001)):

$$\Psi^{-1*}_{GEV}(\Psi) = \mu^* - \frac{\sigma^*}{\xi^*}\left[1 - \{-\log(\Psi)\}^{-\xi^*}\right] \tag{3.56}$$

$$\Psi^{-1}_{GEV}(\Psi, t) = \mu^* - \frac{\sigma^*}{\xi^*}\left[1 - \{-\log(\Psi)\}^{-\xi^*}\right] + b \cdot t + b_0 \tag{3.57}$$

Using Equation 3.55, return level plots for year of interest can be inferred by selecting the appropriate value for t. The selected value for t must be between 0 (corresponding to the start of the considered temporal horizon) and T_{hor} (the length of the temporal horizon). For this, the assumption is made that the temporal evolution of extreme H_{m0} can described using a linear trend defined by b and b_0.

3.5.4.1. Updated stationarity

Under non-stationary models, the non-stationarity of the distribution parameters is applied to extrapolate design values throughout the evaluation period (e.g. the considered lifetime of a structure). Under updated stationary models (UST, Luke et al. (2017)), the values of the GEV parameters corresponding to the end of the observation period are used to project extreme H_{m0} return levels under the assumption that H_{m0} can be described by a stationary process after the observation period has ended. UST models show promise in cases where physical changes have affected the extreme wave climate over the observation period or when the future non-stationary behavior of the design parameter is uncertain (Luke et al., 2017).

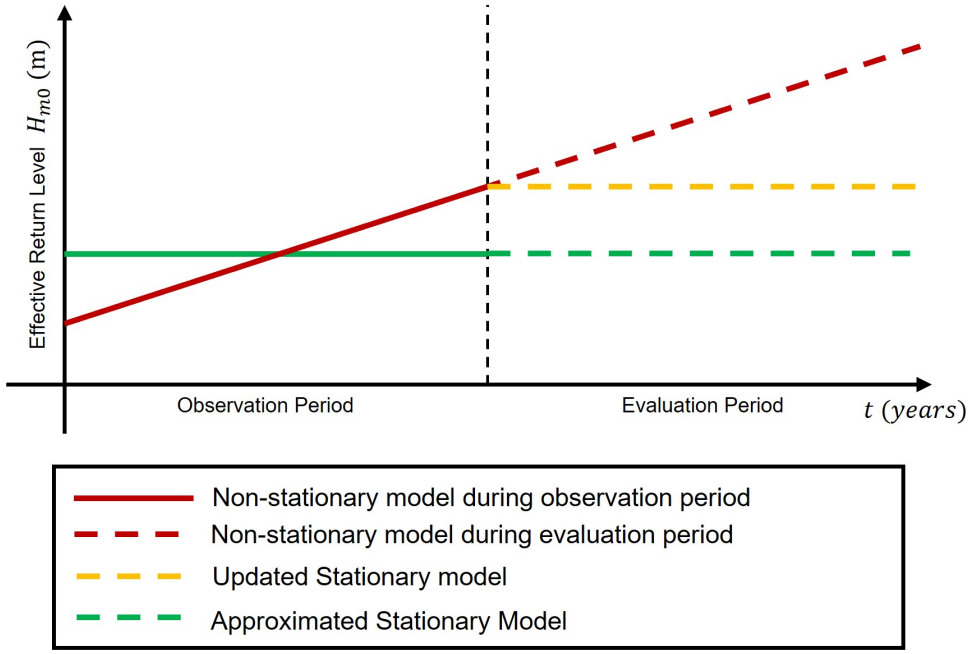


Figure 3.14: Illustration showing the temporal evolution of the effective return levels for non-stationary (red), updated stationary, and approximated stationary modelling.

3.5.4.2. Approximated Stationarity

When the applicability of non-stationary models in certain, stationary modelling is potentially preferred. Applying the derived representative model is challenging because the value of the location parameter changes over time (t). To derive return level return period curves as in the stationary case, this research proposes approximated stationarity (AST). In AST modelling, the GEV parameters corresponding to the median of the observation period are used to approximate H_{m0} return levels for the evaluation period under the assumption of stationarity. The major assumption is the median value of the linear trend in effective return levels during the observation period can be used to derive design values under stationary conditions. This means that applying t corresponding to the median of the observation period describes the average conditions of extreme H_{m0} return levels over the observation period. If H_{m0} does not exhibit significant trends, then the AST model are expected to project extreme H_{m0} return levels that are very close to the return levels as derived using conventional stationary modelling.

3.5.4.3. Non-Stationary Parameters

The question that remains is how to select the appropriate values for b and b_0 for the five representative time-series. In this research, the values for b and b_0 are selected in accordance with the method used to derive the representative time-series. That means that for the five representative time-series, the following values for b and b_0 are selected:

1. For series aggregation, the median values of b and b_0 between all intra-cluster grid points are selected;
2. For parallel maximum aggregation, the maximum values of b and b_0 between all intra-cluster grid points are selected;
3. For parallel median aggregation, the median values of b and b_0 between all intra-cluster grid points are selected;
4. For the clustering centroid, the values of b and b_0 corresponding to the clustering centroid are selected;
5. For the grid point that yields the maximum return level (before detrending), the corresponding values for b and b_0 to the grid point are selected.

4

Exploratory Analysis North Sea

This chapter presents an exploratory analysis of the North Sea. Section 4.1 provides an overview of the ERA5 Reanalysis dataset that provides time-series for wave and wind parameters at over 1200 grid points in the research domain. This includes a quality control check assessing the applicability of ERA5 in light of this research. Furthermore, temporal trend analyses are performed for the extreme significant wave height (Section 4.2.2) and extreme wind speeds (Section 4.3) to get a better understanding about non-stationarity of waves and wind in the research domain. Finally, Section 4.4 computes the input values for the wave parameters of interest in the Hierarchical Agglomerative Clustering (HAC) analysis.

The significant wave height (H_{m0}) is the main variable of interest in this research. It is the most important parameter in design and risk assessment of hydraulic infrastructure and is included in numerous failure modes. Furthermore, H_{m0} is the parameter for which extreme value models at cluster level are derived. Therefore, most analyses presented in this chapter are focused on H_{m0}. This includes the quality control of the ERA5 Reanalysis dataset and the temporal trend analysis in Section 4.2. The peak period (T_P) and the mean wave direction (θ) are not considered in the quality control check and temporal trend analysis (see also Section 1.2. However, they are included in the HAC analysis as it enables clustering results based on a more complete representation of the extreme wave climate in the North Sea domain.

4.1. ERA5 Reanalysis Dataset

For the analysis of wave and wind parameters in the North Sea, time-series were obtained through ERA5 Reanalysis (Hersbach et al., 2020). ERA5 is the fifth generation atmospheric and oceanic reanalysis of the global climate produced by the European Center of Medium Range Weather Forecasts (ECMWF). ERA5 provides hourly estimates for a large number of atmospheric, oceanic and land-surface parameters between 1950 and the present (Hersbach et al., 2020). Currently the ERA5 reanalysis data is split into two parts, with a dataset that spans the period between 1979 and the present and a back extension of the data between 1950 and 1978. The ERA5 reanalysis data is available on a longitude-latitude grid providing spatial resolution of 0.25 degrees for atmospheric data and 0.5 degrees for ocean wave parameters. The back extension is only available as a preliminary version, since it encounters problems with tropical cyclones around Australia and the Equator (Hersbach et al., 2020). However, the data in the rest of the world (including the North Sea) is reliable and therefore will be used in this research. A higher resolution for the ocean wave data can be achieved by interpolation (Wang et al., 2021). in this research, the resolution of the ocean wave data is interpolated to match the spatial resolution of the atmospheric data.

Reanalysis combines global observations with numerical model data to construct globally complete and consistent dataset that respect the laws of physics (Hersbach et al., 2020). In the present, forecasts are combined with newly available observations to produce a new best estimate of the state of the atmosphere. This is called analysis. Using this technique, the ECMWF models can be updated to give the best description of the atmospheric and oceanic climate. Reanalysis works in much the same way.

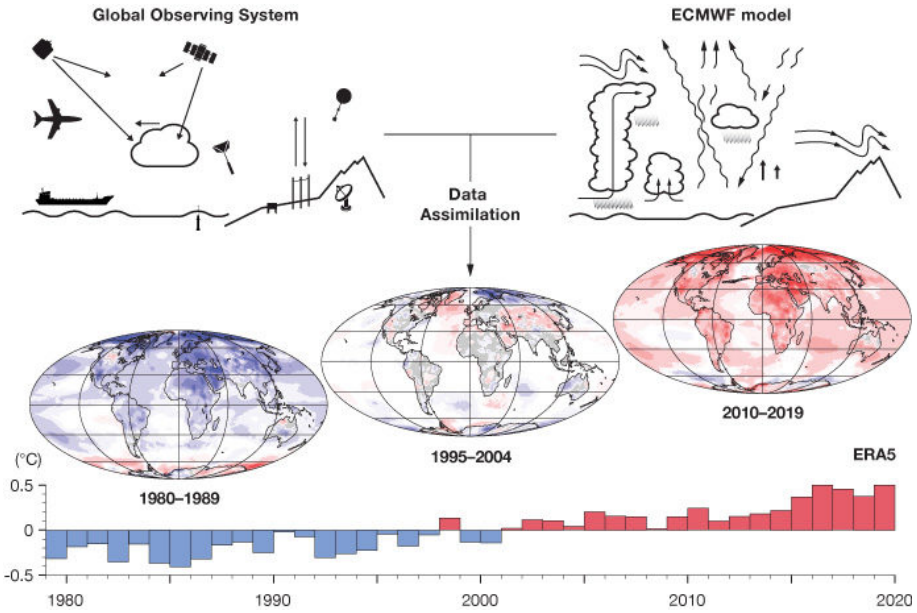


Figure 4.1: Schematic overview of ERA5 Reanalysis data assimilation (Hersbach et al., 2020).

The modern forecasting numerical models are used to connect historical observations at a point in time, thereby producing a reconstruction of the historical global climate (Hersbach et al., 2020). A schematic overview of the Reanalysis process is given in Figure 4.1.

For this research, various parameters describing the wave and wind climate in the North Sea have been acquired. An overview of the considered parameters is presented in Table 4.1. The boundaries of the dataset have been determined based on the boundaries of the North Sea (see Figure 4.2). The Northern and Western boundary has been set along the Shetland islands. The Eastern boundary has been set such that it includes the Skagerrak. Finally, the Southern boundary has been set such that it includes the English channel. This way, it is possible to have a better understanding of the extreme wave conditions between Dover and Calais. The extreme wave climates in the Atlantic Ocean and the North Sea are likely to differ, resulting in complex extreme wave characteristics in the English Channel. A temporal horizon spanning between 1950 and 2020 is considered in this research. To speed up processing, the wave and wind data have been filtered such that the maximum H_{m0} observations at 4 hour intervals are obtained. subsequently, the corresponding values for $T_{m-1,0}$, T_P, θ, u_{10} and v_{10} are obtained. The oceanic data in ERA5 reanalysis has a spatial resolution of 0.5 degrees in longitude and latitude (Hersbach et al., 2020). A higher resolution has been achieved by interpolation (Wang et al.,

Table 4.1: Overview of wave and wind parameters used in this research that were retrieved from the ERA5 Reanalysis dataset. The variable column corresponds to the variable name given by the ECMWF. The symbol corresponds to the nomenclature used in this research.

Variable	Symbol	Unit	Interpolation	Description
SWH	H_{m0}	m	yes	Significant wave height from wave energy spectrum
MWP	$T_{m-1,0}$	s	yes	Mean wave energy period from wave energy spectrum
PP1D	T_P	s	yes	Peak Wave Period derived from wave energy spectrum
MWD	θ	^{o}N	yes	Mean incoming wave direction, mean over all frequencies and directions of the two-dimensional wave spectrum
U10	u_{10}	m/s	no	Eastward component of 10m elevation windspeed
V10	v_{10}	m/s	no	Northward component of 10m elevation windspeed

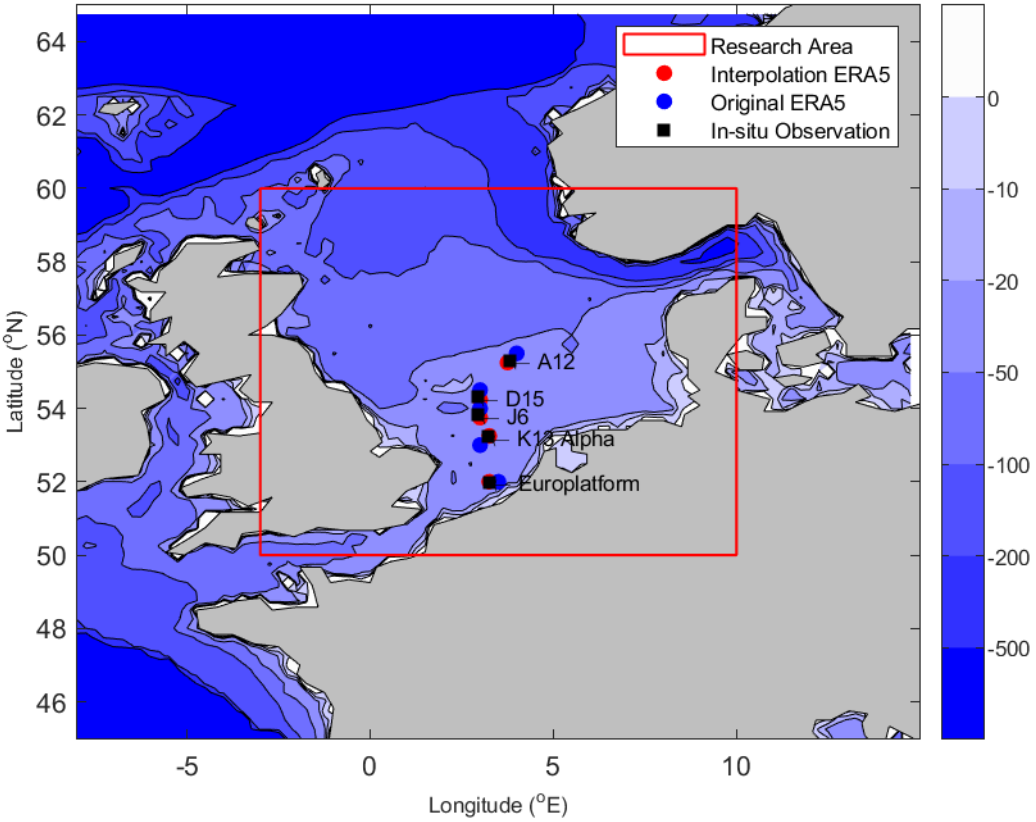


Figure 4.2: Bathymetry map of North Western Europe showing the North Sea research domain in red. Also, the locations used for the quality control check of H_{m0} are given

2021), increasing the spatial resolution to 0.25 degrees. ERA5 provides windspeed data in Northward (v_{10}) and Eastward components (u_{10}) of the 10 meter elevation windspeed. The windspeed data is converted to a more conventional notation describing the absolute magnitude of the 10 meter elevation windspeed (U_{10}) and the incoming wind direction (ϕ) in degrees with respect to true north (see appendix D).

To assess the applicability of ERA5 Reanalysis in light of this research, time-series for H_{m0} from the ERA5 datasets were compared against in-situ H_{m0} measurements performed at offshore platforms along the Dutch coast (see Figure 4.2). Several goodness-of-fit metrics are employed. The correlation coefficient (R) is employed to measure the linear correlation between ERA5 and the in-situ observations. Bias ($bias$) was calculated to study whether EAR5 Reanalysis tends to under- or overestimate H_{m0} observation. Finally, the Root Mean Square Error ($RMSE$) reflects the deviation of ERA5 Reanalysis compared to the in-situ observations. An elaborate description of the quality control of the ERA5 dataset is given in Appendix D. Regarding the performance of ERA5, the following conclusions can be made:

- The scatterplots (Figures D.1 and D.2) show that ERA5 reanalysis has good agreement with the in-situ observations, with the highest density of the scatter points on or very close to the identity line. For extreme H_{m0} observations, the dispersion is a bit larger, but still within a reasonable limit ($RMSE$ percentage is within 5% for higher observations);
- The bias shows very good agreement with the in-situ observations, with station J6, K13 and Europlatform showing bias values $\leq 5\,cm$. ERA5 time-series near A12 and D15 tend to overestimate H_{m0}, but are still within reasonable values ($\leq 20\,cm$);
- Comparing the original and interpolated H_{m0} datasets, it can be observed that interpolation is beneficial, because R, $bias$ and $RMSE$ for nearly all 5 station show better results. Only $bias$ at station J6 does not give better results compared to the original data. However, with a bias of $-0.037\,m$, the bias is still very much within acceptable limits.

Based on these considerations, it can be concluded that the interpolated ERA5 reanalysis data is suitable for the analysis of the extreme wave climate in the North Sea. Quality control for the wind data was not performed. Rijkswaterstaat (RWS) does not provide historical data for the wind speed at sea and the KNMI only provides land based measurements for wind speed. For land based observations, care should be taken when comparing ERA5 with in-situ observations, because wind observations typically vary on (very) small spatial and temporal scales and are affected by the local terrain (Hersbach et al., 2020). This means the error between ERA5 reanalysis and the land based in-situ observations is potentially quite large. The exact parameters for wind speed and direction are not important in this research. Rather, the wind data must allow for the derivation of long-term temporal patterns in wind climate as a potential cause of changing storm climate in the North Sea. As a result, a comprehensive goodness-of-fit analysis for the wind climate was deemed unnecessary for the aims of this research and was not performed.

4.2. Temporal Trend Analysis Significant Wave Height

This section presents the results of the temporal trend analysis of the significant wave height in the North Sea. Whether a non-stationary approach for extreme value modelling is a useful alternative to the stationary approach, depends on whether a statistically significant trend can be detected in the data. Because of the extent of the North Sea basin, it is expected that temporal trends in the North Sea vary on a spatial scale. Therefore a temporal analysis is executed over the entire North Sea domain. Also, due to the variability of the storm climate over the last 70 years (see Chapter 2.1), it is expected that non-stationarity depends on the considered temporal horizon. To this end, two temporal horizons will be considered in this analysis. The first is the so-called short-term horizon which spans between 1990 and 2020. The second horizon is the long-term horizon and spans between 1950 and 2020.

4.2.1. Mean Significant Wave Height

First, the temporal trends in mean H_{m0} in the North Sea domain are studied. The annual mean value for H_{m0} is selected at each grid point. To detect the presence of significant monotonic trends, the Mann-Kendall (MK) test is employed (see Section 3.2.1) with a significance level of $\alpha = 0.05$. The Theil-Sen (TS) test (See section 3.2.2) is employed to calculate the slope of the detected significant trends. The major assumption made in this research is that potential non-stationarity can be described by a linear trend in time. Figure 4.3 presents the values for the Theil-Sen estimator for mean H_{m0} observations (denoted by $b_{H_{m0},mean}$) in regions where the MK-test detected a significant monotonic trend.

Figure 4.3 shows that in recent history (1990-2020), the Mann-Kendall tests found significant downward trends for mean H_{m0} in large areas of the North Sea, especially near the British Coast. Despite the

Table 4.2: Statistics for the comparison between interpolated H_{m0} observations obtained through ERA5 Reanalysis and in-situ H_{m0} observations in the research domain. Table contains the location of the offshore platforms. The values in brackets give the coordinates of the considered ERA5 grid-point. The number of observations included in the quality control is given by N. Next, the correlation coefficient (R), bias ($bias$) and root mean square error ($RMSE$) are given. For more information, please see Appendix D.

Station	Latitude	Longitude	Duration	N	R	bias (m)	RMSE (m)
A12	55.30 (55.25)	3.80 (3.75)	2009 -2018	84790	0.979	0.102	0.222
D15	54.31 (54.25)	2.93 (3.00)	2009 - 2017	67128	0.967	0.158	0.277
J6	53.82 (53.75)	2.94 (3.00)	2009 - 2018	82136	0.971	-0.037	0.212
K13 Alpha	53.22 (53.25)	3.22 (3.25)	1990 - 2018	249814	0.970	0.023	0.214
Eurplatform	51.98 (52.00)	3.26 (3.25)	1990 - 2018	250688	0.968	-0.012	0.193

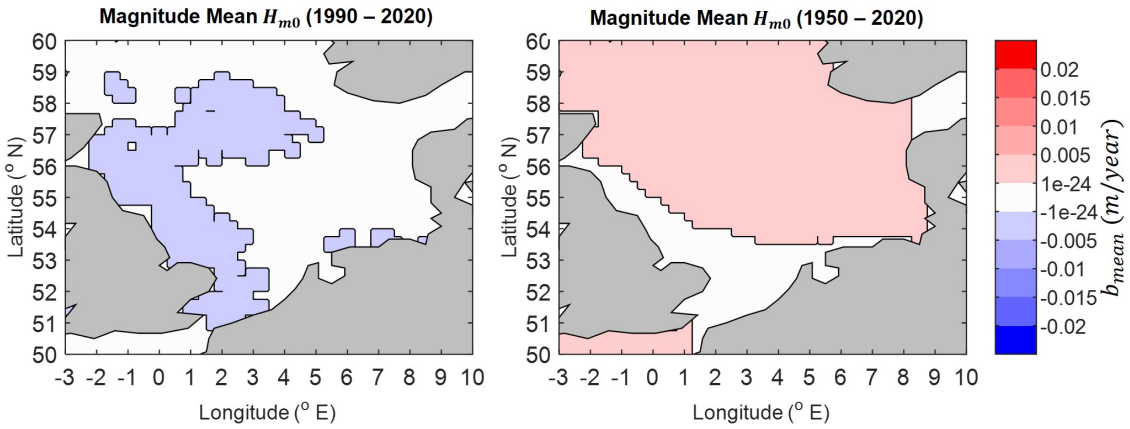


Figure 4.3: Results for the temporal trend analysis for annual mean values of H_{m0}. The figures show the magnitude and direction for a trend ($b_{H_{m0},mean}$) here the MK-test detected the presence of a monotonic trends. Left figures show the results for 1990-2020 and right figures show the period 1950-2020. Blue indicates a downward trend and red indicates an upward trend.

presence of downward monotonic trends, the magnitude of $b_{H_{m0},mean}$ is quite small, with all values for $b_{H_{m0},mean}$ lying below $0.5\ cm/year$. Shifting our attention to the long term temporal horizon (1950-2020), it there is a significant upward trend in the mean H_{m0} throughout the majority of the North Sea basin. Especially for the regions towards the North of the research domain and towards the south of the English Channel. Also for the long term temporal horizon the magnitude of $b_{H_{m0},mean}$ lies below $0.5\ cm/year$.

4.2.2. Extreme Significant Wave Height

The temporal trend analysis for extreme H_{m0} is performed using extreme events identified using annual maxima (AM) and peak over threshold (POT). For the selection of extreme events, the dominant wave directions are not accounted for. At this stage, there is no knowledge about the dominant wave directions at the grid points. For the POT analysis, the values for the threshold (u) and declustering time lag (δ) are carefully selected based on the diagnostic tests for mean residual life (MRL), dispersion index (DI), extremal index (EI) and the associated stability of the GPD parameters. This process is explained in more detail in Appendix E. The appropriate values for u and δ are determined for 7 grid-points throughout the research domain. For these 7 grid points, the 99.5 percentile of H_{m0} and 72 hours for u and δ respectively were determined to be most appropriate. The assumption is that these values for u and δ are appropriate for all grid-points in the research domain. The values for u and δ are subsequently used for the POT selection at all grid points. Apart from analyzing H_{m0} intensity (i.e. magnitude of the peak values), POT analysis is also used to study non-stationarity of the frequency and the duration (T_{dur}) of extreme H_{m0} events. The frequency is defined as the number of annual threshold exceedances and is denoted by N_{exc}. The duration is defined as the time between the start and end of the threshold exceedance, and is denoted by T_{dur} (See figure 3.1). It should be noted that for the temporal trend analysis of N_{exc} and T_{dur}, only the MK-test is considered. The results for the temporal trend tests can be seen in Figure 4.4. The parameters $b_{Hm0,AM}$ and $b_{Hm0,POT}$ are the TS-estimator (i.e. trend slope) of extreme events selected using AM and POT respectively.

The temporal trends for AM H_{m0} (top-left figures in Figure 4.4) show that for the short temporal horizon (1990-2020), downward trends are observed in large areas of the Western and Central North Sea domain. Also, a small region near the German and Danish coast is identified showing a downward trend in AM H_{m0}. The presence of downward trends for this temporal horizon is in accordance with the downward trends identified in mean H_{m0}. In the areas where the MK-test detected the presence of a trend, the magnitude of $b_{Hm0,AM}$ exceeds $3\ cm/year$ and even exceeds $4\ cm/year$ in the central areas of the North Sea, suggesting a large decrease in extreme wave magnitude over the last thirty years.

Shifting our attention to the long temporal horizon (1950-2020), an upward trend for AM H_{m0} is ob-

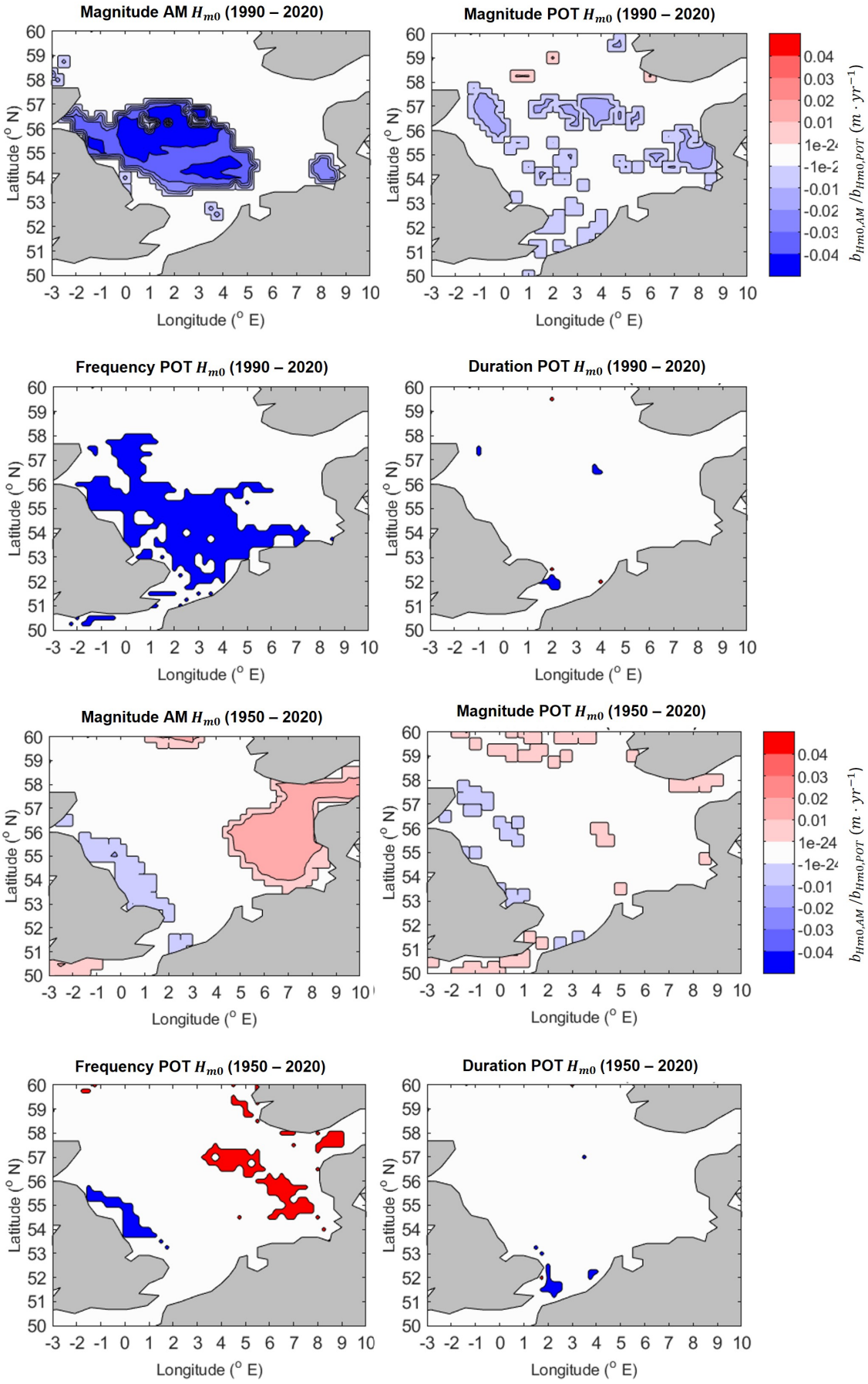


Figure 4.4: Results of trend analysis of extreme H_{m0} identified using AM and POT in the North Sea. The top and bottom sets of figures represent the short and long term analysis respectively. For each set, top figures show magnitude of trends for AM and POT H_{m0} respectively in regions where the MK-test detected a significant trend. Bottom left and right figures show MK-test results for N_{exc} and T_{dur} respectively. Red indicates an upward trend and blue a downward trend.

served in the Eastern areas of The North Sea (near the Danish Coast and The Skagerrak), while a downward trend is observed in the western Areas (British Coast). Furthermore, increasing trends can be seen South of the United Kingdom and along the Northern boundaries of the research domain (West of Norway). The magnitude of the downward trend near the British coast does not exceed $1\,cm/year$, while magnitude of the increasing trend in the Eastern North Sea is higher, showing values between 1 and $2\,cm/year$.

For the temporal trend analysis for extreme H_{m0} observations selected using POT analysis, comparable trends can be observed as for AM H_{m0} events. For the short temporal horizon, strictly downward trends can be observed. However, values for $b_{Hm0,POT}$ are lower than values for $b_{Hm0,AM}$. The values for $b_{Hm0,POT}$ do not exceed $2\,cm/year$. Also, the regions where the significant trends for POT H_{m0} are detected are smaller than for AM H_{m0} and show more spatial spreading. For the long temporal horizon, downward trends are observed in the Western North Sea, while upward trends are observed in several other regions. This is consistent with the detected trends in AM H_{m0}. The only exception being near the Danish coast, where almost no statistically significant trends were detected. Similar to the findings for the short term analysis, the extent of the areas where the MK-test detected a trend in POT selected H_{m0} extremes is small compared to AM H_{m0} observations. Furthermore, values for b_{POT} are smaller than for b_{AM} with the magnitude of the detected trends not exceeding $1\,cm/year$.

The magnitude of the temporal trends for extreme H_{m0} observations selected using POT is smaller than for AM. This phenomenon can be explained by the nature of the POT analysis and the fixed threshold applied in this research. Whereas AM selects the most extreme H_{m0} observation for every year, the POT analysis only selects extreme H_{m0} observations exceeding the threshold (u). Subsequently, the MK and TS tests only consider the magnitude of the H_{m0} observations exceeding u. This means that all values included in the temporal trend analysis have a minimum value for H_{m0} equal to u, making it more difficult to detect and quantify a statistically significant trend, especially for longer time scales. If H_{m0} occurrences no longer exceed u for a given period within the investigated temporal horizon, the presence of long-term trends may be obscured. The implications of using a fixed threshold on the temporal trend analysis, and suggestions to overcome this problem, are further elaborated upon in the discussion.

Apart from the peak magnitudes of H_{m0}, the POT analysis was also applied to detect the presence of temporal trends in the frequency (N_{exc}) of extreme events and the average duration (T_{dur}) of extreme H_{m0} events (see Figure 4.4). The temporal trends in $Nexc$ are consistent with the trends in extreme $Hm0$ events selected using AM and POT. For the short temporal horizon, a strictly downward trend in N_{exc} can be observed in large regions of the North Sea domain. For the long temporal horizon, an upward trend in the Eastern areas of the North Sea, and a downward trend in the Western regions, can be observed. Regarding T_{dur}, almost no trends are observed for both the the short and long temporal horizons. Notable is that for both temporal horizons, the few observed trends are strictly downward. However, no explanation for this phenomenon has been found.

In conclusion then, extreme H_{m0} shows different temporal trends depending on the location and considered temporal horizon. For the short temporal horizon, strictly downward trends for both the magnitude of H_{m0} and the frequency of extreme events (N_{exc}) are observed. For the long temporal horizon, various regions throughout the research domain show an upward trend for the magnitude of H_{m0} and N_{exc}, especially near the Danish coast. Downward trends for H_{m0} and N_{exc} are observed near the British coast. Because the direction and magnitude of temporal trends in AM $Hm0$ depend on the considered location and temporal horizon, the extreme wave climate in the North Sea does not exhibit monotonic non-stationarity. The magnitude and direction of the temporal trends are important factors to consider in the (non)-stationary extreme value analysis of H_{m0} as will be further explained in Chapter 5.2.

4.3. Temporal Trend Analysis Extreme Wind Speeds

Wind is the physical driver of wave growth (see Appendix A). A temporal trend analysis of extreme wind speed (U_{10}) is performed to study if non-stationarity in extreme U_{10} can (partially) explain the detected non-stationarity of H_{m0}. This potentially provides a better understanding of the non-stationarity of ex-

treme H_{m0} in the North Sea.. This section studies the trends for extreme U_{10} selected using Annual Maxima (AM) and Peak over Threshold (POT). A similar selection of the threshold (u) and declustering time lag (δ) as performed in Section 4.2.2 showed that the appropriate values for u and δ are the 99.5 percentile of u_{10} and 96 hours respectively. The results for the temporal trend analysis of extreme U_{10} can be seen in Figure 4.5. In Figure 4.5, $b_{u10,AM}$ and $b_{u10,POT}$ represent the TS estimators (i.e. slope of the temporal trend) for extreme U_{10} events selected using AM and POT respectively.

Comparing the results of the temporal trends for AM H_{m0} and U_{10} for the short temporal horizon shows that a large region in the Central North Sea where a downward trend for AM H_{m0} was observed, is also characterized by a decrease in AM u_{10}. The magnitude of this trend varies between 0 and $0.15m/s/year$. Combined with the knowledge that extreme waves in the North Sea are mostly wind waves (See Appendix D) gives reason to believe that the downward trend in H_{m0} between 1990 and 2020 can partially be explained by decreasing U_{10}. However, there are also areas in the North Sea that show a decrease in wind speed but no decrease in the resulting extreme H_{m0}, for example in The Channel (See figures 4.4 and 4.5).

Shifting our attention to the long-term temporal horizon (1950-2020) for AM U_{10}, we can observe that the large region with an upward trend for AM H_{m0} in the Eastern region of the North Sea is accompanied by a similar region with an upward trend for AM U_{10}. This can also be observed in the Northern parts of the research domain (West of Norway) and the South-West of the research domain (South of England and the English Channel). For these regions, it can be said that the increase in H_{m0} can at least partially be attributed to an increase in extreme U_{10}. However, for AM H_{m0}, a decreasing trend was observed East of the British coasts. This downward trend is not observed for AM U_{10} observations.

The plots for the POT analysis of U_{10} show similar results compared to H_{m0}. For the short term horizon, a large region with a downward trend for U_{10} can be observed in the central North Sea regions. For the long temporal horizon, a downward trend can be observed for the British coast, that was not detected for the AM U_{10}. Despite the trend not being detected for AM u_{10}, the resulting trends for the POT u_{10} are enough reason to suggest that the downward trend for the extreme wave climate in Western parts of the North Sea are caused by a downward trend in wind speed.

Finally looking at the trends for the frequency (N_{exc}) and duration (T_{dur}) of extreme U_{10}, it can be seen that for the short temporal horizon, the regions with temporal trends in N_{exc} and T_{dur} are relatively small and spread out over the research domain. For the long term horizon, a large region with increasing N_{exc} can be observed for the regions West of Denmark and the Skagerrak. Regarding T_{dur}, a downward trend can be observed for the British coast. This suggests that the decrease in extreme H_{m0} near the British coast between 1950 and 2020 is not only caused by a decrease in the magnitude of U_{10} but also by a decrease in storm duration.

In conclusion, it can be said that the observed temporal trends in extreme H_{m0} can at least partially be explained by trends in the extreme wind speeds. The observed trends for extreme H_{m0} are accompanied by similar trends for U_{10}. An attribution study to better understand the underlying cause of the non-stationarity of H_{m0} and U_{10} is not performed in this research. It is assumed that the detected temporal non-stationarity is a result of climate variability.

4.4. Wave Parameters for Hierarchical Clustering Analysis

This research assesses the practicability of Hierarchical Agglomerative Clustering (HAC) analysis to cluster grid-points in the North Sea based on similar extreme wave characteristics. This requires the careful selection of wave parameters and representative input values for the HAC analysis. The motivation behind the selected wave parameters is explained in Section 4.4.1. Next, the derivation of the input values representing the wave parameters in the HAC analysis is explained in Section 4.4.2.

4.4.1. Motivation for Wave Parameters

The primary goal of this research is to study the methodology how extreme wave loading conditions are derived for design and risk assessment of hydraulic infrastructure in light of climate change. design

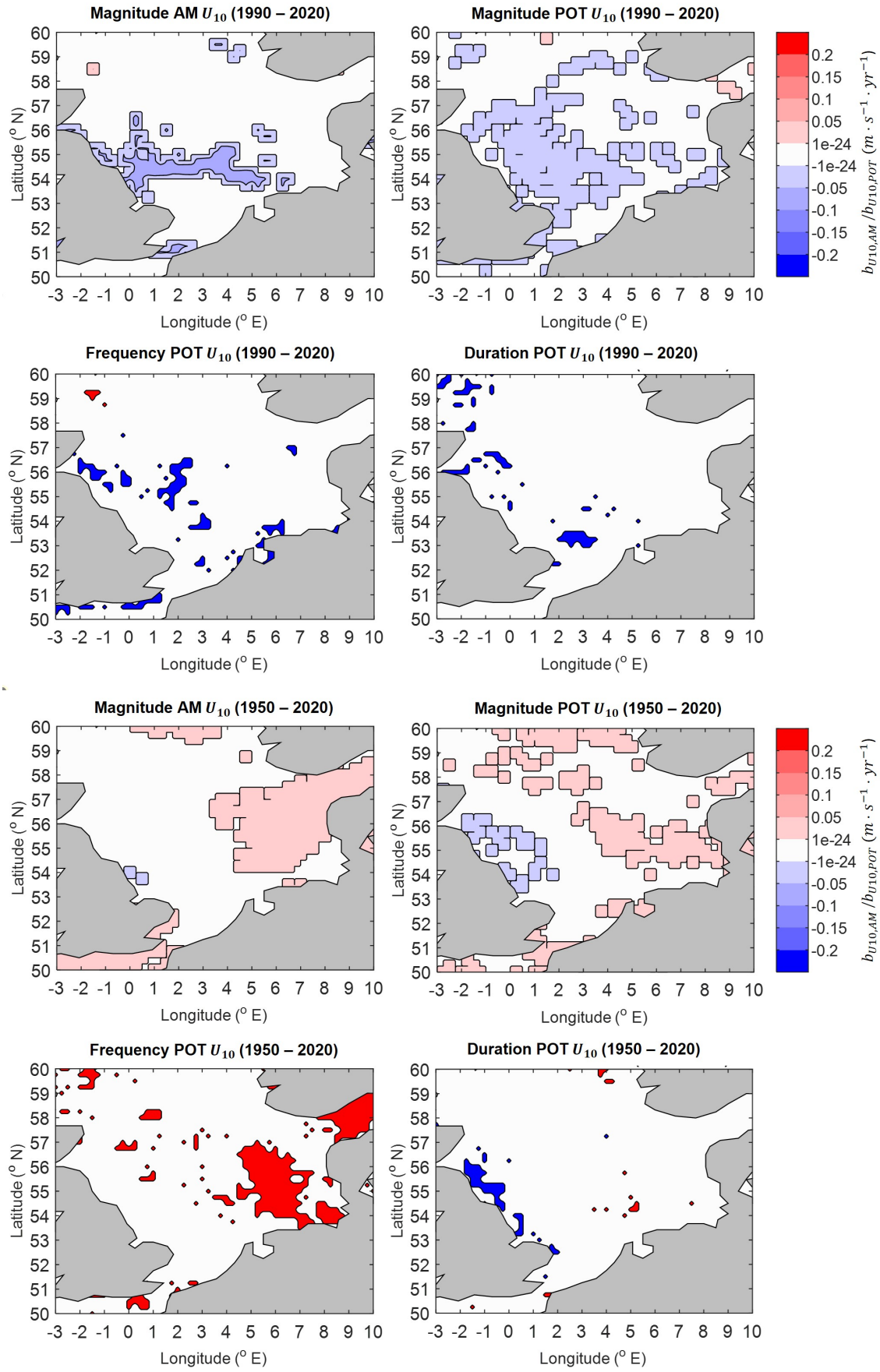


Figure 4.5: Results of trend analysis of extreme U_{10} identified using AM and POT in the North Sea. The top and bottom sets of figures represent the short and long term analysis respectively. For each set, top figures show magnitude of trends for AM and POT U_{10} respectively in regions where the MK-test detected a significant trend. Bottom left and right figures show MK-test results for N_{exc} and T_{dur} for extreme U_{10} events respectively. Red indicates an upward trend and blue a downward trend.

and risk assessment of hydraulic infrastructure concern the most extreme wave conditions a hydraulic structure is subjected to during its lifetime. The clustering output is based on grid points exhibiting similar extreme wave characteristics, thereby helping in the derivation of extreme loading conditions. To obtain useful clustering output requires the careful selection of wave parameters to be used to generate the input for the HAC analysis. This research is aimed at practitioners working on design and risk assessments of hydraulic infrastructure. As a result, wave parameters for the HAC analysis are selected based on the parameters often used in design and risk assessment of hydraulic infrastructure. Engineers and design practitioners use parameters that describe the extreme wave climate directly, such as the significant wave height and wave period. Using clear parameters enables the design practitioner to better interpret the cluster boundaries.

The first wave parameter selected for HAC-analysis is the dominant wave direction (θ_1 and θ_2). The dominant wave direction is especially important for the design for coastal infrastructure because it provides vital information about the incoming wave angle. More importantly, information about the dominant wave directions helps to filter time-series at the grid points, thereby creating subsets of homogeneous and independent extreme observations required for the extreme value analysis (Van Den Bos & Verhagen, 2018). Often, multiple dominant wave directions are relevant for coastal structures. Therefore, the HAC analysis also includes a secondary dominant wave direction (θ_2) to identify clusters for which multiple dominant wave directions must be considered.

The second wave parameter is the significant wave height (H_{m0}). The H_{m0} is one of the most important parameters when performing risk assessments on hydraulic infrastructure. It is used in most limit state functions describing the failure modes associated with hydraulic infrastructure (Jonkman et al., 2018; Voorendt & Molenaar, 2019). Also, this research aims to derive extreme value models that describe H_{m0} on the cluster level. Therefore, H_{m0} is a crucial parameter to include in the HAC analysis. Finally, the peak period (T_p) is selected. Apart from H_{m0}, the T_p is often used in the design and risk assessment of hydraulic infrastructure because provides information about the maximum wave energy in the extreme wave field (Van Den Bos & Verhagen, 2018).

The number of input parameters in this analysis is limited to 4 parameters. In hierarchical clustering, there is a trade-off between the number of input parameters and the number of clusters required to give a good representation of the input. One of the primary benefits of clustering analysis is the data volume reduction to study the extreme wave climate throughout the North Sea. More parameters require a higher number of clusters, meaning that potentially the goal of data volume reduction is not met. Therefore, it is decided to keep the number of input parameters limited to 4.

4.4.2. Derivation of HAC Input Values

Now that the wave parameters for the Hierarchical Clustering (HAC) analysis are selected, the input values must be derived. Due to the nature of HAC Analysis, each wave parameter must be characterized by a single representative value at the grid point. This means that 4 input values representing the 4 wave parameters of interest must be derived for each grid point. The main focus is to derive wave characteristics associated with extreme wave events. This requires the careful selection of an extreme threshold based on the significant wave height. All wave events with H_{m0} exceeding the threshold are selected with corresponding values for T_P and θ. From this subset, the HAC input values for the dominant wave direction (θ_1 and θ_2), the significant wave height ($H_{m0,P99.5}$) and the peak period (T_p) are selected.

4.4.2.1. Definition of Extreme Wave Climate

For design and risk assessment of hydraulic infrastructure, the most extreme wave loading conditions that the structure will be subjected to during its lifetime are of interest. These extreme loading conditions will occur during storm conditions. This requires a careful definition of an extreme wave climate. This research defines the extreme wave climate based on the appropriate threshold (u) for significant wave height (H_{m0}) in a Peak over Threshold (POT) analysis. The value of u in a POT analysis can be decided based on the diagnostic tests for mean residual life (MRL), dispersion index (DI), extremal index (EI) and the stability of the GPD parameters. The selection of the appropriate storm threshold is described in detail in Appendix E. Here a short summary is presented about the most important as-

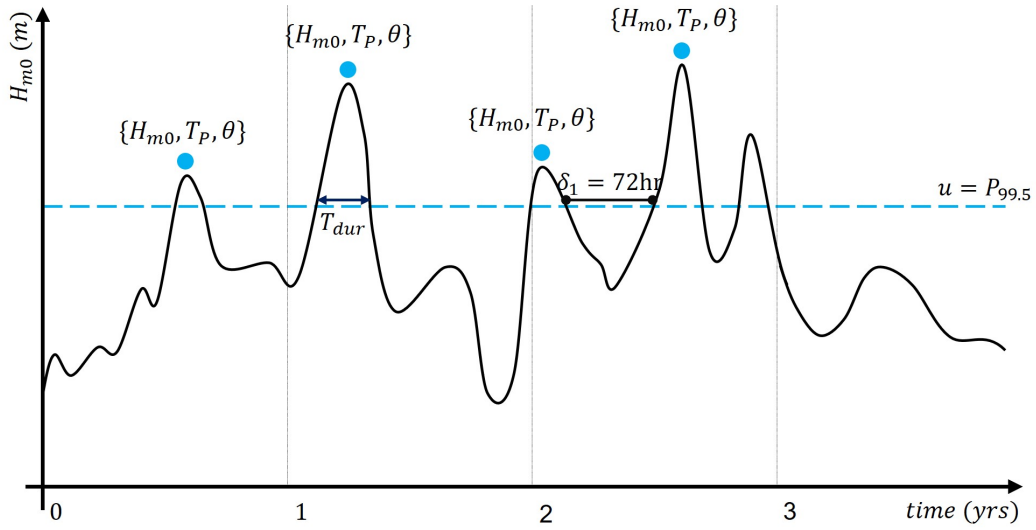


Figure 4.6: Peak over threshold analysis to determine an extreme wave threshold for the derivation of input values representing the parameters in the HAC analysis.

sumptions and simplifications made for the definition of the extreme wave climate.

In a POT analysis, extreme observations should be both homogeneous and independent (see Section 3.1.2). Because we have little prior knowledge about the dominant wave directions at each grid point, the complete directional wave spectrum was used to determine the extreme wave threshold. It should be noted that this violates the requirement of homogeneity, for which the selected extreme wave events should come from similar directions. However, given the limited understanding about the dominant wave direction(s) prior to this investigation, it was determined that it was a necessary compromise to make at this time. An appropriate declustering time lag was selected using MRL, DI, and EI, meaning that the requirement of independent observations in the POT analysis is still satisfied.

Because performing a full POT analysis would require too much processing, the diagnostic tests for MRL, EI, DI and the stability of the GPD parameters are performed at 7 grid locations throughout the research domain. These 7 grid points are selected to give the best representation of potentially different extreme wave climates throughout the research domain. The aim is to find one value for u and one value for δ that are appropriate at all 7 grid points. The assumption is that the determined values for u and δ at the 7 grid points are appropriate at all grid points in the research domain. In the end, it was found that the 99.5 percentile for (H_{m0}) is an appropriate threshold (u) at all 7 locations and the declustering time lag (δ) should be equal to 72 hours at all 7 locations.

Using the established values for u and δ for the significant wave height, a POT analysis is performed giving information about the wave parameters (H_{m0}, T_P and δ) at all grid points in the North Sea domain. The obtained observations for H_{m0}, T_P and δ are subsequently used to determine the input values for the HAC analysis.

4.4.2.2. Dominant Wave Directions

The first wave parameter included in the clustering analysis is the dominant wave direction. Knowledge about the dominant wave direction in each location helps to create homogeneous and independent subsets of extreme H_{m0} observations that can be used in the temporal trend analysis and the intra-cluster extreme value analysis (See Section 3.1). The input value representing the dominant wave direction(s) in the HAC analysis is determined as follows. For every grid point in the research domain, the values for H_{m0} exceeding the storm threshold and the corresponding value for θ are filtered (see Figure 4.7). Next, the marginal probability density functions (PDF) are fitted over the observations of θ. It should be noted that the parameters for mean wave directions are circular on the domain [0,360]. This means

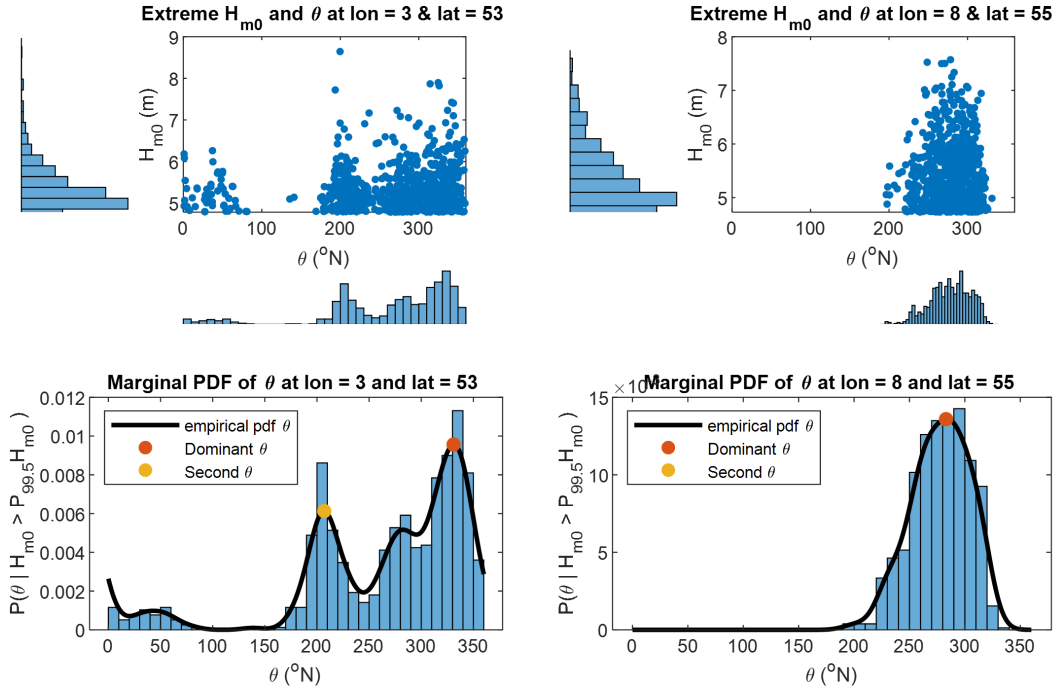


Figure 4.7: Top row: Extreme observations for significant wave height and corresponding mean wave directions for the Dutch coast (left) and Danish coast (right). Bottom row, marginal distributions of the mean wave direction and selected dominant wave directions based on the highest probability density.

that the PDF must be continuous for the wave directions from the north.

Finally, the direction with the highest probability density is selected as the primary dominant wave direction (θ_1). This means that θ_1 is defined as the extreme wave direction with the highest probability of occurrence. In case there are multiple directions that are sufficiently far apart to be considered separate wave directions (difference between peaks of the PDF for θ should be $\geq 45^o$), both dominant wave directions are included in the analysis. The secondary dominant wave direction is denoted by θ_2. This process is repeated for every grid point in the research domain. Figure 4.8 shows the resulting map of dominant wave directions in the North Sea domain.

4.4.2.3. Significant Wave Height

The input values for the significant wave height (H_{m0}) for the HAC analysis are based on the established extreme wave threshold. At each grid point, the 99.5 percentile values for H_{m0} are determined based on the entire wave time-series. This value is denoted by $H_{m0,P99.5}$ and serves as the input value for the HAC analysis. It should be noted that non-stationarity in H_{m0} is not accounted for whilst deriving the HAC input values. Therefore, the value of $H_{m0,P99.5}$ is potentially affected by the non-stationarity of H_{m0} at the grid points. The computed values of $H_{m0,P99.5}$ are a representation of the examined temporal horizon and are thus (partially) dependent on the non-stationarity in H_{m0}. The HAC algorithm potentially links grid points together with similar $H_{m0,P99.5}$ values whilst ignoring non-stationarity. The implications of this on the research are further elaborated upon in the Discussion.

$H_{m0,P99.5}$ varies smoothly across the North Sea domain, meaning that it is a continuous parameter. Clustering spatially continuous data is potentially problematic because of the small dissimilarities between spatially neighboring grid points that may result in the Chaining Problem (see Section 3.3.1.2). To overcome the chaining problem, the values for $H_{m0,P99.5}$ are rounded down to the nearest integer. This way, the dissimilarity between different locations is increased allowing for a better HAC analysis. Also, regions can be identified based on a minimum value for $H_{m0,P99.5}$ to account for during extreme wave conditions. The resulting map for $H_{m0,P99.5}$ in the North Sea domain can be seen in Figure 4.9.

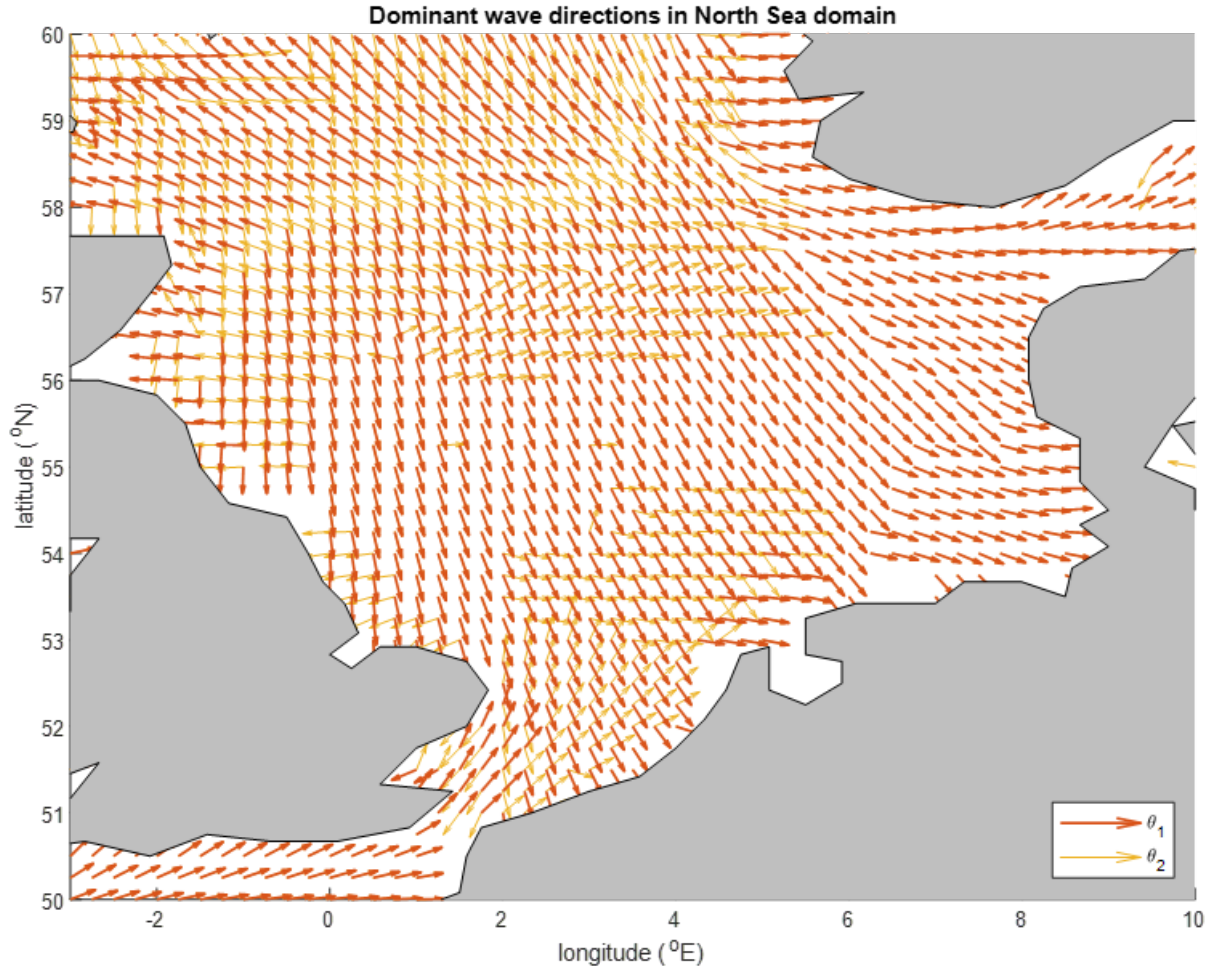


Figure 4.8: Dominant wave directions in the North Sea domain as calculated in Section 4.4.2.2. Orange quivers correspond to θ_1, yellow quivers correspond to θ_2.

4.4.2.4. Peak Wave Periods

In this research, clustering of extreme wave characteristics is also based on the peak periods (T_p) related to extreme wave conditions. Only including H_{m0} is not sufficient to have a complete understanding of the extreme wave climate, and information about T_P is required for the design and risk assessment of hydraulic infrastructure (De Leo et al., 2021). Furthermore, T_p corresponds to the waves where the maximum amount of energy is stored (see Appendix A). To account for potential fetch and wave dispersion effects related to waves from different (dominant) directions, the characteristic value for T_P in each grid point is determined based on the primary dominant wave direction (θ_1). First, the time-series for the extreme wave observations have been filtered so that all extreme wave events with a mean wave direction $\theta_1 \pm 45^o$ remain. The minimum value of T_P from the filtered time-series of extreme waves has been selected to serve as HAC input value. Because H_{m0} and T_P are positively correlated, it is assumed that the minimum value for T_p is closest related to the value for $H_{m0,P99.5}$. Similar to $H_{m0,P99.5}$, T_p is a continuous parameter with small dissimilarities between spatially neighboring grid points, making clustering challenging because of the Chaining Problem. To overcome the Chaining Problem, it decided to round the obtained values down to the nearest integer. This gives information on a minimum value for T_P to account for during storm conditions in different regions of the North Sea. The characteristic values for T_P can be seen in Figure 4.9.

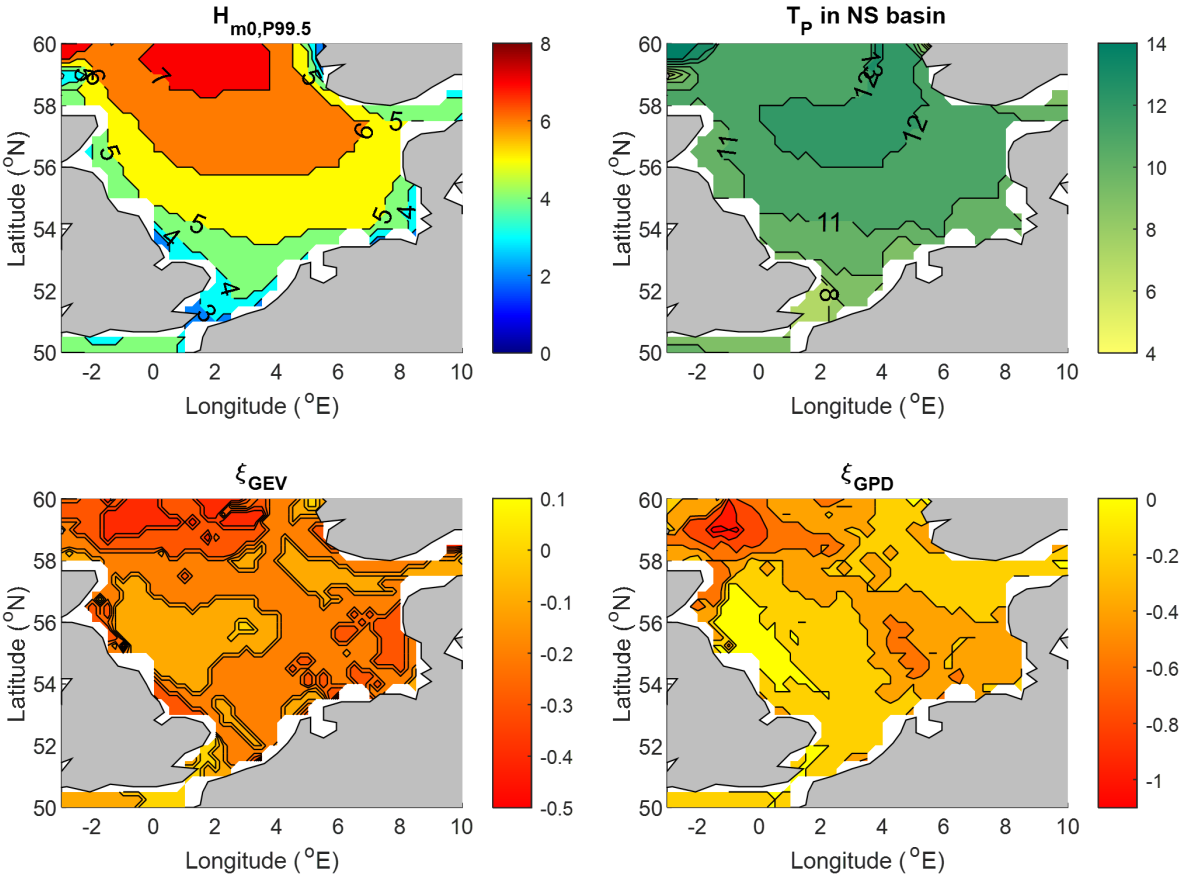


Figure 4.9: Spatial distributions of the characteristic values for $H_{m0,P99.5}$ (top left), T_P (top right), ξ_{GEV} (bottom left) and ξ_{GEV} (bottom right) in the North Sea basin.

4.5. Summary

This chapter presents an exploratory analysis of the North Sea and the research domain. Section 4.1 presents an overview of the ERA5 Reanalysis that provides the time-series for wave and wind parameters used in this research. The ocean wave data was interpolated to increase the spatial resolution and the wind parameters are converted to a more conventional notation including the absolute windspeed and (U_{10}) and wind-direction (θ) with respect to true North. The quality control check of the significant wave height (H_{m0}) against in-situ measurements showed that the interpolated ERA5 data is suitable for this research because it gives a sufficiently accurate representation of the wave climate.

Section 4.2.2 provides a temporal trend analysis of H_{m0} to get a better understanding of the non-stationarity in the North Sea domain. Temporal trend analysis was carried out by means of selecting extreme H_{m0} observations using Annual Maxima (AM) and Peak over Threshold (POT). The presence of statistically significant trends was tested using the Mann-Kendall (MK) test with a significance level of $\alpha = 0.05$. The slope of statistically significant trends was calculated using the Theil-Sen (TS) test. Furthermore, two different temporal horizons are studied. The short term horizon primarily showed downward trends for H_{m0} whilst the long term horizon showed a combination of upward trends in the East and downward trends in the Western parts of the research domain. From the opposite trends detected for the two temporal horizons, it is concluded that the non-stationarity of H_{m0} is not monotonic but more likely shows long term oscillatory cycles.

Temporal trends for the extreme wind speeds (U_{10}) showed similar results, i.e. strictly downward trends for the short temporal horizon and a combination of upward trends (East) and downward trends (West) for the long temporal horizons. Albeit not exact matching, the spatial distribution of temporal trends in

U_{10} is very similar to the distribution of temporal trends in extreme H_{m0} observations, for both temporal horizons. With the North Sea extreme wave climate being characterized by locally generated wind waves, there is enough reason to conclude that the temporal evolution of extreme H_{m0} can at least partially be explained by similar temporal evolution in extreme U_{10}.

Finally, in Section 4.4.2, the input values of the wave parameters for the Hierarchical Agglomerative Clustering (HAC) analysis are computed. The wave parameters are used for hierarchical clustering are selected based on parameters often used in the design and risk assessment of hydraulic infrastructure. These include the dominant wave directions (θ_1 and θ_2), the significant wave height (H_{m0}) and the Peak Period (T_P). The derivation of the HAC input values includes the definition of an extreme wave threshold and subsequent POT analysis used to select the extreme wave observations. Fromt the extreme observations, the HAC input values for H_{m0}, T_P, θ_1 and θ_2 are derived. Values for θ_1 and θ_2 are derived based on the frequency of occurrence of θ above the extreme wave threshold. The HAC input value for H_{m0} is the extreme wave threshold $H_{m0,P99.5}$. Finally, the HAC input values for T_P are the lowest observed value for T_P corresponding the primary dominant wave directions. Further processing of $H_{m0,P99.5}$ and T_P is required because of the small spatial dissimilarities between spatially neighboring grid points. Otherwise, the chaining problem might result in the continuous linkage of these grid points because the HAC algorithm is unable to detect a clear cut-off point.

5

Hierarchical Clustering Analysis and Extreme Value Analysis at Cluster Level

This chapter presents the results of the clustering analysis and extreme value analysis at cluster level. Section 5.1 presents Hierarchical Agglomerative Clustering (HAC) analysis in which 63 clusters are derived based on extreme wave characteristics for the significant wave height ($H_{m0,P99.5}$), Peak Wave Period (T_P) and the dominant wave directions (θ_1 and θ_2). A detailed analysis of the derivations of the input parameters is presented in Section 4.4.2. Furthermore, this chapter presents the intra-cluster extreme value analysis. This consists of a temporal trend analysis of extreme significant wave height (H_{m0}) and extreme windspeeds (U_{10}) on a cluster level and the derivation of representative extreme value models describing the significant wave height (H_{m0}) on a cluster level.

5.1. Hierarchical Cluster Map of the North Sea

The first step in this research is to apply Hierarchical Agglomerative Clustering (HAC) to partition the grid-points of the research domain into clusters based on similar extreme wave characteristics. The primary goal is to identify clusters that exhibit similar extreme wave characteristics, which is important knowledge for the design and risk assessment of hydraulic infrastructure. The information from the intra-cluster grid-points can be pooled to improve extreme value models at cluster level. Moreover, clustering analysis potentially results in a reduction of the data volume that must be studied to obtain an understanding of the wave climate in the North Sea. This research performs HAC analysis based on input values of the significant wave height ($H_{m0,P99.5}$), the peak wave period (T_P) and the dominant wave directions (θ_1 and θ_2) corresponding to extreme wave events at the grid-points. More information about the HAC input values is presented in Section 4.4.2.

Before taking a closer look at the clustering results, a quick summary of the steps involved in assessing the clustering output is presented. A comprehensive description is presented in Section 3.3.6. Assessment of the clustering output for each configuration can be divided into the following steps:

1. Compare the performance of the different linkage mechanisms using the cophenetic correlation C and select the 2 linkage mechanisms that yield the highest values for C.
2. Determine the number of clusters ($\mathcal{K}$) that yields the best clustering output for the two linkage mechanisms based. Evaluation of $\mathcal{K}$ is done by assessing the following internal evaluation metrics: Silhouette Coefficient (SC), Calinski Harabasz criterion (CH) and the Davies-Bouldin value (DB). If values for $\mathcal{K}$ are similar between metrics, a choice for $\mathcal{K}$ is straightforward. If values for $\mathcal{K}$ are far apart, a critical analysis of the clustering output for different values of $\mathcal{K}$ is required to select the optimal value for $\mathcal{K}$.
3. For each weight configuration, select the linkage mechanism and value for $\mathcal{K}$ that yields the best clustering solution based on the relative scores for SC, CH and DB and how well the cluster boundaries represent the HAC input values.

4. For each parameter combination, select the weigh configuration, corresponding linkage mechanism and value for $\mathcal{K}$ that yields the best clustering solution. This is based on a critical assessment of the clustering output against the HAC input values for the five weight configurations.

The final clustering output is then subject to further development to improve the cluster quality. This involves the splitting of clusters with grid-points in different geographical locations and the merging of neighboring clusters when the wave parameters provide sufficient statistical evidence between to do so.

One of the goals of clustering is potential data volume reduction. This means that the value for $\mathcal{K}$ must result in a sufficient data volume reduction. In Hierarchical clustering, there is always a trade-off between $\mathcal{K}$ and the quality of the resulting clusters. Too few clusters results in grid-points with different input values to be linked together. Too many clusters will lead to well-defined clusters based on the extreme wave characteristics, but the data volume reduction will not be as efficient. To limit the number of clusters identified by the HAC-algorithm, a maximum possible number of clusters is imposed ($\mathcal{K}_{max} = 50$). A comprehensive motivation for $\mathcal{K}_{max}$ is presented in Section 3.3.5.

5.1.1. Initial Clustering Results

This research presents HAC analysis for two wave parameter combinations to see which combination yields the best clustering results. For each wave parameter combination, the best clustering solution must be identified. A comprehensive description of the complete HAC analysis for each configuration (different parameter combination and assigned weights) is presented in Appendix F. The following sections provide a summary describing how the best performing clustering output for each wave parameter combination is presented. In section 5.1.1.4, the overall best performing cluster analysis is selected that is used for subsequent further development in Section 5.1.2.

5.1.1.1. Wave Parameter Combinations

The specified wave parameters are used in two combinations in the HAC analysis performed in this study. Both combinations use information about the dominant wave directions (θ_1 and θ_2) and the number of dominant wave directions (N_θ). Extreme wave fields are not always uni-directional and may show different dominant wave directions. Knowledge about the dominant wave direction(s) helps

Table 5.1: Weight configurations for the different wave parameter combinations used in HAC analysis. The top table gives the configurations for the first wave parameter combinations. The bottom table gives the weights for wave parameter combination 2. The first weight configurations at more weight to $H_{m0,P99.5}$ (and T_P if applicable), and the last configurations apply more weight to θ_1 and θ_2.

Configuration	Assigned Weights (%)			
	θ_1	θ_1	N_θ	$H_{m0,P99.5}$
1.1	15	15	15	55
1.2	20	20	20	40
1.3	25	25	25	25
1.4	27.5	27.5	25	20
1.5	30	30	25	15

Configuration	Assigned Weights (%)				
	θ_1	θ_1	N_θ	H_{m0}	T_P
2.1	10	10	10	35	35
2.2	15	15	15	27.5	27.5
2.3	20	20	20	20	20
2.4	25	25	25	12.5	12.5
2.5	30	30	30	5	5

to create time-series that contain homogeneous and independent extreme H_{m0} observations, which is critical for the extreme value analysis at the cluster level in Section 5.2. N_θ is added to provide additional steering to the clustering analysis.

Combination 1: Significant Wave Height
The first combination of parameters is a combination of the dominant wave directions (θ_1, θ_2 and N_θ) and the significant wave height (H_{m0}) at each grid-point. The peak period (T_p) is not included in this analysis for the benefit of the clustering output. Including more parameters requires a higher value for $\mathcal{K}$, for the clusters to properly respect the dissimilarities between the input values. Moreover, in design practices, H_{m0} is a more important parameter than T_p (Jonkman et al., 2018). A separate analysis with only T_p is not performed because information about the peak period requires information about the extreme significant wave height, and not vice versa.

Combination 2: Significant Wave Height and Peak Period
In the second parameter combination, both $H_{m0,P99.5}$ and T_P are included next to θ_1, θ_2 and N_θ. Information about H_{m0} alone is typically not sufficient, and information about T_P is required to have a complete understanding of the extreme wave climate (De Leo et al., 2021). Furthermore, T_P is an important parameter in the design and risk assessment of hydraulic infrastructure because it provides information about the waves that contain the most energy. Therefore, it makes sense to divide the North Sea into areas that share similar characteristics for both H_{m0} and T_p. Because the number of parameters is increased, it is expected that the overall quality of the HAC analysis will decrease if the total number of clusters is not increased. The clustering output for combination 2 should therefore be critically assessed to see whether adequate cluster boundaries are still obtained for this wave parameter combinations.

Weight Configurations
The input values are normalized by means of z-score standardization prior to HAC analysis. Therefore, it is possible to assign weights to the different parameters (See Section 3.3.1.3). Weight may be assigned to emphasize particular parameters in the clustering that are thought more significant or to add some additional steering to obtain clustering output that best serves the engineering goals. In total, five different weight configurations were assigned for both parameter combinations to the wave parameters. For the first runs of each parameter combination, more weight is assigned to H_{m0} and/or T_P compared to the directional parameters θ_1, θ_2 and N_θ respectively. Then for each subsequent configuration, more weight is removed from H_{m0} and T_P and is added to θ_1, θ_2 and N_θ respectively. The applied weight configurations are given in Table 5.1.

5.1.1.2. Combination 1: Significant Wave Height

The first combination of parameters included θ_1, θ_2, N_θ and $H_{m0,P99.5}$. For a complete overview of the values of the diagnostic parameters, see Table 5.2. For configuration 1.1, it was decided that the best clustering results were obtained using Ward linkage with $\mathcal{K} = 50$. Ward linkage yields superior values for CH compared to Average linkage. Moreover, Ward linkage provides better cluster boundaries based on θ_1 and θ_2. Similar results where found for configuration 1.2. Average linkage scored much better compared to Ward linkage when comparing the values for C. Also, the range of $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ is much smaller for Average linkage than it is for Ward linkage (see Table 5.2). Nevertheless, Ward linkage still produced the better cluster divisions that respected the boundaries between θ_1 and θ_2. In both configurations 1.1 and 1.2, average linkage placed to much emphasis on $H_{m0,P99.5}$.

When studying configuration 1.3, it was quickly found that due to to the superior scores for SC, CH and DB, the clear agreement between $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ and the well-defined cluster divisions based on θ_1, θ_2 and $H_{m0,P99.5}$, the ideal clustering solution for this configuration is Ward linkage clustering with a total of 50 clusters ($\mathcal{K}$).

For configuration 1.4, Ward linkage provides the highest values for SC, CH and DB. Despite this, Average linkage was found to give better clustering results for configuration 1.4. Average linkage yields better cluster boundaries, higher values for C and $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ for Average Linkage give a clear agreement on the ideal value for $\mathcal{K}$ ($\mathcal{K} = 50$). For configuration 1.5, Ward linkage yields better

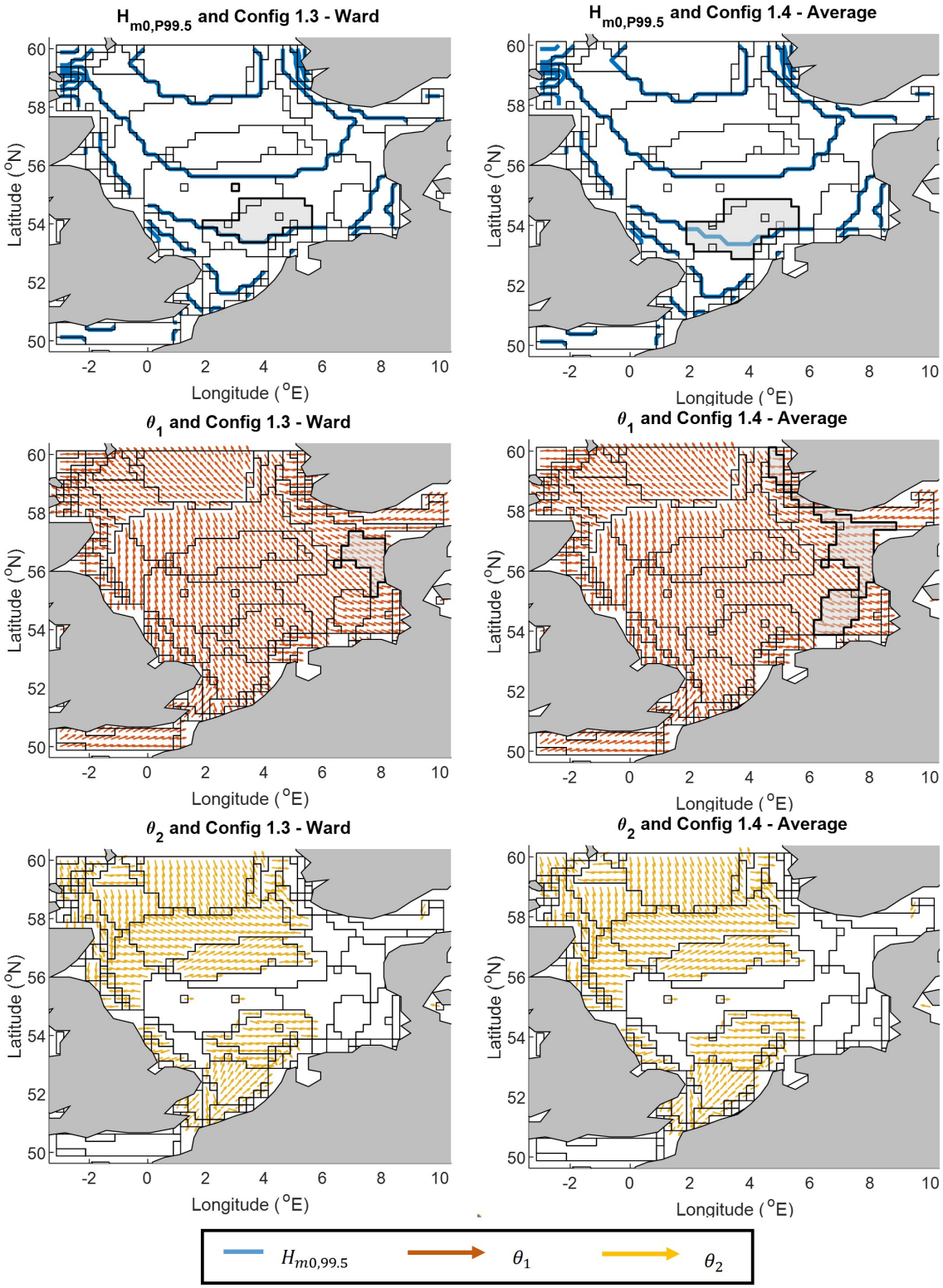


Figure 5.1: Comparison between the results for configuration 1.3 clustered with Ward linkage (left), and configuration 1.4 with average linkage (right). Top figures compare the clustering output with the contours of $H_{m0,P99.5}$ and the middle and bottom compare the clustering output with θ_1 and θ_2 respectively.

results than Average linkage. However, neither Ward and Average linkage provides good clustering solution. Both Linkage mechanisms place high emphasis on θ_1 and θ_2. This can be explained by the weight configuration. Moreover, no clear clustering solution can be identified based on the values for

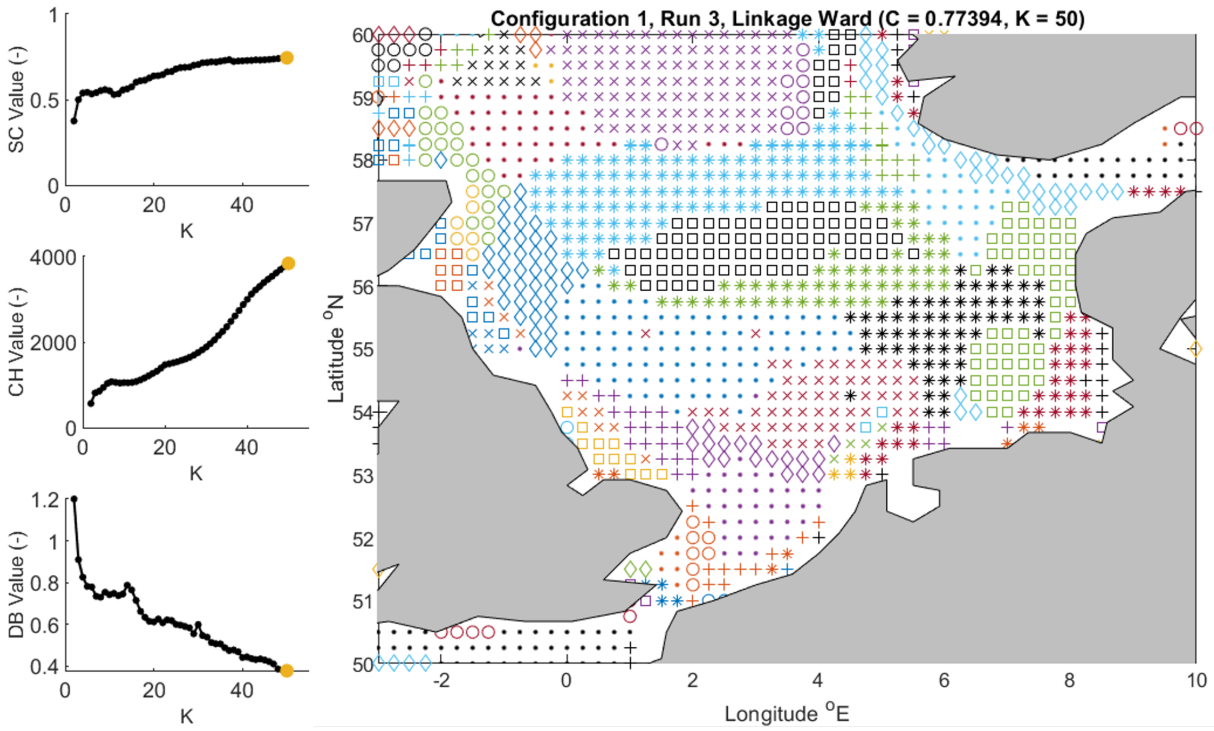


Figure 5.2: Optimal cluster solution for wave parameter combination 1. Figures on the left shows the plots for SC, CH and DB for different values of $\mathcal{K}$. The final selected value for $\mathcal{K}$ is givnen in yellow. The right figure shows the resulting clustering output for wave parameter combination 1.3 using Ward linkage and $\mathcal{K} = 50$.

SC, CH and DB.

Based on the initial analysis of parameter combination 1, it was decided that the best clustering solutions for this wave parameter combination are given by configurations 1.3 using Ward linkage, and configuration 1.4 using average linkage. These configurations gave the best cluster boundaries based on the input data for θ_1, θ_2, and $H_{m0,P99.5}$ (see Figures F.3 and F.4). Now a choice must be made between these two solutions. The optimal clustering solution was assessed by comparing the clustering output against the input data for θ_1, θ_2, and $H_{m0,P99.5}$. In the end, Configuration 1.3 using Ward linkage yield the best results for wave parameter combination 1 because of its superior ability to define cluster boundaries between the input values for θ_1, θ_2, and $H_{m0,P99.5}$. Two notable regions where configuration 1.3-Ward performed better are:

- Between Denmark and Norway at the entrance of the Skagerrak. Here 1.3-Ward found a cluster boundary based on θ_1. 1.4-Average assigned these regions to the same cluster despite there being evidence that the values for θ_1 are not completely similar throughout the cluster. Furthermore, there is also a larger difference between the local water depth (See Figure 5.1), further supporting cluster divisions in this region;
- North West of the Dutch coast. Here 1.3-Ward identified a cluster boundary based on $H_{m0,P99.5}$ that was not identified by 1.4-Average.

From the comparative analysis of the clustering output between the different weight configurations, it becomes apparent that Average linkage is very sensitive for different weight configurations, because it emphasizes clustering divisions based on the input parameters that were assigned the most weight. Only for configuration 1.4, Average linkage provides well-defined cluster divisions based on θ_1, θ_2, N_θ and $H_{m0,P99.5}$. This is due to the characteristics of the linkage mechanisms. This further explained in the discussion.

5.1.1.3. Combination 2: Significant Wave Height and Peak period

The second parameter combination included θ_1, θ_2, N_θ, $H_{m0,P99.5}$ and T_P. The main goal of this parameter combination was to see whether the amount of information presented to the HAC algorithm

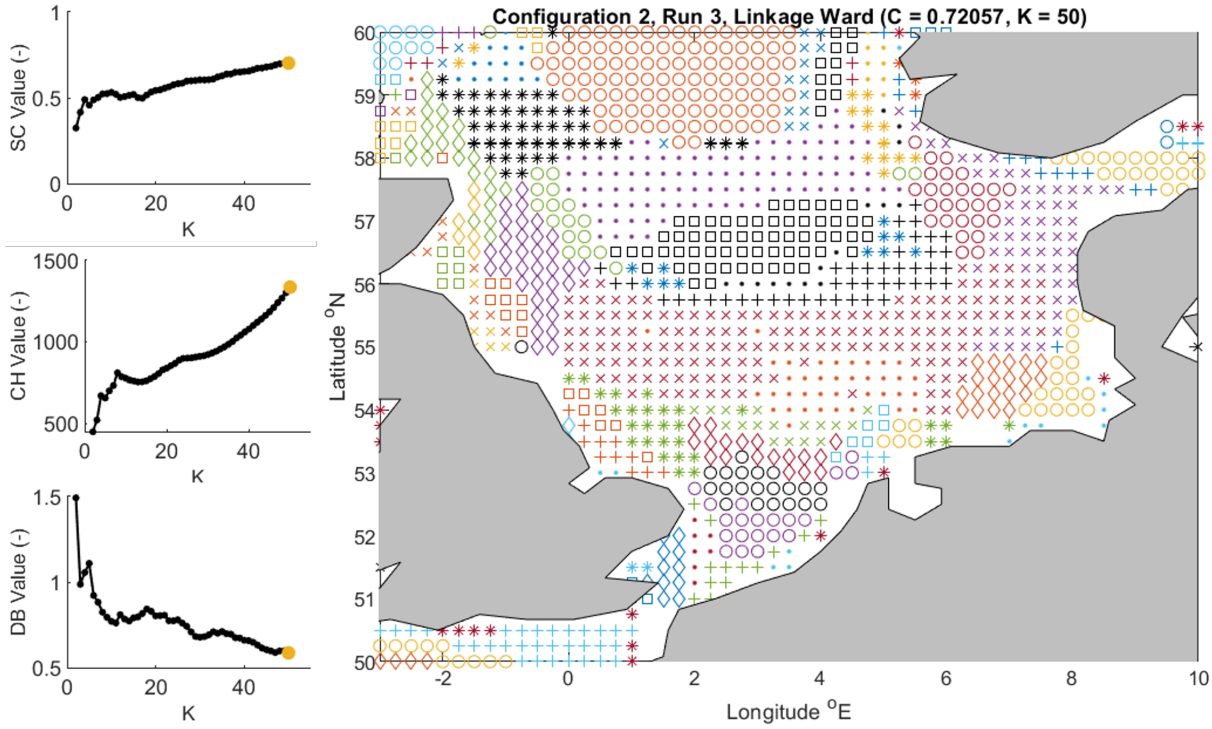


Figure 5.3: Final clustering solution for wave parameter combination 3.

could be increased without a significant loss of the clustering quality. The peak period (T_P) because of its importance in the design of hydraulic infrastructure next to H_{m0}. For configurations 2.1 and 2.2, Ward linkage yields the best clustering results with values for $\mathcal{K}$ equal to 50. Ward linkage provides better values for CH and better cluster divisions based on the input parameters. Moreover, Average linkage is very sensitive to the higher weights assigned to $H_{m0,P99.5}$ and T_P in configurations 2.1 and 2.2. It must be said that neither linkage mechanism scores particularly well for configuration 2.1 (see Appendix F for more information).

For configuration 2.3, Ward linkage with $\mathcal{K} = 50$ yielded the best clustering results. This is mainly because of better values for CH and SC and the better defined cluster divisions detected in the clustering output (see Figure F.8). Combined with the fact that values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ are closer for Ward linkage than for Average linkage, we can be certain that this is the best clustering solution for this configuration.

For configuration 3.4, Ward linkage yields the best results for $\mathcal{K}$ equal to 50. However, both Ward and Average linkage show significant bias regarding cluster boundaries based on θ_1 and θ_2. This can be explained by the fact that more weight was assigned to these parameters. However, more divisions for different values of $H_{m0,P99.5}$ and T_P are desired due the different design conditions that may be normative in these regions. During pre-processing, the input values for $H_{m0,P99.5}$ and T_P are rounded down to the nearest integer. This results high intra-cluster ranges for $H_{m0,P99.5}$ and T_P (in the order of 1 meter 1 second respectively). To ensure that the intra-cluster range of $H_{m0,P99.5}$ and T_P does not become to high, cluster boundaries based on these variables are desired. A close investigation of configuration 2.5 (see Figure F.10) quickly led to the conclusion that neither Ward nor Average linkage yielded suitable clustering results. Both linkage mechanisms result in strong bias towards cluster divisions based on θ_1 and θ_2. Cluster divisions based on $H_{m0,P99.5}$ and T_P were completely neglected.

In conclusion, for this parameter combination, only one clustering solution provided suitable results in light of this research, namely clustering solution 2.3 computed using Ward linkage with $\mathcal{K}$ equal to 50. The clustering output for this configuration can be seen in Figure 5.3.

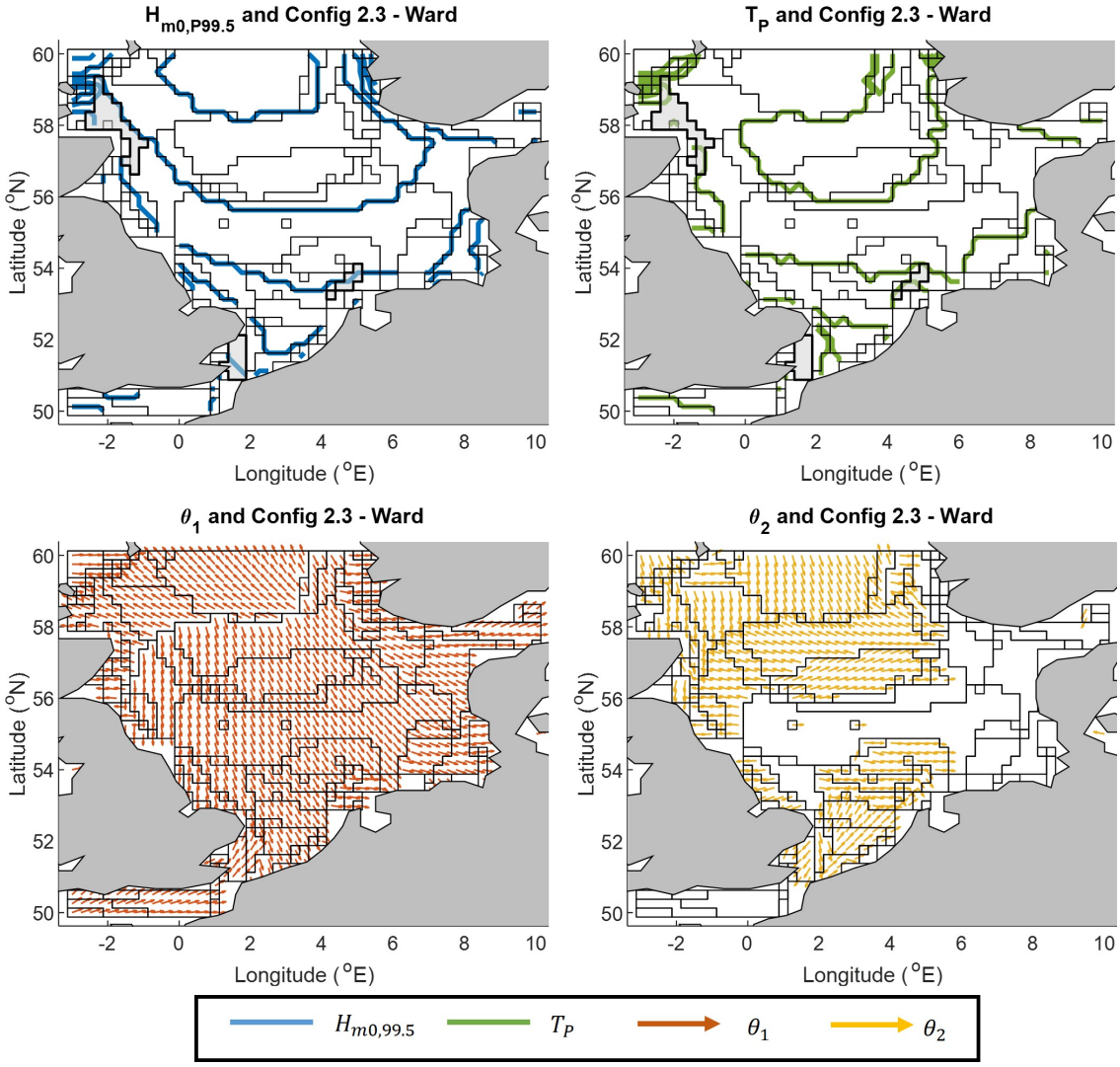


Figure 5.4: Comparison of the cluster output of configuration 2.3-Ward with the input parameters $H_{m0,P99.5}$ (top left), T_P (top right), θ_1 (bottom left) and θ_2 (bottom right). Top panels show contour lines for $H_{m0,P99.5}$ and T_P. The dominant wave directions (bottom panels) are represented by quivers. Areas of interest where the cluster boundaries do not respect the dissimilarities between the input parameters are given in shaded grey.

5.1.1.4. Selection Final cluster solution

In the end, the optimal clustering solution between the two parameter combination must be selected for further analysis. Preferably, the selected clustering output provides as much information as possible about the extreme wave characteristics in the North Sea. This allows engineers to have a better understanding about the relevant hydraulic boundary conditions throughout the North Sea. The main question is whether the inclusion of T_P as a clustering parameter next to $H_{m0,P99.5}$ does not reduce the quality of the clustering output too much. Because the internal evaluation metrics cannot be used to compare the quality of different parameter combinations (see Section 3.3.6), the best clustering solution is selected based on how well the clustering output respects the dissimilarities between the wave parameter input values. For parameter combination 1, configuration 1.3 using Ward linkage and $\mathcal{K} = 50$ gives the best clustering solution (see Figure 5.2). In this section, a careful analysis of the clustering output of configuration 2.3 is performed to study whether the divisions in the input data between θ_1, θ_2, $H_{m0,P99.5}$ and T_p are sufficiently respected by the cluster boundaries detected by the HAC algorithm.

The top left panel in Figure 5.4 shows that the clustering output for configuration 2.3 match the input divisions between $H_{m0,P99.5}$ throughout the research domain. The only region where this is not respected is found in the English Channel. Here the cluster (indicated by the grey shaded cluster in Figure 5.4)

is intersected by the contour representing the boundary between the input of $H_{m0,P99.5}$. The top right panel of Figure 5.4 shows that differences in T_P are well respected by the cluster boundaries in configuration 2.3. A close inspection of the T_P contours shows that there are two smaller regions where the cluster division do not match with the input divisions of T_P. The first region can be seen North-East of the Scottish coast, and the second region is North-West of the Dutch coast (indicated by the grey shaded cluster in Figure 5.4. Finally, a closer inspection of θ_1 and θ_2 shows that the cluster boundaries match the input boundaries in θ_1 (bottom-left panel). The bottom-right panel in Figure 5.4 shows that the cluster divisions for configuration 2.3 match very well with the values for θ_2 in the regions where secondary dominant wave directions are present.

Looking at the three areas where the cluster boundaries for configuration 2.3 did not respect the dissimilarities between the input values for $H_{m0,P99.5}$ or T_P, it can be observed that the clusters do respect input values ford θ_1 and θ_2. This means that these clusters provide extreme waves that typically result from the similar storms. Based on these findings, it was established that the missed boundaries based on values for $H_{m0,P99.5}$ or T_P are not problematic. The dominant wave directions are deemed more, because information about θ_1 and θ_2 helps to create homogeneous subsets that be used in extreme value analysis on the cluster level. Despite the clusters that miss divisions based on $H_{m0,P99.5}$ or T_P, configurations 2.2 provides more information about the hydraulic boundary conditions throughout the North Sea domain. This information is critical for the design and risk assessments in different areas of the North Sea. Furthermore, only 3 clusters were identified where the cluster boundaries did not respect the dissimilarities between the input values for $H_{m0,P99.5}$ and/or T_P. This number is not deemed problematic. It is therefore decided to use configuration 2.3-Ward in the remainder of this study.

5.1.2. Further Development of Cluster Map

In Section 5.1.1.4, we determined that the best initial clustering solution was given by configuration 2.3 using Ward linkage with $\mathcal{K}$ being equal to 50. Configuration 2.3 provides the most information about extreme wave characteristics for the significant wave height (H_{m0}) peak wave period (T_P), and dominant wave directions (θ_1 and θ_2). A closer inspection of Figure 5.3 shows that especially along in coastal regions, several smaller clusters that only contain very few grid-points can be identified. Furthermore, often grid-points are assigned to the same cluster even though they are in different geographical locations. The aforementioned observations potentially cause ambiguity when using the clustering output for the risk assessment of hydraulic infrastructure. Because this research is primarily aimed towards engineers, it is important to remove this ambiguity. A critical assessment of the clustering output is required to study whether the clustering output can be improved.

The first step is to remove grid-points outside the research domain. Next, clusters containing grid-points in different geographical locations will be addressed. The clusters will be split if there is reason to believe that the extreme wave characteristics differ between the geographical locations. Finally, clusters in the same geographical area will be compared to study whether these clusters share similar extreme wave characteristics so that these clusters can be merged.

5.1.2.1. Removing Points Outside Research Domain

Due to the selected boundaries of the study domain, two locations are included in the dataset while they are not of interest to this Research. This includes the points in the Irish Sea (West of Wales) and Kiel Bay (East of Denmark). These grid-points are removed from the final clustering solutions

5.1.2.2. Splitting Clusters Based on Geographical Locations

The next step is to split clusters that contain grid-points in different geographical locations. The applied HAC algorithm determines cluster boundaries based on input values for the significant wave height ($H_{m0,P99.5}$), peak period (T_P), and dominant wave directions (θ_1 and θ_2). Each of these parameters are represented by a single value. More information, either in the form of more wave parameters or more values representing the selected wave parameters, was not included in the HAC algorithm. Including more parameters comes with a loss of clustering quality. As a result, if the input values of H_{m0}, T_p, θ_1 and θ_2 are similar, they are assigned to the same cluster, regardless of their geographical locations. However, this does not necessarily mean that these regions share the same similar extreme wave characteristics. It is only a result of the limited information presented to the HAC algorithm.

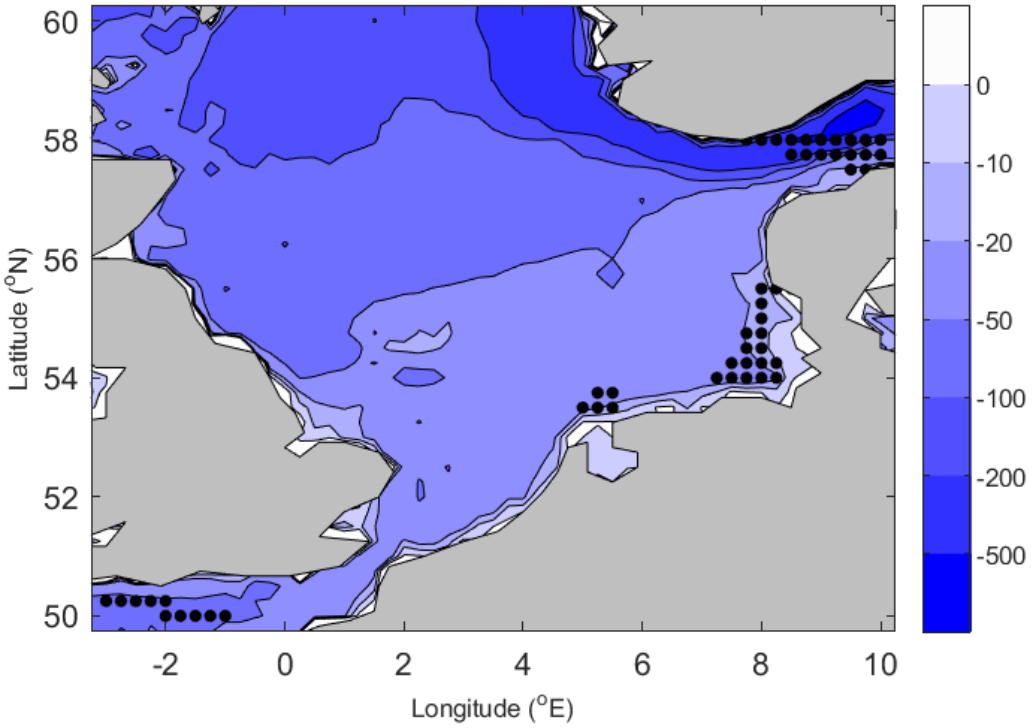


Figure 5.5: Example of a clusters with grid-points in different geographical locations. The map shows the bathymetry of the North Sea and grid-points belonging to Cluster 43. Notice how the local water depth differs for each sub-cluster.

The storm climate in the North Sea depends on the location of the low pressure area of the extra tropical cyclones (See chapter 2.1) This means that extreme wave events in different locations are typically the result of different storms. In this research, the time of occurrence of the extreme waves in different locations is not considered. As a result, there is no evidence that extreme waves in different regions of the North Sea are resulting from the same storm events. Furthermore, there are other parameters that were not included in the HAC analysis that are very important for extreme wave development. For example, local water depth influences wave development because larger depths can accommodate higher waves. Regions with larger depths have the potential to be affected by more extreme waves than shallower regions. The HAC algorithm considered the 99.5 percentile value of H_{m0} to represent H_{m0}. This means that the HAC algorithm has no knowledge about the most extreme waves at the grid-point.

In this research, clusters are split based on the geo-locations of the intra-cluster grid-points. A subsequent HAC analysis is employed with the coordinates of the grid-points within the established clusters as input values. The values for the geo-locations are normalized through z-score standardization to account for the differences in magnitude between latitude and longitude. The selected linkage mechanism is Average so that the considered dissimilarities is based on the centers of the clustered objects. The HAC algorithm then determines whether the different “sub-clusters” are too distant to be considered the same cluster. An example is presented in Figure 5.5. The intra-cluster grid-points have different geographical locations. The local water depth at each of these sub-clusters differs significantly, varying between approximately 20 meters and 500 meters depending on the location. The HAC analysis determined to split this cluster into four sub-clusters (see Figure 5.5).

5.1.2.3. Merging Clusters with Similar Extreme Wave characteristics

The final step is to merge smaller clusters with larger clusters if there is enough statistical evidence to do so. Especially in coastal regions, numerous clusters can be identified that consist of very few grid-points (see Figure 5.3). This potentially causes ambiguity to as to which cluster to select for extreme wave analysis. To complicate matters further, these smaller nearshore clusters typically have extreme wave statistics that are comparable to bigger, surrounding clusters. It is therefore decided to study the coastal

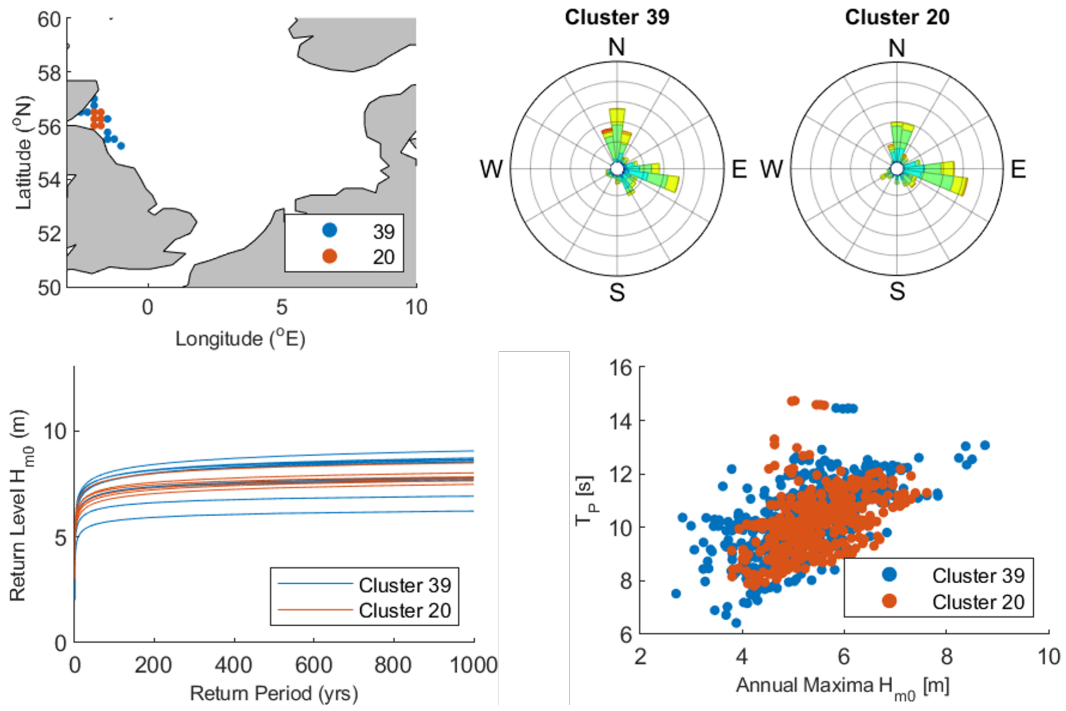


Figure 5.6: Example of two clusters that can be merged based on the same extreme wave statistics. The figure shows the locations of the considered cluster (top left), the dominant wave directions (top right), the return level plots for AM H_{m0} (bottom left) and the scatter plots for the peak period (bottom right).

regions to determine which clusters can be merged. The HAC algorithm identified clusters based on representative values for H_{m0}, T_P, θ_1 and θ_2. As a result, the extreme statistics of these parameters are used to determine whether clusters can be merged. A further requirement is that the clusters must have equal water depths. The latter requirement is especially important near the Norwegian trench, where the water depth nearshore rapidly increases to depths exceeding 500 meters. For each of the studied set of clusters, the Annual Maxima (AM) H_{m0} are computed at each grid-point with the corresponding values for T_p and θ. Using these observations, the following plots are produced:

- The AM H_{m0} observations are used to infer the distribution parameters of the Generalized Extreme Value (GEV) distribution using Maximum Likelihood Estimation (MLE) at each of the intra-cluster grid-points. The GEV parameters are subsequently used to produce return level plots for the extreme significant wave heights at the grid-points.
- The AM H_{m0} and corresponding θ observations are plotted on wave roses to compare the dominant wave directions between the considered clusters.
- The values for AM H_{m0} and corresponding T_p values are plotted on scatterplots to compare the T_p corresponding to extreme wave events in the considered clusters.

If the extreme waves in both clusters share similar statistics for H_{m0}, T_P and θ, both clusters are merged. An illustration is given in Figures 5.6. Here, illustrations are given for two clusters near the Scottish coast. The clusters of interest share very similar extreme wave statistics. Therefore, it was decided to merge these clusters. This process is repeated until no other neighbouring clusters can be found that share similar extreme wave statistics. It should be noted that this step is very subjective and requires a critical assessment by the modeller. No diagnostics checks where performed to validate the resulting clustering output.

5.1.3. Final Clustering Output

The final clustering output is presented in Figure 5.7. After the process of splitting and merging clusters has been completed, the North Sea domain is partitioned in 63 different clusters. This means that the value for $\mathcal{K}$ is higher than $\mathcal{K}_{max}$ that was imposed prior to running the HAC algorithm. This can

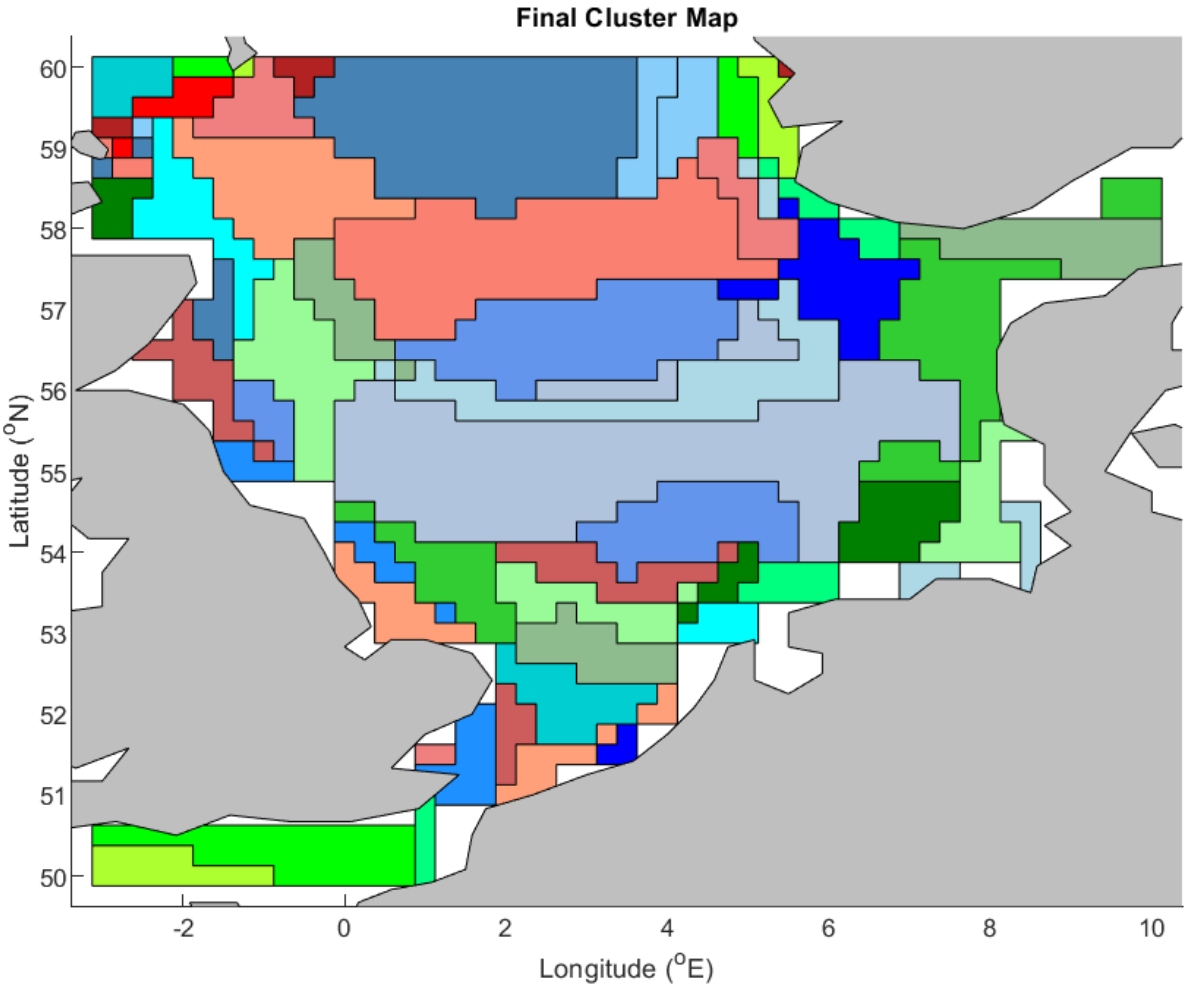


Figure 5.7: Final cluster map that is used in the remainder of this research.

be explained by the fact that the HAC-algorithm detects partitions grid-points in different geographical locations into clusters. These clusters were subsequently split in Section 5.1.2. It could be suggested to increase $\mathcal{K}_{max}$ prior to running the algorithm. However, because the grid-point coordinates were not included as input, the HAC algorithm still only detects cluster boundaries based on the input values for $H_{m0,P99.5}$, T_P, θ_1 and θ_2. Increasing $\mathcal{K}_{max}$ will result in cluster boundaries based on θ_1 and θ_2, which is not necessarily required. This is explained in more detail in the Discussion.

Apart from the final cluster map, it is also interesting to study the nature of the cluster boundaries. In the HAC analysis, the clusters were defined based on representative values for H_{m0}, T_p, θ_1 and θ_2. This means that each cluster boundary must represent a boundary between different values for H_{m0}, T_p, θ_1 and/or θ_2. Figure 5.8 shows the final cluster map with the contours for the input values for $H_{m0,99.5}$, T_p, θ_1 and θ_2. Figure 5.8 can be used to determine whether the cluster boundaries between different clusters are based on $H_{m0,99.5}$, T_p, θ_1 or θ_2. What can be seen in Figure 5.8 is that every cluster boundary represents at least one division based on the input values.

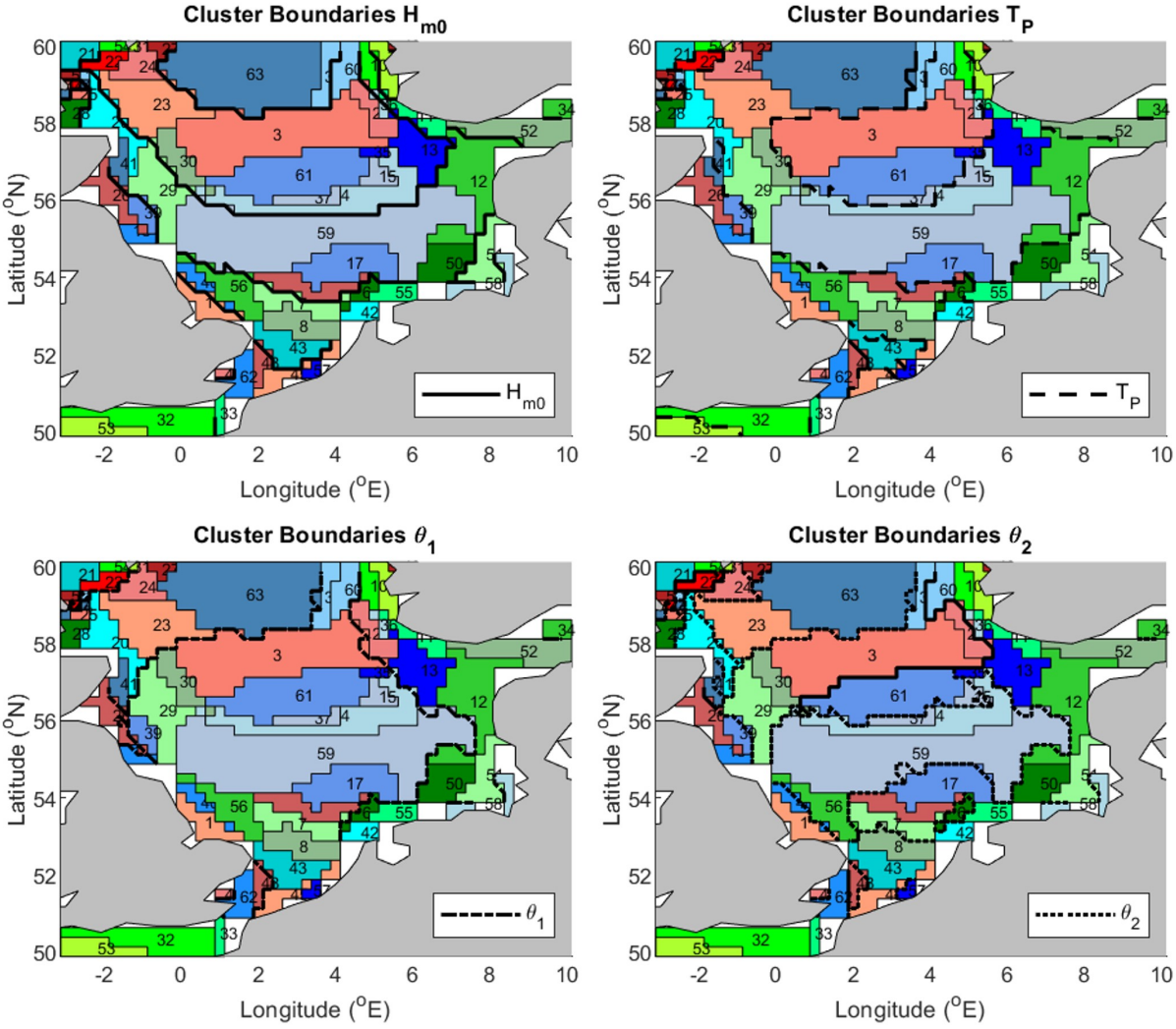


Figure 5.8: Final cluster map with the cluster boundaries based on H_{m0} (top left), T_p (top right), θ_1 (bottom left) and θ_2 (bottom right).

Table 5.2: Clustering results for parameter combination 1 (θ_1, θ_2, N_θ & $H_{m0,P99.5}$). From left to right, first, the cophenetic correlations (C) are given for the linkage mechanisms Ward (W), Average (A), Complete (C), and Single (S). Then for the two best linkage mechanisms, the values for Silhouette coefficient (SC), Calinski-Harabasz criterion (CH) and the Davies-Bouldin value (DB) are given and the ideal number of clusters as given by the evaluation criterion. For both linkage mechanisms, the ideal number of clusters is marked as $\mathcal{K}_{opt}$. For each configuration, the linkage mechanism with the best clustering results in marked in bold.

Config.	C				**L1**	SC		CH		DB		$\mathcal{K}_{opt}$	**L2**	SC		CH		DB		$\mathcal{K}_{opt}$
	W	A	C	S		$\mathcal{K}_{SC}$	SC	$\mathcal{K}_{CH}$	CH	$\mathcal{K}_{DB}$	DB			$\mathcal{K}_{SC}$	SC	$\mathcal{K}_{CH}$	CH	$\mathcal{K}_{DB}$	DB	
1.1	0.73	0.78	0.70	0.61	A	42	0.82	50	4986	50	0.19	45	W	35	0.79	50	12290	49	0.27	50
1.2	0.63	0.80	0.62	0.61	A	50	0.80	50	2101	50	0.22	50	W	38	0.77	50	5257	50	0.31	50
1.3	0.77	0.89	0.61	0.67	A	50	0.69	46	1227	50	0.35	48	W	50	0.74	50	3839	50	0.38	50
1.4	0.80	0.90	0.65	0.70	A	50	0.70	50	2205	50	0.37	50	W	46	0.71	50	3869	49	0.44	48
1.5	0.84	0.90	0.82	0.71	A	8	0.59	47	1813	30	0.39	30	W	43	0.67	50	4206	42	0.49	43

Table 5.3: [Clustering results for parameter combination 2]Clustering results for parameter combination 2 (θ_1, θ_2, N_θ & T_P). From left to right, first, the cophenetic correlations (C) are given for the linkage mechanisms Ward (W), Average (A), Complete (C), and Single (S). Then for the two best linkage mechanisms, the values for Silhouette coefficient (SC), Calinski-Harabasz criterion (CH) and the Davies-Bouldin value (DB) are given and the ideal number of clusters as given by the evaluation criterion. For both linkage mechanisms, the ideal number of clusters is marked as $\mathcal{K}_{opt}$. For each configuration, the linkage mechanism with the best clustering results in marked in bold.

Config.	C				**L1**	SC		CH		DB		$\mathcal{K}_{opt}$	**L2**	SC		CH		DB		$\mathcal{K}_{opt}$
	W	A	C	S		$\mathcal{K}_{SC}$	SC	$\mathcal{K}_{CH}$	CH	$\mathcal{K}_{DB}$	DB			$\mathcal{K}_{SC}$	SC	$\mathcal{K}_{CH}$	CH	$\mathcal{K}_{DB}$	DB	
2.1	0.63	0.82	0.69	0.52	A	49	0.64	4	681	50	0.48	40	W	44	0.79	50	2471	50	0.57	50
2.2	0.63	0.82	0.67	0.62	A	50	0.58	34	440	50	0.56	44	W	50	0.76	50	1447	49	0.53	50
2.3	0.72	0.86	0.48	0.72	A	50	0.56	50	611	46	0.53	50	W	50	0.70	50	1336	50	0.59	50
2.4	0.83	0.92	0.74	0.78	A	12	0.59	46	1007	31	0.45	48	W	49	0.63	50	2174	50	0.59	50
2.5	0.84	0.92	0.82	0.76	A	7	0.71	25	3092	15	0.33	25	W	5	0.69	50	5794	15	0.53	50

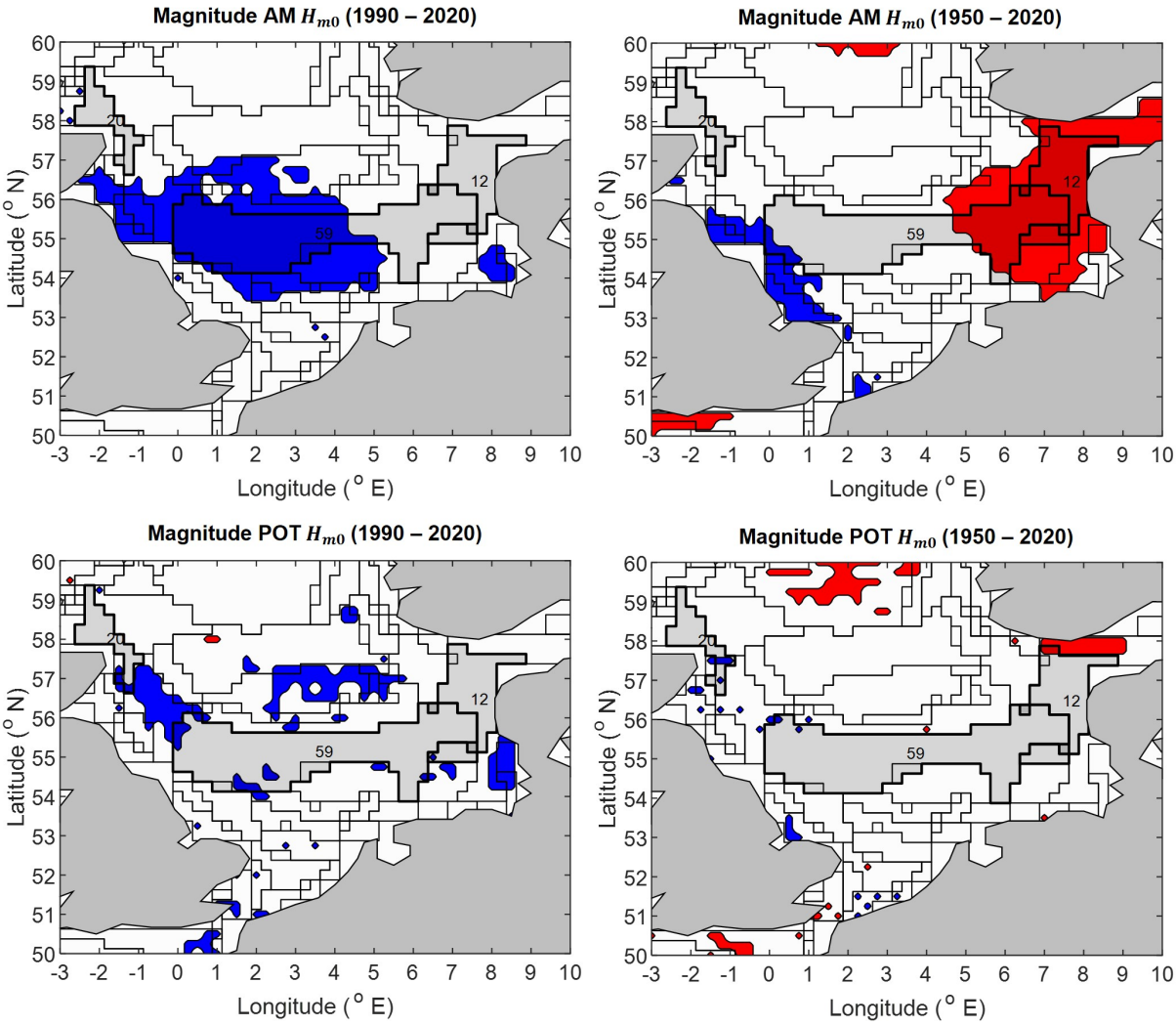


Figure 5.9: Results of the Mann-Kendall test for extreme H_{m0} as determined in Section 4.2.2 and clustering output developed in Section 5.1.

5.2. Intra-Cluster Extreme Value Analysis

In this Section, extreme value analysis (EVA) at the cluster level will be evaluated in relation to EVA at grid-point level to study the applicability of a clustering approach to evaluate the extreme wave statistics in the North Sea. The extreme value models will be inferred by means of fitting the Generalized Extreme Value (GEV) distribution using Bayesian Inference. Moreover, a non-stationary component is added to the extreme value models to account for the non-stationarity in the extreme wave climate. Computing representative extreme value models at cluster level requires a critical assessment of the extreme value properties at the intra-cluster grid-points. Therefore, the non-stationarity of the extreme significant wave height (H_{m0}) and extreme wind speeds (U_{10}) in studied in more detail within the established clusters from Section 5.1. The temporal trend analysis at cluster level is presented in Section 5.2.1.

Five techniques will be evaluated to model the extreme H_{m0} at cluster level. The extreme H_{m0} observations will be detrended and representative time-series are derived before the parameters of the Generalized Extreme Values (GEV) are inferred using Bayesian Inference. After the GEV-models have been derived, a non-stationary component is added to the GEV-models to account for potential non-stationarity extreme H_{m0}. The results of the representative extreme value models are presented in Section 5.2.2.

Figure 5.9 shows the established clustering output from 5.1 and the results for the temporal trend analysis of H_{m0} for the North Sea as was performed in Section 4.2.2. Three different clusters will be evaluated in this section. Each cluster has unique characteristics in terms of its location in the North Sea domain and the degree of non-stationarity it contains. The following clusters are included:

- Cluster 12 (Danish Coast) which shows an upward trend for the extreme wave climate between 1950 and 2020 (see Figure 5.9);
- Cluster 20 (Scottish coast): Here no temporal trend was detected for extreme H_{m0} in both considered temporal horizons in Figure 5.9;
- Cluster 59 (Central North Sea), which is the biggest cluster. Furthermore, Figure 5.9 shows that this cluster has opposite trends for extreme H_{m0} between 1950 and 2020.

5.2.1. Intra-Cluster Temporal Trend Analysis

Before analyzing the representative extreme value models for H_{m0} at the cluster level, it is important to have an understanding about the intra-cluster non-stationarity of H_{m0}. For studying the non-stationarity of H_{m0}, extreme observations are selected using Annual Maxima (AM) and Peak over Threshold (POT). The presence of a statistically significant trends is detected using the Mann-Kendall (MK) test and the linear slope of the trends is estimated using the Theil-Sen (TS) test.

Information about the dominant wave directions (θ_1 and θ_2) can be applied to improve extreme value selection to form independent and homogeneous subsets of extreme H_{m0} observations in the clusters. Hereby meeting the requirements for POT (see Section 3.1). Information about θ_1 is used to filter the H_{m0} time-series so that only waves travelling in similar directions are maintained.

New values for the threshold (u) and declustering time lag (δ) are selected for the filtered time-series using the diagnostic tests for Mean Residual Life (MRL), Extremal Index (EI), Dispersion Index (DI) and the stability of the parameters of the Generalized Pareto (GP) distribution. It is assumed that the found values for u and δ at the centroid of the cluster are representative for the entire cluster. A detailed description of the selection of u and δ for this analysis is presented in Appendix E. The boundaries for the dominant wave directions and the newly obtained values for u and δ are given in Table 5.4. The directionally filtered H_{m0} observations are also used for AM selection of extreme H_{m0}. The intra-cluster temporal trends for the extreme wind climate were also studied. For the extreme value selection of U_{10}, only AM was applied. The time-series for U_{10} was filtered to construct a time-series containing only U_{10} observations with a direction matching the range of dominant wave directions. The sections below present the results of the temporal trend analysis for each of the three clusters of interest.

5.2.1.1. Cluster 12: Danish Coast

First, the results for the temporal trend analysis for Annual Maxima (AM) H_{m0} are studied. It can be observed that there are no differences in the detected temporal trends when accounting for the intra-cluster dominant wave directions compared to the findings in Section 4.2.2. This can be explained by the fact that cluster 12 is characterized by only one dominant wave direction. This means that the selected AM H_{m0} observations in the general trend analysis (Section 5.9) and the intra-cluster temporal

Table 5.4: Selected Boundaries for the dominant wave directions (θ), threshold (u) and declustering time lag (δ) for the selected clusters subjected to intra-cluster extreme value analysis. The subscripts LB and UB correspond to the lower and upper bound of the applied dominant wave direction respectively.

Location	Cluster	θ_{LB}	θ_{LB}	u	δ
Dutch Coast	8 (θ_1)	270	360	$P_{99.0}\, H_{m0}$	2
	8 (θ_2)	180	270	$P_{99.5}\, H_{m0}$	2
Danish Coast	12	210	330	$P_{99.5}\, H_{m0}$	5
Scottisch Coast	20	60	180	$P_{99.5}\, H_{m0}$	2
Central North Sea	59	260	360	$P_{99.5}\, H_{m0}$	5

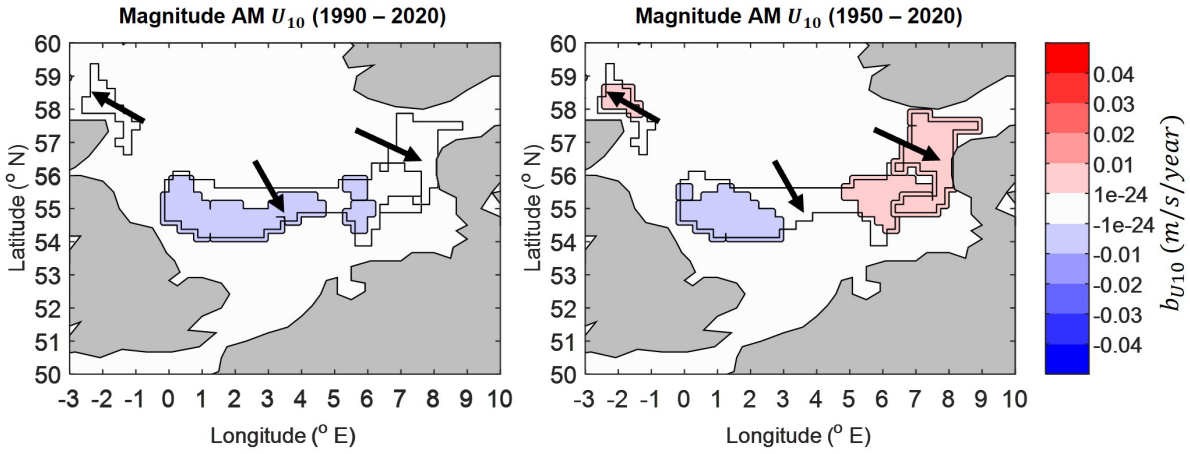


Figure 5.10: Intra-cluster temporal trend analysis for AM U_{10}. Left and right show the results of the short and long term analysis respectively. Black quivers represent the dominant wave direction in each cluster.

trend analysis relate to the same wave directions. Hence, similar temporal trends are detected. More interesting are the results of the POT analysis. In Section 4.2.2, both the long- and the short-term analyses showed no temporal trends in the region of cluster 12 (see Figure 5.9). However, when accounting for the intra-cluster dominant wave direction, small areas with an upward trend for extreme H_{m0} can be observed for the long-term analysis. The magnitude of the trends varies between 1 and $2\,cm/year$. The observed non-stationarity is accompanied by an increase in Frequency (N_{exc}). This suggests that accounting for dominant wave directions influences the detection of statistically significant trends.

Shifting attention to the temporal trend analysis for the extreme wind speeds (U_{10}). When comparing extreme U_{10} and H_{m0}, it can be observed that comparable temporal trends are detected. This includes no statistically significant trends between 1990 and 2020 and an upward trend for extreme U_{10} between 1950 and 2020. This means that the upward trend for extreme H_{m0} between 1950 and 2020 in cluster 12 can at least partially be explained by an increase in the extreme wind climate.

5.2.1.2. Cluster 20: Scottish Coast

In Section 4.2.2, Cluster 20 showed no statistically significant temporal trends in extreme H_{m0} for both the short and long term horizon (see also Figure 5.9). When accounting for θ_1, the short term trends for extreme H_{m0} selected using AM show few grid-points where a downward trend in extreme H_{m0} can be observed. For the long term, also a few grid-points can be observed where a downward trend was detected. This is consistent with our findings in Section 4.2.2, where downward trends for extreme H_{m0} were observed along the British coast for both temporal horizons. These findings suggest that when a grid-point is characterized by multiple dominant wave directions, accounting for the dominant wave directions helps to detect statistically significant temporal trends in extreme H_{m0}. Despite the fact that few grid-points show statistically significant trends, the slope of the trend in all grid-points is below $5\,mm/year$. It therefore possible to state that cluster 20 is not characterized by steeps trends in extreme H_{m0}.

Looking at the temporal trend analysis of extreme U_{10}, no statistically significant trends were detected for the short term horizon. For the long term horizon, few grid-points exhibit an upward temporal trend, which is surprising given the downward trend for extreme H_{m0}. It appears that the small upward trend in extreme $U10$ does not transfer into an upward temporal trend for extreme $Hm0$, meaning that other processes are responsible for the downward trend in extreme H_{m0}.

5.2.1.3. Cluster 59: Central North Sea

The general temporal trend analysis (Section 4.2.2) showed a strong downward trend for H_{m0} in the Western regions of cluster 59 for the short term analysis (see Figure 5.9). For the long term analysis, an upward trend for extreme H_{m0} was observed in the East, and a downward trend was observed at

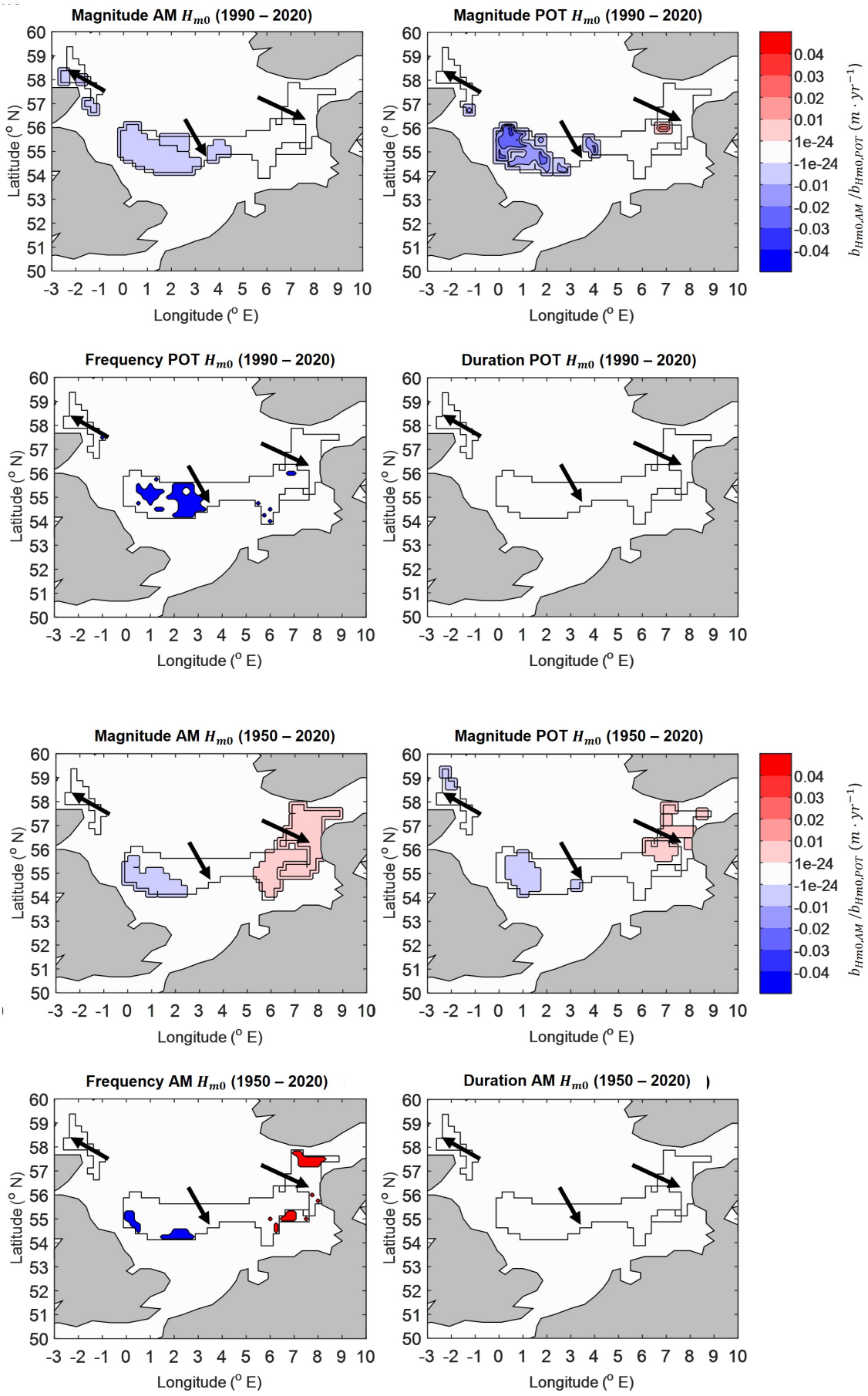


Figure 5.11: Results of the intra-cluster temporal trend analysis. Top and bottom figures show the results for the long and short term analysis respectively. Top-left figures show temporal trends in magnitude of AM H_{m0}. Top-right figures show trends in magnitude of POT H_{m0} extremes. Bottom-left figures show trends in the threshold exceedance frequency of H_{m0} (N_{exc}) and bottom-right shows temporal trends for average duration of threshold exceedance (T_{dur}). Black quivers represent the dominant wave direction selected for the cluster.

the Western border of cluster 59. Accounting for the dominant wave directions enhances the presence of these trends. For the short term analysis, a downward trend (magnitude between 1 and $2\,cm/year$) is observed in cluster 59. For the long term, it is observed that the larger region in the East of cluster 59 is affected by an upward trend in extreme H_{m0} compared to the general trend analysis. Meanwhile, the Western part of cluster 59 is characterized by a larger area exhibiting a decreasing trend in extreme H_{m0}. The findings for the temporal trends in extreme H_{m0} are further supported by similar trends in extreme U_{10}. The short term horizon shows a predominantly downward temporal trend throughout the cluster (approximately $0.1\,m/s/year$). The long term analysis shows a predominantly upward trend for U_{10} in the East, and a downward trend in the West. The magnitudes of these trends are approximately also $0.1\,m/s/year$.

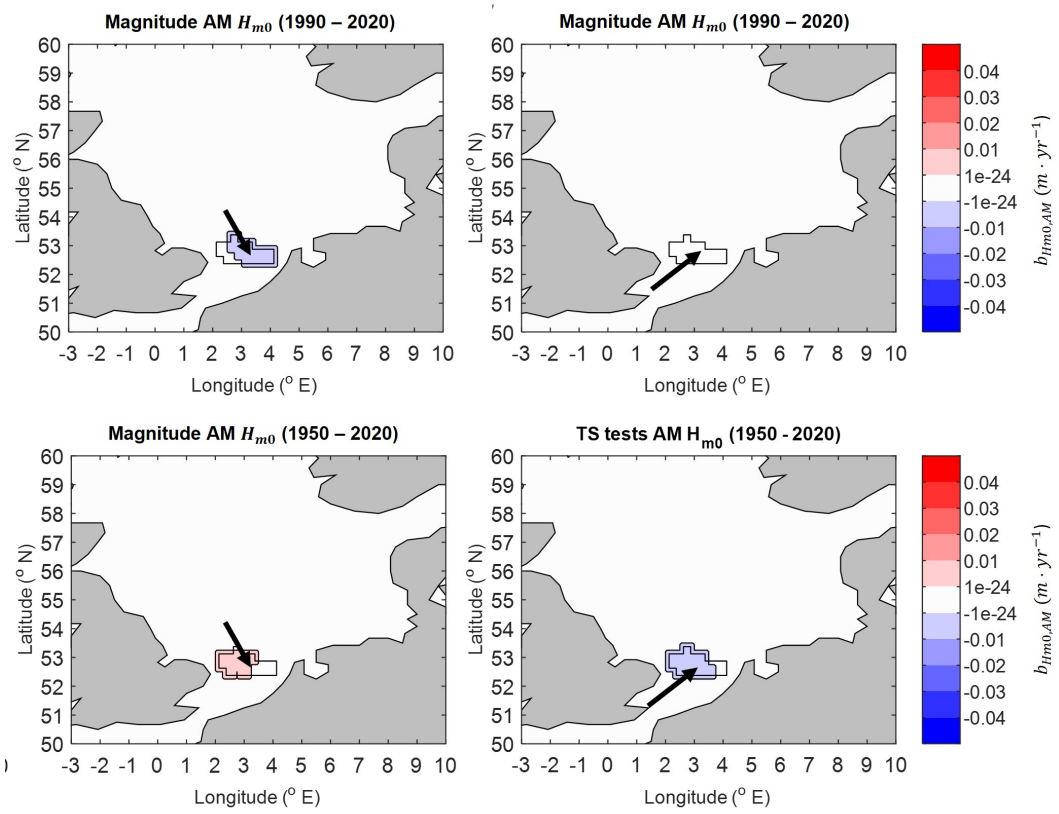


Figure 5.12: Results for the intra-cluster temporal trend analysis for cluster 8. Top and bottom figures show the short and long temporal horizons, respectively. Left and right figures represent the primary (θ_1) and secondary (θ_2) dominant wave direction, respectively. Black quivers represent the dominant wave directions.

5.2.1.4. Effects of Multiple Dominant Wave Directions

In Section 5.2.1.2, it was noticed that accounting for the dominant wave direction (θ_1) potentially results in the detection of statistically significant temporal trends at grid-points where initially no temporal trend was detected in Section 4.2.2. To study this phenomenon in more detail, it was decided to study another cluster where initially no trends are detected. For the design and risk assessment of hydraulic infrastructure along the Dutch coast, typically two dominant wave directions must be considered, namely waves from the Northwest (θ_1 in Figure 4.8) and from the Southwest (θ_2). To this end, cluster 8 was selected. This section provides a short case-study studying the different effects of θ_1 and θ_2 on the temporal trend analysis. Table 5.4) presents the relevant boundaries for θ_1 and θ_2 as well as the corresponding threshold (u) and declustering time-lag (δ). Figure 5.12 presents the results of the temporal trend analysis of cluster 8.

Figure 5.12 shows that accounting for the dominant wave directions has an important effect on the detection of statistically significant temporal trends. More importantly, it can be observed that different temporal trends can be present depending on the considered dominant wave directions. For the

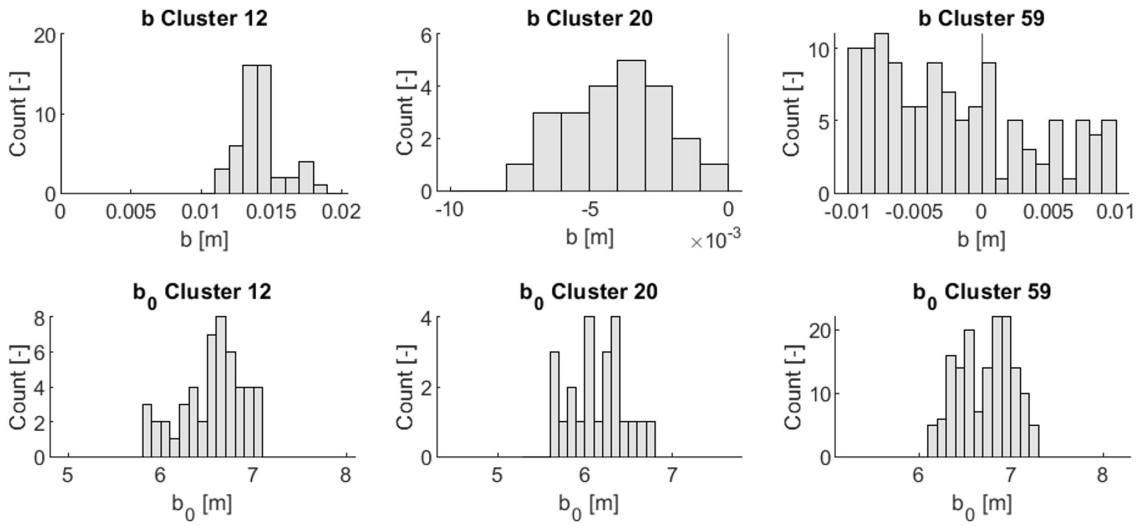


Figure 5.13: Histograms showing the TS estimators b and b_0 fitted to the AM H_{m0} observations used for detrending.

short temporal horizon, it can be seen that a downward trend exist for the AM H_{m0} coming from the Northwest. The magnitude of this trend is approximately $1\ cm/year$. These findings are consistent with our findings for the short term analysis in Section 4.2.2. However, the AM H_{m0} coming the from Southwest show no statistically significant trend. Shifting our attention to the long term analysis, it can be seen that an upward trend is present for waves coming from the Northwest, but a downward trend is observed for waves coming from the Southwest. For both wave directions, the magnitude of the trends is approximately $1\ cm/year$.

This section shows that different temporal trends could exist for different dominant wave directions in the same cluster. The presence of temporal trends in extreme H_{m0} affects the computation of non-stationary extreme value models. Although the specific effects of the two dominant wave direction are not further investigated in this research, opposing trends are important to consider when designing hydraulic infrastructure as they potentially result in a shift in normative wave direction, and subsequently design conditions. The potential impacts of different (and opposing) temporal trends for different wind directions will be further explored in the Discussion.

5.2.2. Extreme Value Models at Cluster Level

This section evaluates the performance of the derived Extreme Value Models at Cluster Level. Prior to fitting the models, AM H_{m0} observations are detrended to them homogeneous. Five methods to construct representative time-series are evaluated for the three clusters of interest. More information about the representative time-series can be found in Section 3.5.2. As mentioned in Section 3.5.3, Bayesian Inference results in posterior distributions for the GEV distribution describing detrended significant wave height:

$$H^*_{m0} \sim GEV\left(\mu^*, \sigma^*, \xi^*\right) \tag{5.1}$$

The temporal trend information (b and b_0) is added to μ^* to account for potential non-stationarity, resulting in the following extreme value distribution:

$$H_{m0} \sim GEV\left(\mu^* + \left(b \cdot t + b_0\right), \sigma^*, \xi^*\right) \tag{5.2}$$

First, the performance of the GEV-models is assessed for $\Psi^*_{GEV}(H^*_{m0})$ to study which of the five representative time-series results in an extreme value model that best describes H_{m0} at cluster level. After that, the performance of the $\Psi_{GEV}(H_m 0)$ models is evaluated in light of the assumed temporal trends.

5.2.2.1. Detrending Annual Maxima Significant Wave Height

Before evaluating the extreme value models, a close inspection of the intra-cluster values of the TS estimators b and b_0 is presented. The TS estimators b and b_0 used to detrend the data can be found in

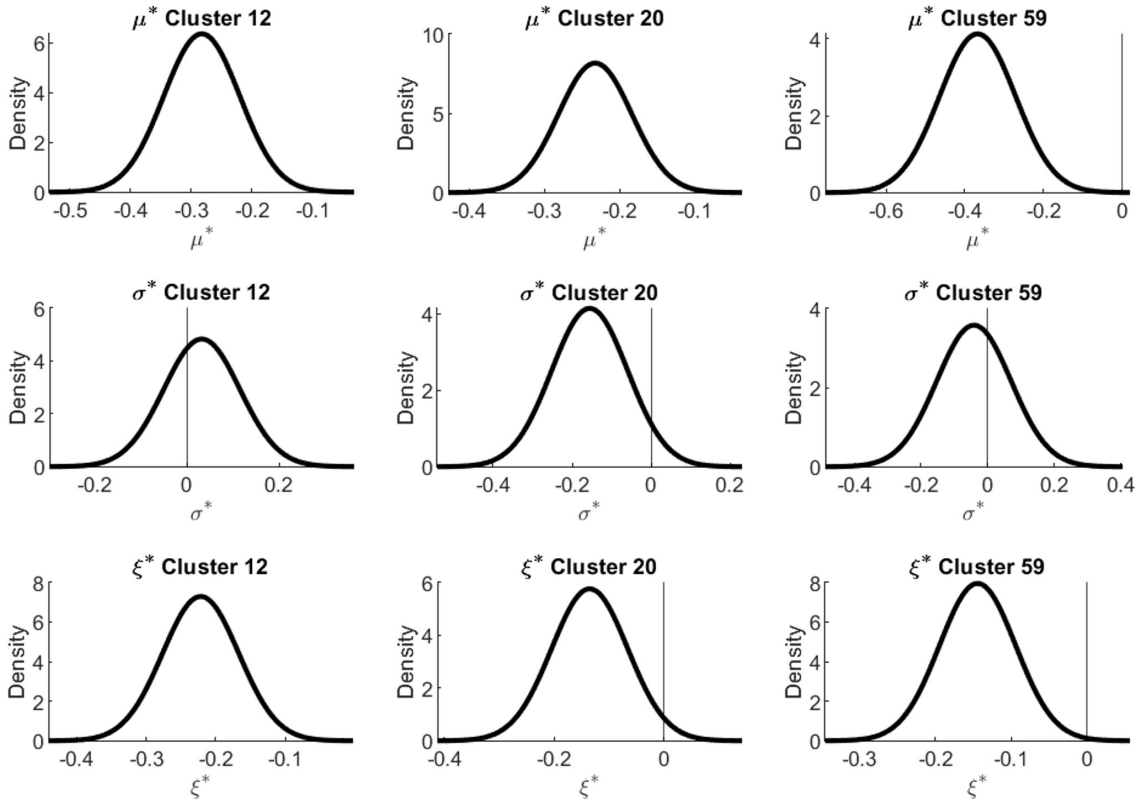


Figure 5.14: Normal priors for the location (μ^*), scale (σ^*) and shape (ξ^*) parameters for the considered clusters.

Figure 5.13. The values for b_0 have a spread in the order of a meter for all clusters. This is a result of pre-preprocessing of the HAC input before clustering. The parameters representing H_{m0} were rounded down to the nearest integer. This means that the range for extreme waves is still relatively large. For b, it can be verified that Cluster 12 is characterized by an upward trend only, with values for b between 1 and $2\ cm/yr$. Despite the fact that for cluster 20, the MK-test did not detect significant monotonic trends (see Section 5.2.1), some downward trends can be seen in Figure 5.13. However, the majority of these trends are smaller than $5\ mm/yr$. Cluster 59 is characterized by a opposite trends varying between -1 and $1\ cm/yr$).

5.2.2.2. Prior Distributions

The GEV parameters of the representative models are determined by means of Bayesian Inference (BI). The normal priors ($p(\theta^*)$ are constructed using the intra-cluster information about the GEV parameters (μ^*_{gp}, σ^*_{gp} and ξ^*_{gp}) at the grid-point level. See Section 3.5.3.1 for more information. Figure 5.14 shows the informative priors for the considered clusters. It can be observed that all values for μ^*_{gp} are negative as a result of the detrending process. Furthermore, it can be noticed that nearly all values for ξ^*_{gp} are negative indicating that most GEV distributions at the grid-point level have an upper bound. However, few grid-points in cluster 20 show positive values for ξ^*, indicating that for these grid-points, the GEV distributions do not have an upper bound. Potentially this is the result of the deeper water depth prevailing in the region of this cluster. Deeper waters allow higher waves to exist, thereby removing the upper bound from the GEV distributions.

5.2.2.3. Posterior Distributions

The posterior distributions for the 5 representative extreme value models in clusters 12, 20 and 59 can be seen in Figure 5.15. The median of the posteriors is selected to represent the best fit, and the 5th and 95th percentile values are selected to construct the 90% confidence interval.

Figure 5.15 shows that between the 5 representative extreme value models, parallel maximum aggregation yields the highest values for μ^*. This is a direct result of the fact that parallel maximum aggregation

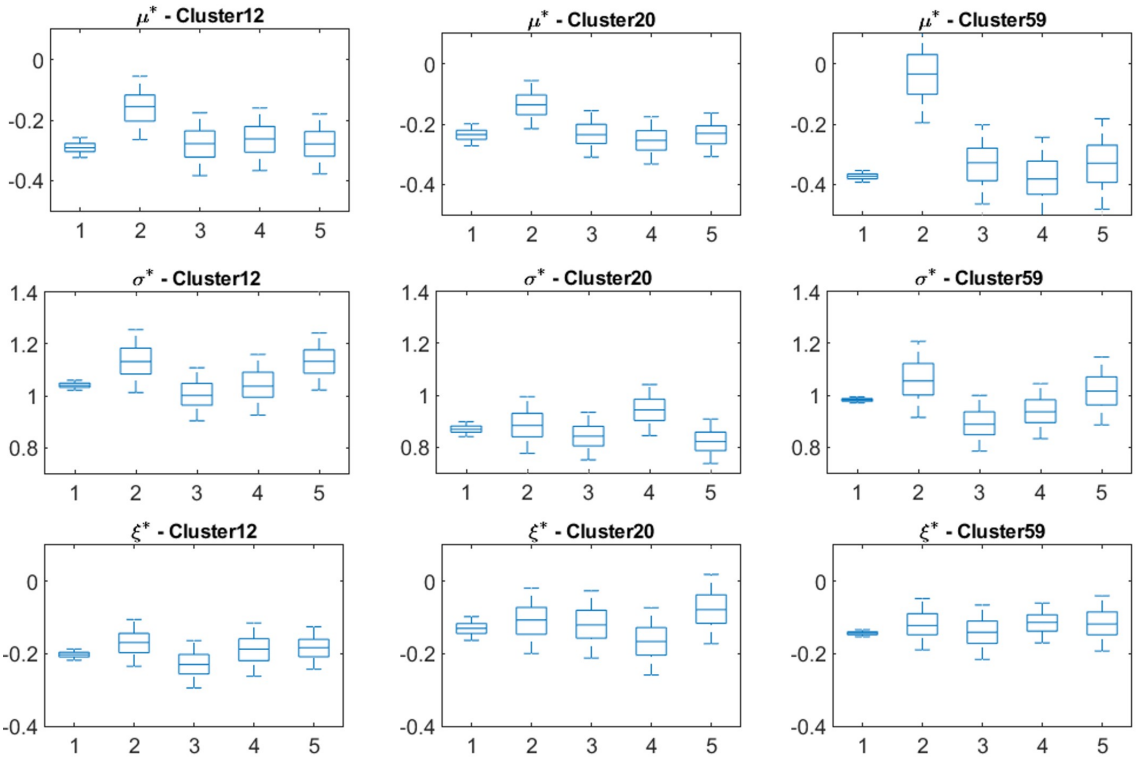


Figure 5.15: Box plots showing the posterior distributions for the five representative $\Psi^*_{GEV}(H^*_{m0})$ models in the different clusters. The center of the box indicates the best fit value (P_{50}) for μ^*, σ^* and ξ^*. The top and bottom whiskers indicate the width of the 90% confidence interval (i.e P_5 and P_{95}) for μ^*, σ^* and ξ^*.

selects the highest intra-cluster H^*_{m0} for every year to construct the representative time-series. Comparing the three clusters, it can be seen that the values for the values μ^* are all in the same order of magnitude, despite the wave heights varying between the different clusters. This can be attributed to the fact that prior to Bayesian Inference, the AM H_{m0} observations have been detrended. The "amount of detrending" depends on b_0. Higher values for AM H_{m0} correspond to higher values for b_0. This can also be seen in Figure 5.13. Clusters 12 and 59 have higher values for extreme H_{m0} than cluster 20, and therefore also have higher values for b_0. After detrending, all values for H^*_{m0} are in the same order of magnitude, meaning that the values of μ^* after Bayesian Inference are more likely to show the same order of magnitude between the different clusters.

Looking at ξ^*, it can be seen that nearly all clusters and representative time-series show values for ξ^* below zero, indicating that the GEV models are unbounded. Only parallel maximum and the selection of the grid-point with the highest return level for cluster 20 show that the 95th percentiles of ξ^* exceed zero, meaning that the upper confidence bounds of these models are unbounded. Furthermore, the values for ξ^* in cluster 20 are higher compared to clusters 12 and 59. Cluster 20 is located Northeast of the Scottish coast, where higher water depths prevail compared to clusters 12 and 59 (see Figure 4.2), Possibly, extreme waves in cluster 20 are less affected by physical depth limitations, resulting in relatively high values for ξ^*.

Furthermore, it can be noticed that the confidence interval of the posterior distributions for series aggregations is much narrower than for the other representative time series. This is caused by the larger sample size and the fact that the variability of samples is relatively low. Detrending the AM H_{m0} observations reduces the variability of the H^*_{m0} values, because the value for b_0 is higher for grid-points exhibiting higher AM H_{m0} values. The reduced variability and the relatively large sample size ($N_{gp}{\cdot}T_{hor}$) cause the confidence interval to become narrow. Naturally, the width of the confidence interval also depends on the cluster size. Figure 5.15 also shows that the confidence interval for series aggregation of cluster 59 is smaller than for clusters 12 and 20.

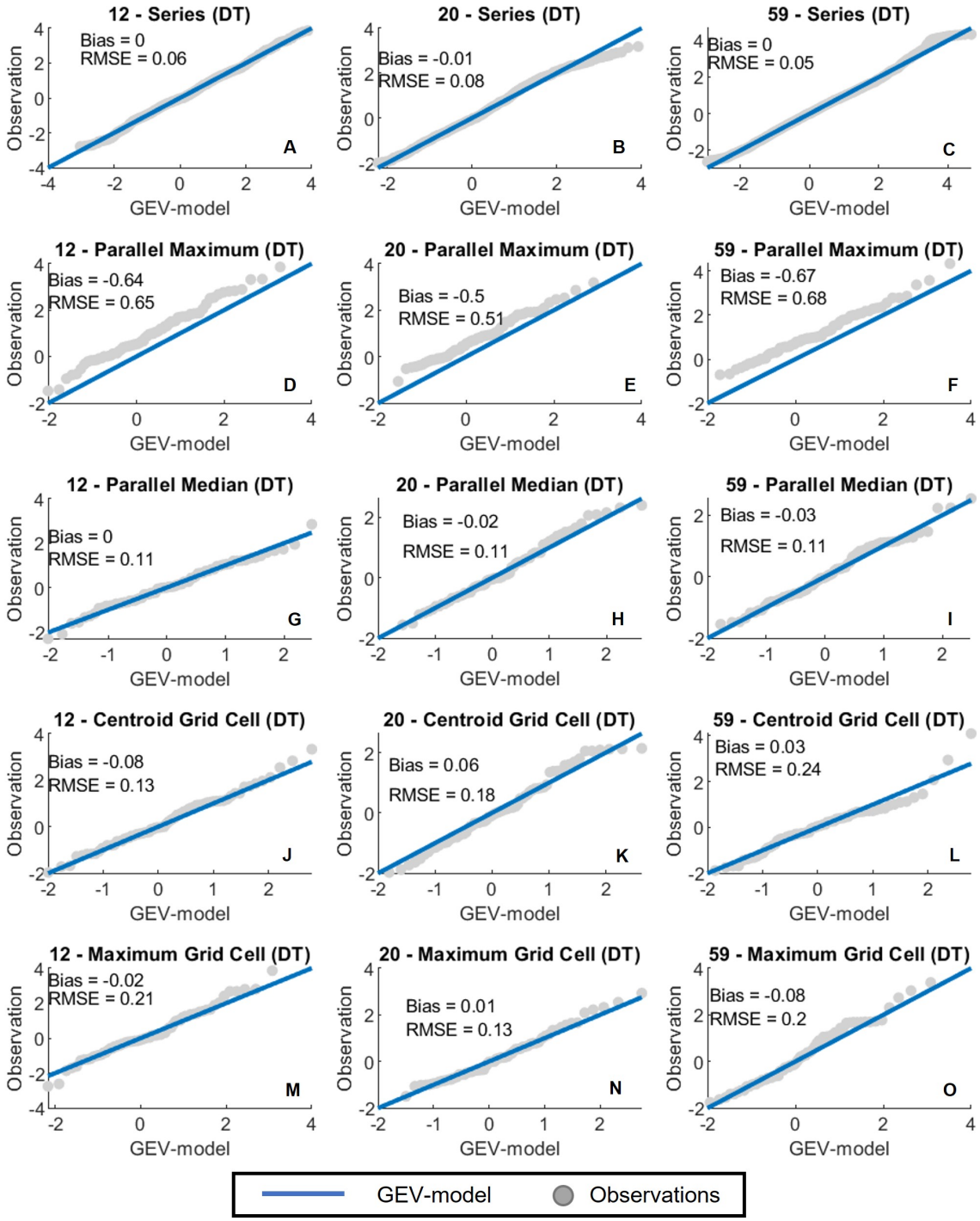


Figure 5.16: QQ-plots comparing $\Psi^*_{GEV}(H^*_{m0})$ models with the representative H^*_{m0} time-series used to construct the models within the considered clusters.

5.2.2.4. Model Checking Against Representative Time-Series

First, we must check the performance of the obtained GEV-distributions ($\Psi^*_{GEV}(H^*_{m0})$). To this end, QQ-plots are used to assess $\Psi^*_{GEV}(H^*_{m0})$ against the representative time-series of H^*_{m0} (See Figure 5.16. Apart from the QQ-plots, the bias and root mean square error (RMSE) are calculated to assess the quality of the GEV distributions.

From Figure 5.16, it can be seen that for series aggregation, the $\Psi^*_{GEV}(H^*_{m0})$ models give a good fit for

all clusters together with the lowest values for bias and RMSE compared to the other representative time-series. This can be attributed to the high number of H^*_{m0} observations and low variability between H^*_{m0} observations used to fit μ^*, σ^* and ξ^*. Only cluster 20 (Figure 5.16.B) gives slightly higher values for bias and RMSE compared to clusters 12 and 59, which can be attributed to the slight deviation of the observations from $\Psi^*_{GEV}(H^*_{m0})$ for the highest quantiles.

Looking at parallel maximum aggregation, it can be seen that $\Psi^*_{GEV}(H^*_{m0})$ gives poor fits when compared to the representative time-series for H^*_{m0}. $\Psi^*_{GEV}(H^*_{m0})$ underestimates the values for H^*_{m0} in all three considered clusters. It seems that selecting the maximum H^*_{m0} observation for each year results in a time-series of H^*_{m0} observations that cannot be adequately fitted to a GEV-model. Potentially, selecting the maximum value at every time step results in a subset consisting of outliers, thereby resulting in an unrealistic set of observations. Selecting the median values ensures that all aggregated observations "fall within the majority of the data," thereby generating a representative time-series for the intra-cluster H^*_{m0} observations. When studying the values for bias and RMSE for $\Psi^*_{GEV}(H^*_{m0})$ for parallel median aggregation, it can already be seen that the fit is already much better than for parallel maximum aggregation.

Shifting attention to the clustering centroid, it can be observed that all GEV models have low values for the bias ($< 10\ cm$) and RMSE. The $\Psi^*_{GEV}(H^*_{m0})$ models for cluster 59 (Figure 5.16.L) has a relatively high value for RMSE compared to the other clusters. This is caused by the deviation of the highest H^*_{m0} observations from the $\Psi^*_{GEV}(H^*_{m0})$ identity line. Here, $\Psi^*_{GEV}(H^*_{m0})$ underestimates the magnitude of the most extreme events, which were found correspond to the 1953 and 1990 extreme H_{m0} events. Finally looking at the models representing the grid-points that yield the highest return levels before detrending, $\Psi^*_{GEV}(H^*_{m0})$ gives a good fit with low values for bias ($< 10\ cm$) and low values for RMSE for the three clusters. There are no big deviations of H^*_{m0} from the $\Psi^*_{GEV}(H^*_{m0})$ identity line to be considered for all clusters.

5.2.2.5. Model Checking Against Intra-Cluster Grid-Points

Apart from verifying $\Psi^*_{GEV}(H^*_{m0})$ by comparing the models with the representative time-series for H^*_{m0}, the performance of $\Psi^*_{GEV}(H^*_{m0})$ must be evaluated by comparing them against all intra-cluster H^*_{m0} observations. This is done to study how well $\Psi^*_{GEV}(H^*_{m0})$ represents the H^*_{m0} observations at the grid-point level. Figure 5.17 shows the QQ-plots for $\Psi^*_{GEV}(H^*_{m0})$ and all intra-cluster H^*_{m0} observations.

First observations of Figure 5.17 show that the scatters representing H^*_{m0} observations at grid-point level are tightly packed together. This tight packing is caused by detrending the AM H_{m0} observations. Figure 5.13 shows that the values for b_0 have a range in the order of $1\ m$. Detrending causes the intra-cluster variability of H^*_{m0} to be reduced because detrending depends on the value of b_0 (see Equation 3.50). Furthermore, it can be observed that $\Psi^*_{GEV}(H^*_{m0})$ for Cluster 59 generally show higher values for the RMSE. This can be explained by the fact that cluster 59 contains more grid-points (156) compared to clusters 12 and 20 (50 and 36 respectively), causing the variability of intra-cluster H^*_{m0} to be higher.

The GEV models corresponding to series aggregation results in the lowest values for bias and RMSE when compared to the other aggregation methods. This can be explained by the fact that series aggregation considers all H^*_{m0} observations in the cluster, thereby maximizing the available information about H^*_{m0} in the cluster.

Parallel maximum aggregation gives the highest overestimation as indicated by the bias (in the order of $20\ cm$) and RMSE values given in Figure 5.17. This is caused by the fact that parallel maximum aggregation selects the maximum values for H^*_{m0} for every year, thereby constructing a set of the most extreme H^*_{m0} observations in the cluster. Moreover, this results in a set of H^*_{m0} values that potentially contains outliers for the cluster observations when comparing the individual grid-points, further enhancing the overestimation of the extreme wave data. Parallel median aggregation gives a slight underestimation based on the bias when comparing the three clusters. The bias for parallel median aggregation is already less than for parallel maximum aggregation. This is explained by the fact that parallel median aggregation selects the median H^*_{m0} value for every year, thereby always selecting a value that falls "within the majority" of the H^*_{m0} data.

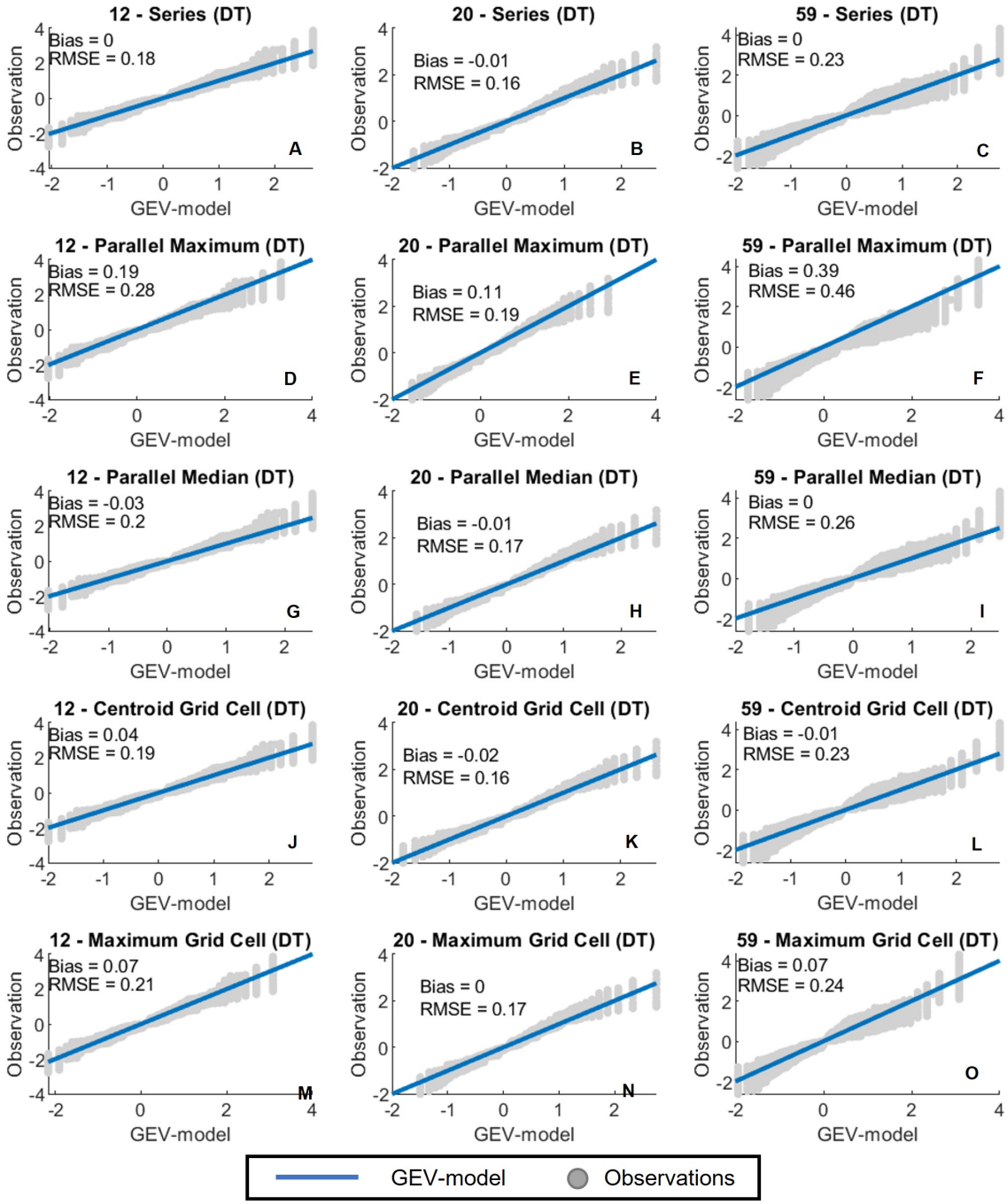


Figure 5.17: QQ-plots comparing $\Psi^*_{GEV}(H^*_{m0})$ models against all intra-cluster H^*_{m0} observations for the different representative time-series and considered clusters.

The QQ-plots for $\Psi^*_{GEV}(H^*_{m0})$ corresponding to the H^*_{m0} observations at the clustering centroid gives very low values for bias ($< 0.05\ m$). Finally, looking at $\Psi^*_{GEV}(H^*_{m0})$ for the grid-point that gives the highest return levels H_{m0} before detrending, does not give a significant overestimation for the detrended data based on the value for bias. This can be explained by the detrending process, which reduces the intra-cluster variability for H^*_{m0} based on the relative magnitude for b_0.

5.2.2.6. Performance for Extreme Return Levels

Not only is the performance of $\Psi^*_{GEV}(H^*_{m0})$ evaluated by comparing them with H^*_{m0} observations. Also, the extreme return levels of H^*_{m0} exceeding the available H^*_{m0} observations are studied. Return level

plots are constructed that compare the return levels of $\Psi^*_{GEV}(H^*_{m0})$ against the return level plots of the GEV distributions at grid-point level. The GEV distributions at grid-point level have been computed by fitting the GEV-parameters to the H^*_{m0} observations using Maximum Likelihood Estimation (MLE). The return level plots are presented in Figure 5.18

Interestingly, Figure 5.18 shows that for cluster 20, not only the upper confidence bounds of $\Psi^*_{GEV}(H^*_{m0})$ are unbounded, but also the GEV distributions at several intra-cluster grid-points. This is verified by the fact that for several grid-points, the values for ξ_{gp} exceed zero (see Figure 5.14). As explained, cluster 20 is located of the Northeast Scottish coast where higher water depths prevail than in clusters 12 and 59 (see Figure 4.2). The fact that several grid-points show unbounded GEV distributions further supports the claim that the AM H_{m0} events in cluster 20 are not always depth limited, meaning that some grid-points do not show an upper bound for the GEV-model.

Figure 5.18, shows that for series aggregation, the return level curves of $\Psi^*_{GEV}(H^*_{m0})$ for high return periods are roughly in the middle compared to the individual grid-points for all three clusters. Furthermore, it can be observed that the 90% confidence bounds of $\Psi^*_{GEV}(H^*_{m0})$ for series-aggregation are narrower than for other representative time-series. This is the result of the high number of observations and the relatively small variety between the intra-cluster H^*_{m0} values used to compute $\Psi^*_{GEV}(H^*_{m0})$. This narrow confidence interval means that the GEV model for series aggregation fails to capture the most extreme intra-cluster return values for H^*_{m0}. This may potentially be problematic when design values for H_{m0} in design and risk assessment of hydraulic infrastructure are based on the confidence interval yielded by the representative GEV-models. This is especially troublesome when hydraulic infrastructure is located in regions where return values for H_{m0} is more extreme than what the confidence interval of $\Psi_{GEV}(H_{m0})$ projects.

Figures 5.18.D, E and F shows the $\Psi^*_{GEV}(H^*_{m0})$ models derived using parallel maximum time-series. It can be seen that for return periods up to approximately 500 years, $\Psi^*_{GEV}(H^*_{m0})$ yields the highest return values for cluster 12 and 59. For cluster 20 however, $\Psi^*_{GEV}(H^*_{m0})$ does not yield the highest return levels for H^*_{m0} despite the fact that parallel maximum aggregation was applied. Some individual grid-points yield higher return levels due to the positive values for ξ_{gp}, resulting in unbound GEV models. Furthermore, it can be observed that the GEV model for parallel aggregation has broader CI bounds compared to series aggregation. Figure 5.15 shows that for parallel maximum aggregation in cluster 20, ξ_{95} exceeds zero, meaning that the upper confidence bound of $\Psi^*_{GEV}(H^*_{m0})$ even captures the return level curves for the grid-points yielding unbound GEV distributions. This means that $\Psi^*_{GEV}(H^*_{m0})$ in cluster 20 still provides information about the most extreme return values of H^*_{m0}, preventing underestimation. For clusters 12 and 59, the 90% confidence bound lies well above the RL-curves of the individual grid-points. $\Psi^*_{GEV}(H^*_{m0})$ models. This potentially results in overestimation of the design values for H_{m0}, resulting in overly conservative designs. As a result, the costs of constructing hydraulic infrastructure may be too high, resulting in projects that are not economically feasible. The implications of this overly conservative designs are further elaborated in Section 5.2.3.

The $\Psi^*_{GEV}(H^*_{m0})$ models for parallel median aggregation show relatively low return values compared to the individual grid-points. Moreover, the upper bound of the 90% confidence interval does not capture the most extreme return levels for H^*_{m0} at grid point level. This despite $\Psi^*_{GEV}(H^*_{m0})$ for parallel median aggregation having broader confidence interval compared to series aggregation. During the design or risk assessment of hydraulic infrastructure in some areas within the clusters, the extreme H_{m0} conditions may be underestimated leading to unsafe designs. The return level plots for $\Psi^*_{GEV}(H^*_{m0})$ derived using the clustering centroid show that the best-fit estimate of $\Psi^*_{GEV}(H^*_{m0})$ follows a middle trajectory compared to the individual intra-cluster grid-points. Also, the upper confidence bounds capture the most extreme return levels of H^*_{m0} within the cluster, meaning that the GEV-model includes information about the most extreme H^*_{m0} within the cluster. Some care must be taken in cluster 20, as the RL-plot for one grid point exceeds the upper confidence bound of $\Psi^*_{GEV}(H^*_{m0})$.

Finally, the return level curves for $\Psi^*_{GEV}(H^*_{m0})$ corresponding to the grid-point with the highest return level (before detrending) are analyzed. The RL-values for $\Psi^*_{GEV}(H^*_{m0})$ are lower than for several intra-cluster grid-points. As mentioned in Section 5.2.2.5, this can be attributed to the detrending process,

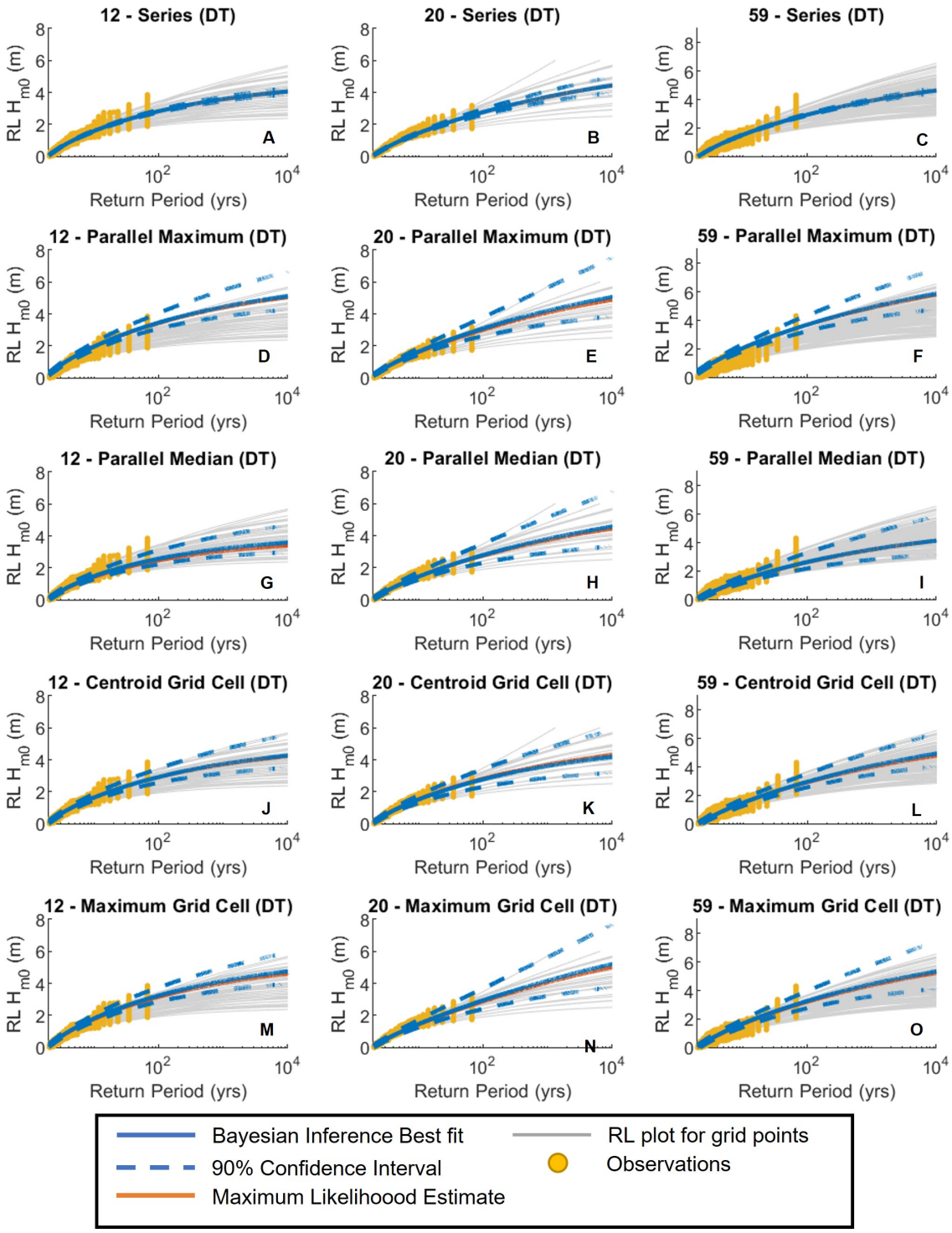


Figure 5.18: Return Level Plots comparing the return level plots of $\Psi^*_{GEV}(H^*_{m0})$ at cluster level to the GEV distributions at grid-point level.

where the intra-cluster variability between AM H_{m0} observations is reduced. Nevertheless, the upper confidence bound of $\Psi^*_{GEV}(H^*_{m0})$ captures the most extreme return levels for the intra-cluster grid-points very well. This means that $\Psi^*_{GEV}(H^*_{m0})$ provides information about the most extreme $H_{[m0^*}$ return levels that can be expected in the cluster.

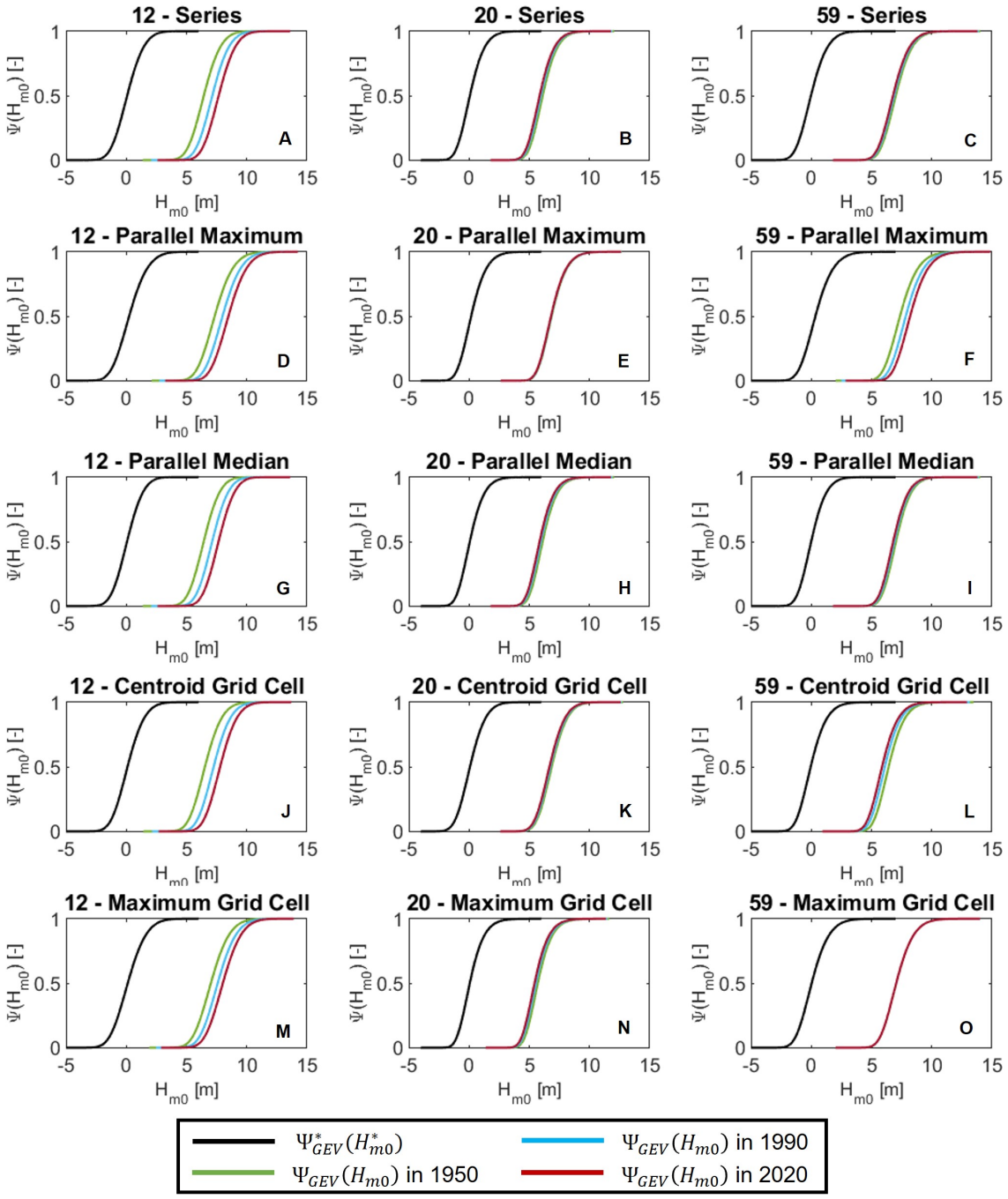


Figure 5.19: Cumulative Density Functions (CDF) showing the Detrended GEV models ($\Psi_{GEV}(H^*_{m0})^*$) and the models with the reinstated temporal trends ($\Psi_{GEV}(H_{m0}, t)$) for different years during the observation period.

5.2.2.7. Reinstating the trend in the GEV-models

After the $\Psi^*_{GEV}(H^*_{m0})$ models have been computed for H^*_{m0}, the temporal trend information of the extreme wave climate is reinstated into the extreme value distributions, following equation 3.55, yielding $\Psi_{GEV}(H_{m0} \mid t)$. This section studies the effect of accounting for the non-stationarity of AM H_{m0} in the representative $\Psi_{GEV}(H_{m0} \mid t)$ models by evaluating their performance in the different clusters. Accounting for the temporal trend means that $\Psi_{GEV}(H_{m0} \mid t)$ yields the following extreme value model:

$$\Psi_{GEV}(H_{m0} \mid t) \sim GEV\left(\mu^* + (b \cdot t + b_0), \sigma^*, \xi^*\right) \tag{5.3}$$

Equation 5.3 shows that the cluster representative extreme value model no longer is a function of

H_{m0} only, but also a function of time (t in years). The main assumption behind this model is that the non-stationarity of H_{m0} between 1950 and 2020 can be described by a linear trend. The cumulative distribution functions (CDFs) of $\Psi_{GEV}(H_{m0} \mid t)$ relevant to any year between 1950 and 2020 can be derived by substituting the corresponding value for t into Equation 5.3. For this, the selected value t must be between 0 (yielding the CDF corresponding to 1950) and 70 (yielding the CDF corresponding to 2020). Figure 5.19 shows the CDF plots for $\Psi_{GEV}(H_{m0} \mid t)$ per cluster. The green, blue and red lines in Figure 5.19 represent the CDFs corresponding to 1950, 1990 and 2020 respectively. The shift in CDFs indicates the non-stationary imposed into $\Psi_{GEV}(H_{m0} \mid t)$ by adding b and b_0 to the location parameter (μ^*) in Equation 5.3.

Figure 5.19 shows that the temporal shift of the CDFs depends on the presence of a temporal trend in the cluster. Section 5.2.2.1 showed that all grid-points in cluster 12 have a significant upward trend for the extreme wave climate. This subsequently means that over time, the CDFs of $\Psi_{GEV}(H_{m0} \mid t)$ for cluster 12 shift towards the right, representing an increase in extreme H_{m0} over time. In cluster 20, almost no significant trends were detected by the MK-test (see Fiugre 5.11. However, Section 5.2.2.1 showed that the non-stationarity in cluster 20 can be characterized by very small negative values for b. As a result, the CDFs of $\Psi_{GEV}(H_{m0} \mid t)$ for cluster 20 in Figure 5.19, show either almost no temporal shift or a small shift towards the left, indicating a decrease of extreme H_{m0} over time.

Shifting attention to Cluster 59, the CDFs of $\Psi_{GEV}(H_{m0} \mid t)$ in Figure 5.19 show different horizontal shifts over time. In Section 5.2.1, it was discovered that cluster 59 is characterized by a significant upward trend for AM H_{m0} in the Eastern parts and a downward trend in the Western parts of the cluster. The presence of heterogeneous non-stationarity affects $\Psi_{GEV}(H_{m0} \mid t)$ when accounting for the non-stationarity. For example, the CDFs of $\Psi_{GEV}(H_{m0} \mid t)$ for series aggregation, parallel median aggregation and the grid-point yielding the highest return level show almost no shift over time (see Figure 5.19). However, the CDF of $\Psi_{GEV}(H_{m0} \mid t)$ for parallel maximum aggregation (where the highest values for b and b_0 are substituted into $\Psi_{GEV}(H_{m0} \mid t)$) show a shift towards the right over time, indicating an upward trend in extreme H_{m0}. Contrarily, the CDF of $\Psi_{GEV}(H_{m0} \mid t)$ corresponding to the clustering centroid indicated a downward trend in extreme H_{m0}. After investigation, it was found that the grid-point yielding the clustering centroid can be found in the Western parts of cluster 59, where negative values for b can be observed.

The selection of the appropriate values of b and b_0 to be substituted into $\Psi_{GEV}(H_{m0} \mid t)$ naturally has implications on the return level curves of $\Psi_{GEV}(H_{m0} \mid t)$. Especially with respect to the return level plots at grid-point level. Figure 5.20 shows the return level plots of $\Psi_{GEV}(H_{m0} \mid t)$ corresponding to 2020 ($t = 70$). The grey lines show return level plots corresponding to GEV models derived at grid-point level. The GEV models derived at grid-point level in accordance with 3.55. The substituted values for b and b_0 are the values observed at the grid-points.

For $\Psi_{GEV}(H_{m0} \mid t)$ corresponding to series aggregation, it can be seen that similar results are obtained as in Figure 5.18. Non-stationarity was accounted for by substituting the median values for b and b_0 found in the cluster into $\Psi_{GEV}(H_{m0} \mid t)$. As a result, $\Psi_{GEV}(H_{m0} \mid t)$ shows median return level curves compared to the intra cluster grid-points. The Confidence interval for $\Psi_{GEV}(H_{m0} \mid t)$ remains too narrow, thereby causing $\Psi_{GEV}(H_{m0} \mid t)$ failing to capture the most extreme return levels for H_{m0} characteristic to the cluster. Similar results can be observed for parallel median aggregation. Despite the wider confidence bounds for $\Psi_{GEV}(H_{m0} \mid t)$, they do not capture the most extreme return level plots for H_{m0} for all individual grid-points.

Naturally, the $\Psi_{GEV}(H_{m0} \mid t)$ derived for parallel maximum aggregation yields the highest return levels for H_{m0}. This can be explained by the fact that the maximum values for b and b_0 observed in the cluster are substituted into $\Psi_{GEV}(H_{m0} \mid t)$. As a result, $\Psi_{GEV}(H_{m0} \mid t)$ for parallel maximum aggregation yields the highest return level curves compared to the return level curves at grid-point level, leading to potential overestimation of the extreme H_{m0} for in the respective clusters (see Figure 5.20). This overestimation is further reinforced by the fact that the upper confidence bounds of $\Psi_{GEV}(H_{m0} \mid t)$ far exceed the return level curves at grid point level.

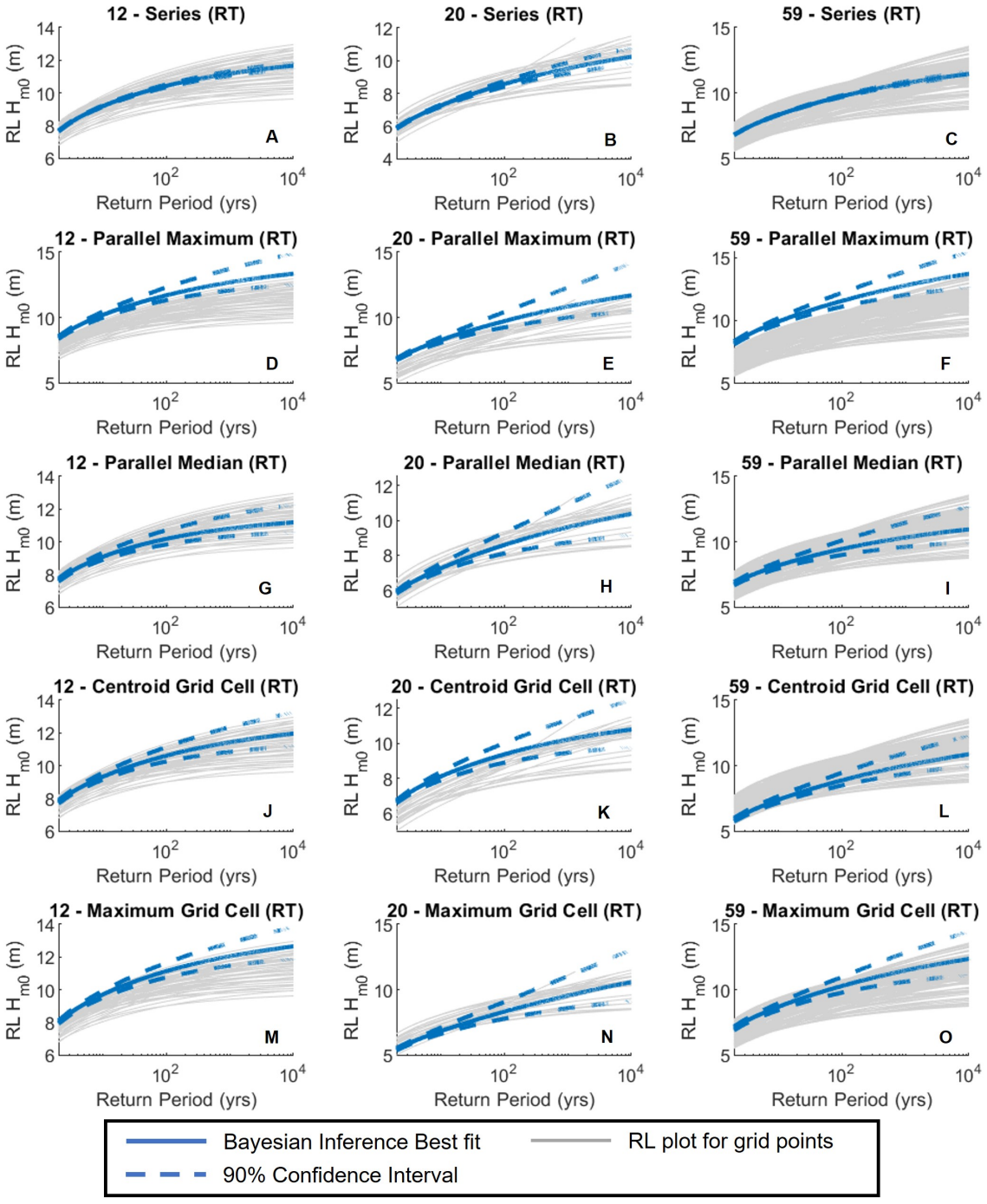


Figure 5.20: Return Level plots for the $\Psi_{GEV}(H_{m0} \mid t)$ models in the year 2020 ($t = 70$) compared against the return level plots at grid point level for the different representative time-series and considered clusters.

For $\Psi_{GEV}(H_{m0} \mid t)$ derived by grid-point selecting (methods 4 and 5), it can be observed that the presence and direction of a temporal trend affects the performance of $\Psi_{GEV}(H_{m0} \mid t)$. Two notable examples can be found in Figure 5.20.L and 5.20.N. The clustering centroid for cluster 59 is located in a region exhibiting negative values for b. When evaluating $\Psi_{GEV}(H_{m0} \mid t)$ in Figure 5.20.L, it can be observed that the negative value for b causes the $\Psi_{GEV}(H_{m0} \mid t)$ to yield low return levels for H_{m0} compared to the return level curves at grid-point level. Figure 5.20.N shows that, despite the MK-test not detecting significant trends in cluster 20, the CDFs actually show a shift towards the left (Figure 5.19.N), indicating a small decrease in extreme H_{m0}. This has the effect that, especially for lower return

periods (< 100 years), $\Psi_{GEV}(H_{m0} \mid t)$ underestimates extreme H_{m0} return levels (see Figure 5.20.N).

5.2.3. Conclusions Regarding Model Performance

The goal of this section is to evaluate the intra-cluster statistical properties of extreme H_{m0} and study how an extreme value model describing extreme H_{m0} statistics at cluster-level can be derived. A linear temporal component ($b \cdot t + b_0$) is added to the location parameter (μ^*) to account for potential non-stationarity of H_{m0}. Five methods to derive representative extreme value models are evaluated. The extreme value model at cluster level ($\Psi_{GEV}(H_{m0} \mid t)$) must be able to give a representative picture of the extreme H_{m0} statistics within the cluster. For this, the representative non-stationary extreme value model must satisfy the following requirements:

- The 90% confidence interval must capture the most extreme H_{m0} return levels at grid point level. Practitioners that apply $\Psi_{GEV}(H_{m0} \mid t)$ for hydraulic infrastructure design will most likely choose a design value for H_{m0} based on the CI of $\Psi_{GEV}(H_{m0} \mid t)$. Capturing the most extreme H_{m0} conditions prevents underestimation of the design conditions;
- The $\Psi_{GEV}(H_{m0} \mid t)$ must not severely overestimate the extreme H_{m0} conditions. Choosing a design value for H_{m0} based on the CI $\Psi_{GEV}(H_{m0} \mid t)$ that severely overestimates the intra-cluster wave conditions can result in overly conservative designs that potentially are not economically feasible.

Based on the results presented in Section 5.2.2, the following conclusions about the five methods to derive $\Psi_{GEV}(H_{m0} \mid t)$ can be made:

1. Series aggregation (method 1): the confidence bounds are too narrow and therefore the representative model fails to capture the most extreme H_{m0} return levels in the cluster, potentially leading to severe underestimation of extreme H_{m0} whilst designing infrastructure;
2. Parallel Maximum aggregation (method 2): The QQ-plots yields poor results when assessing the performance of $\Psi^*_{GEV}(H^*_{m0})$ against the aggregated H^*_{m0} observations. Moreover, the representative model results in severe overestimation of the extreme H_{m0} return levels. The overestimation is further reinforced by selecting the maximum values for b and b_0 to be substituted into the models;
3. Parallel median (method 3): Despite the wider confidence bounds compared to method 1, the 90% confidence bounds fail to capture the most extreme return levels for extreme H_{m0} for the considered clusters. This potentially lead to underestimation of extreme H_{m0} design values;
4. Clustering centroid (method 4): For the clusters where either a monotonic upward trend or no significant trend was detected (i.e. clusters 12 and 20), the representative model shows promising performance. The derived representative models fit the extreme H^*_{m0} observations well and the confidence bounds of the RL-plots capture the most extreme return levels for H_{m0} without resulting in severe overestimation. However, for cluster 59 where mixed non-stationarity is present, the performance of the representative model is compromised;
5. Maximum return levels (method 5): As a result of detrending the AM H_{m0} observations before Bayesian Inference (BI), the representative models do not necessarily yield the highest return levels anymore. This makes the performance of the representative model dependent on the magnitude of b and b_0 at the grid-point in question.

From the above considerations, it is concluded that using the extreme H_{m0} at the clustering grid-point to construct $\Psi_{GEV}(H_{m0} \mid t)$ yields the most promising results. However, there are some considerations to be taken into account. The presence of mixed non-stationarity in clusters results in ambiguity. The temporal trend information (b and b_0) is substituted into the location parameter to account for potential non-stationarity of H_{m0}. Different values for b mean that extreme H_{m0} statistics exhibits different characteristics over time. As a result, $\Psi_{GEV}(H_{m0} \mid t)$ fails to give a representative picture of the extreme H_{m0} statistics in clusters with mixed non-stationarity. Given the current clustering approach and the resulting clustering output, it is advised to not apply non-stationary GEV models in clusters that are characterized by mixed non-stationarity. Conventional GEV models at grid-point level are advised instead. This raises questions regarding the current clustering approach and whether the clustering analysis should account for non-stationarity of extreme H_{m0}. The implications of excluding non-stationarity from the clustering analysis and whether it should be accounted for, is further evaluated in the Discussion.

5.3. Summary

This chapter presents the findings of a Hierarchical Agglomerative Clustering (HAC) analysis of the North Sea domain where grid-points are clustered based on comparable extreme wave characteristics. Furthermore, extreme value analysis at cluster level is performed including intra-cluster temporal trend analysis and the derivation of representative extreme value models describing the significant wave height (H_{m0}) on a cluster level.

The HAC input values include the 99.5th percentile of H_{m0} (denoted by $H_{m0,P99.5}$), the corresponding peak period (T_P) and the most frequent extreme wave directions (θ_1 and potentially θ_2). More information about the derivation of the HAC input is presented in Section 4.4.2. The HAC analysis is performed for two different parameter combinations. Both parameter combinations include θ_1 and θ_2. Combination 1 only includes $H_{m0,P99.5}$, while combination 2 also includes T_P. Five weight configurations are evaluated for each parameter combination. For each weight configuration, the linkage mechanisms are assessed based on the cophenetic correlation (C). The number of clusters ($\mathcal{K}$) that yields the clustering solution for each configuration is selected based on the Silhouette Coefficient (SC) Calinski-Harabasz (CH), Davies-Bouldin ratio (DB) and a careful analysis of how well the clustering output represents the boundaries between the input values. The best clustering solution is provided using $H_{m0,P99.5}$, T_P, θ_1 and θ_2 as input parameters with equal weights and Ward linkage. The value for $\mathcal{K}$ is equal to 50.

A critical assessment of the initial clustering output is required. This includes splitting clusters that contain grid-points in different geo-graphical locations. The HAC algorithm determines cluster boundaries based on the input values for H_{m0}, T_P, θ_1 and θ_2. However, this does not necessarily mean that grid-points in different geographical locations show exactly similar extreme wave characteristics if they are assigned to the same cluster. Also, neighboring clusters are merged when there is sufficient statistical evidence to claim that the clusters share similar extreme wave characteristics. In the end, the North Sea is partitioned into 63 clusters.

The extreme value analysis at cluster level and subsequent temporal trend analysis showed that accounting for the dominant wave direction is important when studying the presence of statistically significant temporal trends. Especially in cluster 59, the number of grid-points exhibiting statistically significant trends increased compared to the general trend analysis in Section 4.2.2. Furthermore, a case-study indicated that even different temporal trends (both in sign and magnitude) may be present within the same cluster depending on the considered dominant wave directions. These are important considerations to account for in the design and risk assessment of hydraulic infrastructure.

For three clusters, representative extreme value models are derived that describe H_{m0} at cluster level. Five different techniques to construct representative time-series of detrended AM H_{m0} observations (denoted by H^*_{m0}) are evaluated to assess their ability to describe extreme H_{m0} at cluster level. Bayesian Inference is employed to derive the GEV parameters describing H^*_{m0}. This yields the following GEV model:

$$\Psi^*_{GEV}(H^*_{m0}) \sim GEV\left(\mu^*, \sigma^*, \xi^*\right) \tag{5.4}$$

Potential non-stationarity is accounted for by adding the TS parameters (b and b_0) to μ^* so that the location parameter becomes a function of time. The extreme value model describing extreme H_{m0} at cluster level yields:

$$\Psi_{GEV}(H_{m0} \mid t) \sim GEV\left(\mu^* + (b \cdot t + b_0), \sigma^*, \xi^*\right) \tag{5.5}$$

Comparing the 5 methods to extreme value models at cluster level, it is concluded that using H_{m0} data from the clustering centroid gives the most promising. The confidence bounds capture the most extreme H_{m0} return levels for the cluster while also not severely overestimating the intra-cluster extreme H_{m0} conditions. However, there are some considerations to be taken into account. Clusters exhibiting different and/or opposing trends are not suitable for deriving representative extreme value models. This raises questions regarding the applied clustering approach and whether the clustering analysis should account for non-stationarity for extreme H_{m0}.

6

Application in Light of Climate Change

This research aims to apply clustering analysis and subsequent extreme value analysis at the cluster level in an attempt to update methodology in which extreme wave conditions are derived in the risk assessment of hydraulic infrastructure. This section studies whether this approach forms a practical alternative to conventional stationary analysis at grid-point level. Section 6.1 presents a practical example where the extreme value models at cluster level are employed to derive typical H_{m0} return values for hydraulic infrastructure. Section 6.2 places the applicability of Non-Stationary Extreme Value Analysis (NEVA) for risk assessment of Hydraulic Infrastructure in broader perspective in light of climate change.

6.1. Extreme Value Models at Cluster Level in Practice

To illustrate the performance of the extreme value models at cluster level for the clustering centroid, an example is presented where we study the temporal evolution of extreme return levels of the significant wave height (H_{m0}) in clusters 12 and 20. Cluster 59 was not considered because the extreme value model does not perform well as a result of heterogeneous temporal trends that characterize cluster 59. The analysis focuses on typical H_{m0} return periods for hydraulic infrastructure. This includes a return period of 50-years representative for offshore structures, such as wind turbines and boring platforms (Jiang et al., 2017), and 1000-year return periods for coastal infrastructure, such as flood defenses (Vergouwe, 2014). This example considers a hypothetical lifespan of 30 years, spanning between 2020 and 2050. This means that the H_{m0} design values in the years 2020 and 2050 will be evaluated.

6.1.1. Effective Return Level Plots

This section explains the derivation of the effective return level (ERL) plots used to analyze the temporal behavior of the H_{m0} return levels in this chapter (Katz et al., 2002). Figure 6.1 depicts the temporal evolution of the 50 and 1000-year return levels of H_{0m0} in clusters 12 and 20 for four distinct example models between 1950 and 2070. The solid lines represent the effective return levels corresponding to the medians of the posterior distributions. The dashed lines indicate the 90% confidence interval constructed using the 5th and 95th percentiles of the posterior distributions. The vertical dashed line indicates the end of the observation period from which the extreme H_{m0} observations were selected to infer the extreme value models. Table 6.1 shows the H_{m0} return levels for the return periods of interest including the boundaries of the confidence intervals.

The GEV parameters for the representative models are inferred using Bayesian Inference so that the GEV parameters read $\theta = \mu^*, \sigma^*, \xi^*$. Subsequently, the temporal trend information (b and b_0) is applied to account for non-stationarity making the location parameter a linear function of time: $\mu = \mu^* + b_0 + b \cdot t$. The non-stationary model is depicted by the red line in Figure 6.1. Secondly, there is the updated stationary model (UST), adopted from Luke et al. (2017). Here the non-stationary GEV parameters at the end of the observation period are used to project H_{m0} return levels for the entire life time of the structure under the assumption that H_{m0} can be described by a stationary process after the observation period has ended. The GEV parameters of the AST model can be found by substituting $t = 70$ in

Equation 3.55. Third is the so-called approximated stationary model. Here, the non-stationary model parameters corresponding to the middle of the observation period are used to approximate H_{m0} return levels over the entire temporal horizon under the assumption of stationary conditions. For this model, the year 1985 is considered, meaning that the value for t is equal to 35 in equation 3.55.

To assess the applicability of non-stationary models at cluster level, they are compared to a stationary GEV model at cluster level. The method to derive the stationary model is similar to how the non-stationary models at cluster level are derived and is adopted from Antonini et al. (2019). The Annual Maxima (AM) H_{m0} observations from the clustering centroid are used to infer the GEV parameters using Bayesian Inference. Furthermore, normal informative priors are constructed using the information about the GEV parameters at grid-point level. The only difference between the non-stationary model derived in Section 3.5 and this stationary model is that for the stationary model, the AM H_{m0} observations are not detrended prior to Bayesian Inference.

6.1.2. Quantitative evaluation

First looking at the results for the 50-year return period in clusters 12 and 20, it can be observed that for all 4 extreme value models, the H_{m0} return levels are higher for cluster 12 than they are for cluster 20. This can be explained by the dominant wave direction of the clusters, which means that cluster 12 has longer fetches for wave growth than cluster 20. The differences in return levels between the non-stationary and stationary models are evident in Cluster 12. The differences are the result of non-stationarity in H_{m0} that was discovered in cluster 12. The best fit line of the non-stationary model results in H_{m0} return levels that are approximately 0.7 m higher than for the stationary model in 2020 (10.31 and 9.65 meters respectively). This difference increases to 1.2 meter in 2050 (10.85 and 9.65 meters for the non-stationary and stationary model, respectively). These are significant differences that must be accounted for in the design and risk assessment of offshore structures. Not accounting for this non-stationary can thus result in significant underestimations of the design conditions, potentially resulting in unsafe structures.

Looking at the confidence bounds of the non-stationary and stationary model, it can be seen that the non-stationary model's lower confidence bound exceeds the stationary model's upper confidence

Table 6.1: Design values for H_{m0} derived from the effective return level plots for 50 and 1000-year return periods (see Figure 6.1)

	Cluster 12 - 50 yr			Cluster 12 - 1000 yr		
	P_{05}	P_{50}	P_{95}	P_{05}	P_{50}	P_{95}
Conventional	9.33	9.65	10.15	10.05	10.62	11.71
Non-stationary (2020)	9.98	10.31	10.76	10.88	11.45	12.4
Non-stationary (2050)	10.53	10.85	11.3	11.42	11.99	12.94
Updated Stationary	9.98	10.31	10.76	10.88	11.45	12.4
Approx. Stationary	9.34	9.66	10.11	10.23	10.8	11.75

	Cluster 20 - 50 yr			Cluster 20 - 1000 yr		
	P_{05}	P_{50}	P_{95}	P_{05}	P_{50}	P_{95}
Conventional	8.69	9.14	9.79	9.61	10.62	12.45
Non-stationary (2020)	8.62	9.01	9.58	9.39	10.13	11.36
Non-stationary (2050)	8.54	8.95	9.49	9.31	10.05	11.28
Updated Stationary	8.62	9.01	9.58	9.39	10.13	11.36
Approx. Stationary	8.71	9.11	9.67	9.49	10.23	11.46

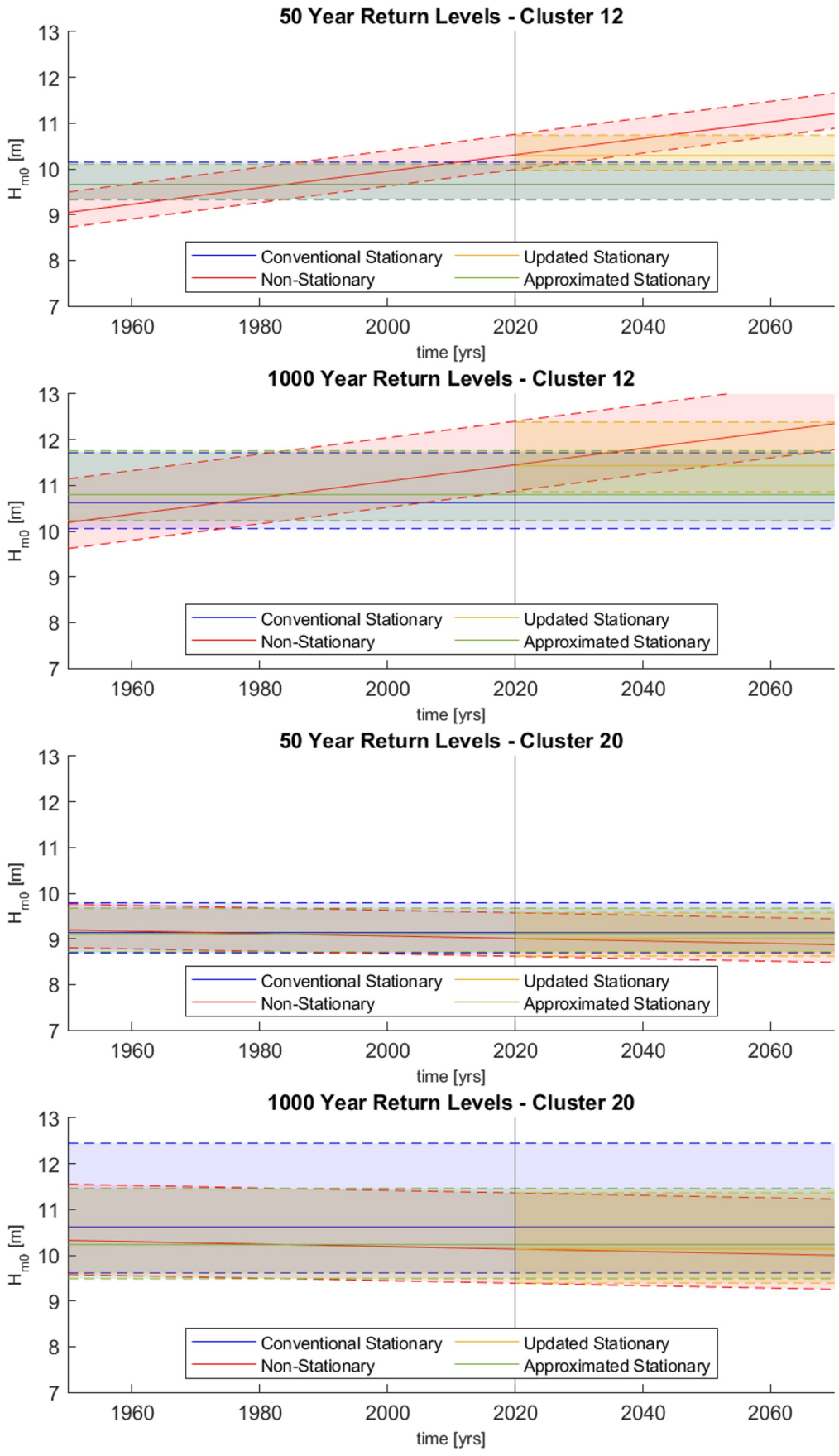


Figure 6.1: Effective return level plots for 50 and 1000-year return levels in Clusters 12 and 20. Solid lines represent the effective return levels corresponding to the medians of the posterior distributions. Dashed lines and shaded area represent the 90% confidence interval constructed using the 5th and 95th percentiles of the posterior distributions.

bound. This indicates that H_{m0} return levels potentially differ significantly between the stationary and non-stationary models in the presence of significant temporal trends.

Cluster 20 was not characterized by statistically significant temporal trends. Despite that, minor differences in the effective return levels between non-stationary and conventional stationary models can be observed in 2020 (9.01 and 9.14 meters) and 2050 (8.95 and 9.01 meters) respectively. Because the values for b in cluster 20 are slightly negative (see Figure 6.1), the ERL-plots for the non-stationary yield slightly lower values than for the stationary model. In this scenario, the non-stationary model to infer design values provides no real benefits compared to stationary modelling and potentially even leads to unsafe designs. Moreover, the confidence bounds of both models share significant overlap, indicating that design values for H_{m0} used to design hydraulic infrastructure are not so different between stationary and non-stationary models in clusters where no temporal trends are present.

Looking at the plots showing the 1000-year return levels in clusters 12 and 20, the H_{m0} ERL are higher than for 50-year return periods. For cluster 12, selecting a stationary model yields best estimate effective return levels in the order of 10.62 meters, whilst the non-stationary model yields 11.45 meters in 2020 and 11.99 meters in 2050. For cluster 20, the H_{m0} RL for the stationary models are 10.62 meters and for the non-stationary model are 10.13 meters in 2020, and 10.05 meters in 2050. Furthermore, the confidence interval (CI) for the 1000-year return levels are wider than for 50-year return levels. For the 50-year return levels, the width of the CI in both clusters is in the order of 0.8-0.9 meters. For 1000-year return levels, the CIs vary between 1.6 and 2.8 meters for the different clusters and considered models. This broader CI is caused by the fact that limited data is available to provide accurate projections surrounding return periods that exceed the length of the observation period. The broader CI result in more uncertainty regarding the choice of design values for H_{m0} when designing infrastructure.

The broader CIs for the 1000-year return levels in cluster 12 mean that for a large part of the period between 2020 and 2050, the CIs of the non-stationary and stationary model overlap. This raises questions regarding the benefit of non-stationary models over stationary models. However, towards the end of the lifetime in 2050, the CI of the non-stationary model almost completely exceeds the confidence interval of the stationary model, resulting in significantly different design conditions between the two models. This supports the statement that, in the case of a statistically significant temporal trend, non-stationary models may be preferable to stationary models for determining extreme wave design conditions.

For cluster 20, the CI of the stationary model is particularly wide, with the upper confidence bound far exceeding the best-fit line. This is caused by the fact that ξ of the upper confidence bound of the stationary model is equal to 0.09. This means that the upper confidence bound is unbounded. Furthermore, the confidence bounds for both the stationary model and the non-stationary model show overlap for the entire lifetime of the structure. This further supports the statement that non-stationary modelling provides no benefit over conventional stationary modelling in clusters where no statistically significant trend is present.

6.1.3. Approximated stationarity

In this research, non-stationary models are obtained by adding b and b_0 to μ^*. This makes the extreme value model describing H_{m0} at cluster level time-dependent. An attempt is made to approximate stationary conditions by considering the GEV parameters corresponding to the median of the observation period (i.e. 1985, $t = 35$). Figure 6.1 and Table 6.1 show that for the 50-year return periods, the H_{m0} effective return levels for the stationary model and the approximated stationary (AST) model are very close, with differences in H_{m0} in the order of 0.01 meter. For the 1000-year effective return levels, the differences between the two models are bigger, with the AST model underestimating H_{m0} by 0.4 meters for cluster 20. Using the AST model in this case may lead to unsafe designs of hydraulic infrastructure. Conclusions about the applied methodology to approximate stationary conditions is debatable due to the limited number of clusters that were tested.

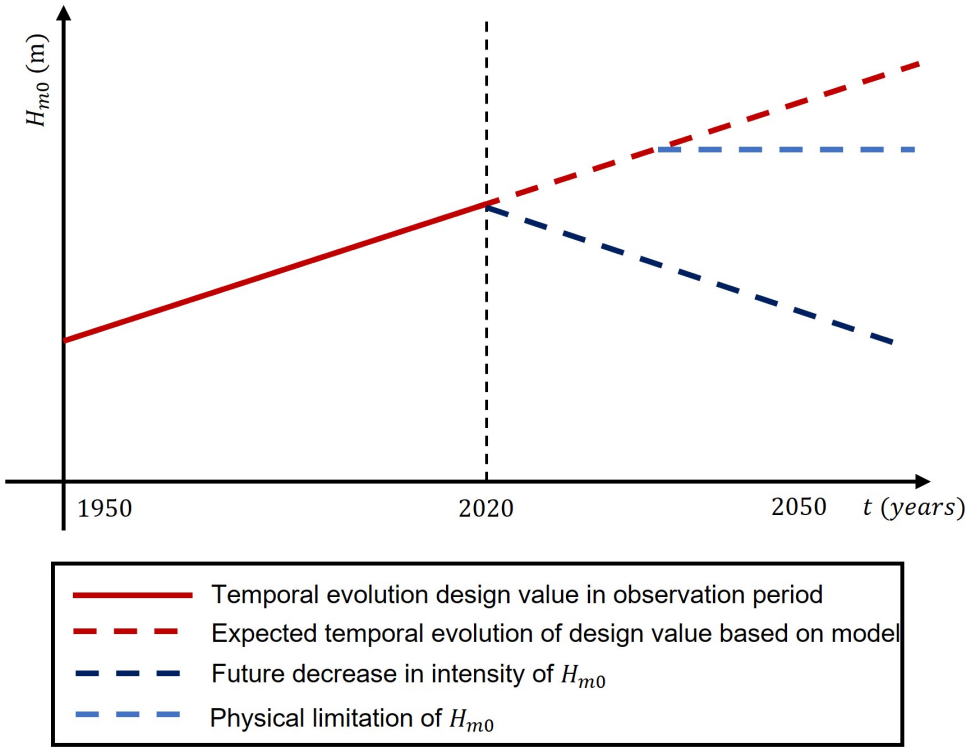


Figure 6.2: Illustration showing the uncertainty surrounding the applicability of non-stationary extreme value models at cluster level to derive design values of H_{m0} in light of climate change.

6.2. NEVA in light of climate change

In chapter 5, representative non-stationary extreme value models describing H_{m0} on the cluster level were derived. These models show promising performance when the clusters are characterized by a uni-directional temporal trend. This Section explores the applicability of non-stationary models in light of a changing climate. Section 6.2.1 describes the considerations surrounding the application of non-stationary models in light of climate change. Section 6.2.3 presents a flowchart for decision making surrounding non-stationary models at cluster level in the design and risk assessment of hydraulic infrastructure. Finally, Section 6.2.4 describes the potential benefit of using an alternative method to derive H_{m0} design conditions in the form of updated stationary models.

6.2.1. Considerations

In Section 5.2, it was explained that heterogeneous intra-cluster temporal trends (either in slope or direction) result in ambiguity surrounding the use of non-stationary extreme value models at cluster level. The temporal trend information (b and b_0) is applied to account for non-stationarity, thereby adding a non-stationary component to the GEV distributions. However, with heterogeneous values for b, the representative models potentially fails to capture the most extreme return levels at the grid-point level. It is advised not to apply non-stationary GEV models in clusters that are characterized by heterogeneous temporal trends. Conventional stationary GEV models at grid-point level are advised instead. See Section 5.2.3 for more information. However, even when a cluster is fully characterized by a uni-directional trend, the use of non-stationary models at cluster level in light of a changing climate remains debatable. Uncertainty surrounding climate change as well as potential physical limitations increase the uncertainty surrounding the applicability of non-stationary models in the design and risk assessment of hydraulic infrastructure.

6.2.1.1. Uncertainty Surrounding Temporal Trends

Accounting for upward trends in extreme H_{m0} is potentially very important in the design and risk assessment of hydraulic infrastructure. The strength requirements for the structure of interest will increase over time. Not accounting for an increase in extreme H_{m0} potentially leads to unsafe design. Section

6.1 shows that using the non-stationary representative models to derive design values accounting for upward trends leads to considerably higher H_{m0} design values than using stationary modelling.

Furthermore, Section 5.2.1 also showed that there are several locations in the North Sea that are characterized by a downward trend for extreme H_{m0}. The question is whether to account for downward trends when deriving design conditions for H_{m0}. Negative values for b result in GEV models that yield decreasing H_{m0} return levels over time. Accounting for a decreasing intensity of H_{m0} may potentially be beneficial in terms of cost-saving because the strength requirements of hydraulic structures will decrease over time. However, caution is required when using non-stationary extreme value models in regions with downward trends for H_{m0}.

The applicability of non-stationary extreme value modelling over conventional stationary modelling in light of climate change is debatable. Matalas (1997) explains that observed trends in the data might only be a small horizon of larger oscillations. As is discovered in Section 4.2.2, the detected temporal trends depend on the considered temporal horizon. The period between 1990 and 2020 primarily showed downward trends for H_{m0} in the research domain, whilst long term analyses (1950-2020) showed a combination of upward trends in the East and downward trends in the West of the North Sea. This potentially means that extreme H_{m0} in the North Sea shows oscillatory behavior instead of a monotonic trend. It is therefore possible that the detected decrease is followed by an increase in H_{m0} intensity. If this oscillatory signal is not properly accounted for, accounting for a downward trend may lead to severe underestimations of H_{m0} design conditions, leading to unsafe designs. Accounting for the downward trend in design practices may be possible if the physical driver behind H_{m0} is properly understood. This includes the non-stationary behavior of the extreme wind climate in the North Sea. If the underlying process causing the non-stationarity of H_{m0} is not well-understood, no clear projections can be made regarding the future non-stationarity of H_{m0}. In short, applying non-stationary extreme value models in regions characterized by a downward trend may potentially be beneficial if the physical process behind the downward trend is well-understood and it is expected that the decrease in extreme H_{m0} continues in the future. If the physical process in not well-understood or when a subsequent increase in H_{m0} is expected, accounting for downward trends in the extreme value model of the extreme wave climate is not advised, and conventional, stationary GEV models are advised instead.

Even when the cluster is characterized by an upward trend only, the success of the GEV models on the cluster-level in not guaranteed. Similar to what is explained above, detected upward trends may also be part of a larger cycle. Figure 6.2 shows potential effects of varying non-stationarity and that how decreasing temporal trends may follow increasing trends beyond the observation period. Accounting for upward trends of H_{m0} in the design and risk assessment of hydraulic infrastructure can therefore lead to over-estimations of the design conditions of H_{m0} if the non-stationarity behind the physical process driving the wave climate is not understood properly. Overestimations of H_{m0} have the consequence that the design and risk assessment of hydraulic infrastructure is potentially overly conservative, resulting in high costs. In worst case scenarios, it potentially means that engineering projects are no longer economically feasible. It is therefore advisable to, even if the region in the North Sea is characterized by an increase in H_{m0} intensity, to closely study the underlying causes of the non-stationarity for example by studying future projections, before making conclusion whether to use non-stationary extreme value models or not.

6.2.2. Future Projections of Wave Climate

Another limitation that hinders decision making surrounding design strategies is that this research does not consider future projections. Future projections provides a better understanding of the non-stationarity of the extreme wave climate. Thereby enabling better decision making surrounding the application of non-stationarity extreme value models in the design and risk assessment of hydraulic infrastructure. The implications of missing future projections on the applicability of non-stationary extreme value models to derive extreme wave conditions are further explained in the Discussion.

6.2.2.1. Physical Limitations

Another factor to account for when determining design conditions for H_{m0} is potential depth limitation. Even if the intensity of the extreme wind speeds increases in the future, the rather limited depth in

various regions of the North Sea may cause H_{m0} to reach a physical limit. Figure 6.2 shows that depth limitations cause design values for H_{m0} to become constant in time. Especially in the design and risk assessment of coastal infrastructure, the limited water depth mean that potentially H_{m0} reaches a physical limit well before the extreme waves reach the coastal structure. Not accounting for physical limitations in coastal waters could potentially result in overestimation of the H_{m0} design conditions. In offshore locations, higher water depths mean that H_{m0} can obtain higher values as a result of potential higher extreme winds in the future. In that case, accounting for physical limitations when deriving design values for H_{m0} is not as relevant as in shallower water depths.

6.2.3. Model Selection Flowchart

The choice for potential cluster specific N-GEV models can be summarized in the flowchart listed in Figure 6.3. In short, it is suggested to first perform a temporal trend analysis in the cluster following the approach mentioned in section 5.2.1. This gives an understanding of the presence of significant trends and corresponding magnitude and direction (b and b_0). If the cluster is characterized by multiple trends varying in direction and/or magnitude, it is advisable to use traditional extreme value models. Furthermore, if the trend can be characterized by downward trends and the underlying cause of the downward trend and no future projections can be made regarding the design variable H_{m0}, traditional extreme value models are preferred. Only when all intra-cluster grid points are characterized by similar trends in both direction and magnitude, cluster representative extreme value models might be preferred over conventional GEV-models. Even then, the applicability of non-stationary extreme value models is debatable. The non-stationarity of the design variable must be properly understood. In case future projections of the design variable are not available, it may be advisable to use updated stationary models describing H_{m0} on the cluster level instead.

The extreme value model on the cluster level must be able to give a representative description of the extreme wave climate within the cluster. For this, the representative (non-stationary) extreme value model must satisfy the following requirements:

- Confidence interval must capture the most extreme H_{m0} conditions within the cluster. It is expected that practitioners using extreme value models on a cluster level for the design and risk assessment of hydraulic infrastructure will base their choice for a design variable on the confidence interval of the representative extreme value model. Capturing most extreme H_{m0} conditions prevents underestimation of the design conditions;
- The representative extreme value model as must not severely overestimate the extreme H_{m0} conditions. Choosing a design value for H_{m0} based on a confidence interval that severely overestimates the intra-cluster wave climate potentially yields overly conservative designs that are not economically feasible.

6.2.4. Updated Stationary Models

In short, whether or not to use non-stationary models in light of climate change is very case dependent and a thorough understanding of the non-stationary of the design parameter as well as the underlying causes of the non-stationarity is required. With a lack of future projections, updated stationary models potentially provide a useful alternative (Luke et al., 2017). In updated stationary (UST) models, the non-stationary GEV parameters at the end of the observation period are used to project H_{m0} return levels under the assumption that H_{m0} can be described by a stationary process in the future. This is elaborated further in Section 3.5.4.1.

Luke et al. (2017) explored UST models whilst studying the applicability of non-stationary extreme value analysis (NEVA) for deriving extreme river discharges (See Section 2.2 for more information). Luke et al. (2017) recommended using UST models for the risk assessments in watersheds that have undergone physical changes (e.g. urbanization). Besides complex bottom dynamics, there are very few physical changes that affect the extreme wave climate in the North Sea. Nevertheless, UST models potentially provide a useful alternative in case the future temporal evolution of the design variable in light of climate change is uncertain.

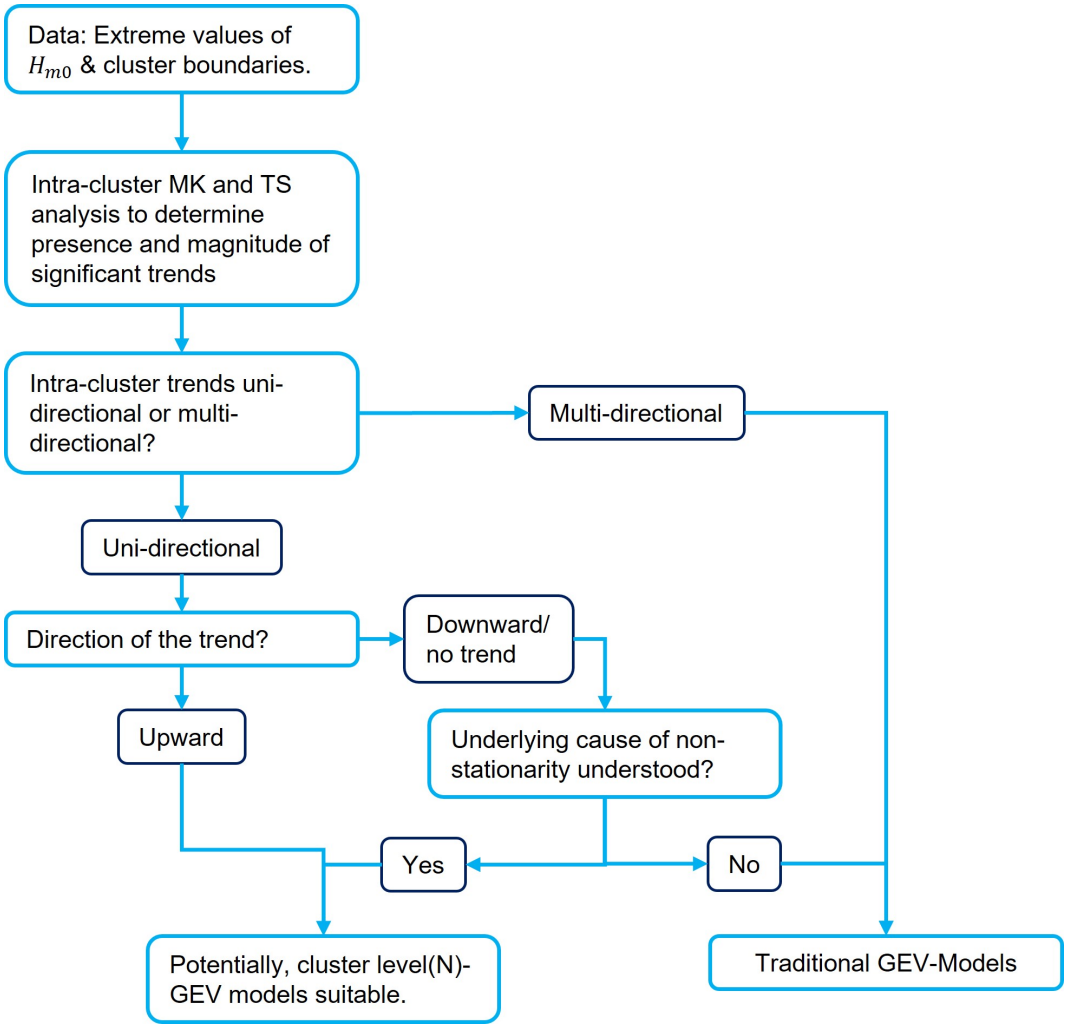


Figure 6.3: Flowchart showing whether applying a cluster characteristic (non-stationary) extreme value model is a feasible solution or not

UST models account for the detected non-stationarity in the sense that it applies the GEV parameters at the end of the observation period to project H_{m0} return levels for the future. However, because UST models do not extrapolate beyond the identified non-stationarity detected during the observation period, future over- and/or underestimations are unlikely if the detected non-stationarity does not continue. Figure 6.1 and Table 6.1 show the 50- and 1000-year return levels of H_{m0} in cluster 12 and 20 for the UST approach. It can be seen that for cluster 12, the UST models provides more conservative estimates than the stationary model. However, given the fact that a continued increase for H_{m0} return levels is uncertain, severe over-estimations are also prevented. When using UST models, it is advised to regularly check and update the model to see if H_{m0} return levels show significant changes in the future and update the design and risk assessment of the structure in question.

7

Discussion and Recommendations

The goal of this research can be summed up by the following points: (1) Study the temporal variations in extreme significant wave height in the North Sea domain, (2) Investigate how hierarchical clustering analysis can be employed to identify regions that exhibit similar extreme wave characteristics, (3) How the obtained clusters and temporal variability can be employed to derive extreme value models describing the extreme H_{m0} at cluster level and (4) assess the applicability of the derived non-stationary models at cluster level to derive design values for design and risk assessment of hydraulic infrastructure. In the discussion, we will evaluate the choices, assumptions and limitations related to this research and their effects on the results. Furthermore, suggestions how this research can be improved in the future will be provided.

7.1. Relevance to Engineering

This section reflects on the relevance of this research with respect to engineering purposes. This compares the obtained results with the motivation of this research. Mainly regarding the performed clustering analysis and extreme value analysis light of design and risk assessment of hydraulic infrastructure. Clustering analysis and subsequent extreme value analysis of the wave climate at cluster level provides several advantages. First of all, clustering analysis can help to reduce the data volume needed to study the extreme wave climate in larger domains, such as the North Sea. In this research we have established a cluster map consisting of 63 clusters, meaning that only 63 time-series representative of the extreme wave characteristics on the cluster level are required instead of the ~ 1200 original grid points to understand the extreme wave climate throughout the North Sea basin. This corresponds to a data volume reduction of 95% compared to an analysis were all grid points are studied individually.

Secondly, clustering analysis provides useful information regarding regions with similar extreme wave regimes throughout the North Sea domain. This can provide useful information regarding hydraulic boundary conditions in early design stages when more accurate data is not readily available. Furthermore, knowledge about extreme wave characteristics is useful when performing analyses on a large spatial scale. Information about wave conditions can potentially help to improve numerical wave models thereby providing more accurate results (Goharnejad et al., 2022). Improved numerical models may contribute to a better understanding of the North Sea's complex bottom dynamics. This in turn is relevant for the stability of bottom foundations of offshore structures including offshore wind turbines and boring platforms.

Non-stationary components were added to the extreme value models at cluster level to account for temporal changes in H_{m0} intensity. If the observed rising trend in extreme H_{m0} for cluster 12 continues, differences in design values for extreme H_{m0} produced by extreme value modeling at cluster level and stationary modeling could be in the order of 1 meter by 2050. The design values for H_{m0} are not particularly different in cluster 20, where no statistically significant trend for H_{m0} was identified. One of the questions this report tries to answer is whether NEVA can be applied for the design and risk assessment of hydraulic infrastructure in light of climate change? Despite the limitations surrounding the use

of non-stationary extreme value analysis (NEVA) models, it is expected that engineers are more likely to apply NEVA when an upward trend for extreme H_{m0} is detected. NEVA will likely not be applied when no temporal trend or a downward trend in extreme H_{m0} is present. In both cases, risk aversion towards extreme H_{m0} conditions is the decisive factor. In case of upward trends, a potential increase in extreme H_{m0} design conditions must be accounted for. In the case of negative trends, the uncertainty surrounding temporal trends means it's unclear if the downward trend for extreme $Hm0$ will continue.

Admittedly, the derived extreme value models at cluster level are likely not applicable in advanced stages of the design of infrastructure. Detailed information about the hydraulic boundary conditions at the location of interest is essential for the selection of design strategies. Furthermore, the hydraulic boundary conditions influence the economic and societal factors that must be considered. However, the obtained clusters and corresponding extreme value information can help to construct more accurate extreme value models at the site of interest. The intra-cluster extreme value information at the grid points can be used to construct informative priors that are subsequently used to infer the extreme value distribution describing design conditions at the location of interest.

7.2. Hierarchical Clustering Analysis

The first step in this research involved the hierarchical clustering of the research domain. The resulting clusters share similar extreme wave characteristics including significant wave height (H_{m0}), peak period (T_p) and dominant directions (θ_1 and θ_2). The goal of the clustering output is to improve extreme value modelling. However, the applied hierarchical clustering (HAC) analysis in this research raises concerns because of the generated output. A critical assessment of the HAC analysis is required to judge its applicability. This includes the wave parameters selected for clustering and the use of Hierarchical Agglomerative Clustering algorithms.

7.2.1. Clustering Parameters

For the hierarchical clustering (HAC) analysis in this research, the following wave parameters were used: significant wave height (H_{m0}, peak period (T_p) and dominant directions (θ_1 and θ_2). These wave parameters were selected because they are typically used in the design and risk assessment of hydraulic infrastructure. Furthermore, promising results were obtained by Goharnejad et al. (2022) who used the same wave parameters in their clustering analysis of the North Atlantic Ocean. This section evaluates the practicability of the selected parameters in the perspective of this research and suggests other potential wave parameters to be used as input in the HAC algorithm.

Evaluation of used parameters

Information about the dominant wave directions helps to identify homogeneous and independent subsets of data, thereby meeting the requirements for extreme value analysis (Van Den Bos & Verhagen, 2018). Also, multiple dominant wave directions can be identified in several regions of the North Sea, which is important for the design and risk assessment of hydraulic infrastructure. Based on this, the dominant wave direction is a critical parameter that must always be included in the HAC analysis.

The significant wave height (H_{m0}) is included in numerous failure modes of hydraulic infrastructure. Therefore, it is critical to include information about H_{m0} in the clustering analysis. In this research, the extreme wave threshold (99.5th percentile of H_{m0}) was used to represent H_{m0} at each grid point. However, using $H_{m0,P99.5}$ causes the HAC algorithm to miss information about the most extreme values of H_{m0} above the threshold. Moreover, as was discovered in Chapter 4, the North Sea basin is characterized by various temporal trends for H_{m0}. These trends were not accounted for whilst deriving $H_{m0,P99.5}$. As a result, multiple clusters are characterized by different or even opposite trends for the H_{m0}, most notably in cluster 59 (see Section 5.2.1). These varying intra-cluster trends form a problem when deriving representative extreme value models on the cluster level (see Section 5.2).

The peak wave period (T_P) provides information about the maximum wave energy corresponding to the extreme events. However, T_P was not subjected to further analysis (i.e. temporal trend analysis and non-stationary extreme value analysis, see Section 1.2), raising questions about the inclusion of T_P in this HAC analysis in light of this research. However, only including H_{m0} is not sufficient to have

a complete understanding of the extreme wave climate, and information about T_P is required for the design and risk assessment of hydraulic infrastructure (De Leo et al., 2021). Typically, design values for T_P are calculated by means of regression between H_{m0} and T_P (Antonini et al., 2019) or using bivariate models using H_{m0} and T_p De Leo et al. (2021). Having comprehensive knowledge about T_P in the clusters increases accuracy whilst deriving the design values for T_P on a cluster level. In future work, it would be interesting to study the temporal behavior of T_P as well as the inclusion of T_P in extreme value models on the cluster level.

Exclusion of Geo-Location Parameters
One of the main problems with the clustering output is that different regions are assigned to the same cluster, even when there is no evidence that the extreme wave characteristics are indeed completely similar or that these locations are impacted by the same storms. This is because in HAC analysis, the generated output is dependent of the input values. Furthermore each wave parameter is represented by a single value. This means that if the input values of H_{m0}, T_p, θ_1 and θ_2 are similar in different locations, they are still assigned to the same cluster. Goharnejad et al. (2022) countered this problem by applying the geo-locations of the grid points as input under the assumption that extreme wave events in different geographical locations are the result of different storms. Similar approaches were attempted in this research. However, it was found that even after standardizing, longitude and latitude were too sensitive for the weights assigned to them. Finding the exact weights that yield the optimal clustering solution lies outside the scope of this research. Therefore, clusters with grid points in different geographical locations were split afterwards. Accounting for the geographical locations is a good way to divide clusters based on the storms the extreme waves originate from. However, the use of geo-locations in HAC analysis should be carefully considered.

Suggestions for Other Parameters
To recap, the significant wave height, peak period and dominant wave directions were selected as clustering parameters because of their use in the design and risk assessment of hydraulic infrastructure. However, other parameters representing extreme wave characteristics could be suggested for HAC analysis. One of the main observed shortcomings is that $H_{m0,P99.5}$ does not capture the complete range of extreme H_{m0} values in a region. Potentially, the extreme value distribution parameters can be used in clustering. The location parameter (μ) of the GEV distribution provides information about the relative magnitude of extreme waves between grid points. The scale (σ) and shape (ξ) parameters can provide additional information about the range of extreme H_{m0} at the grid point. Furthermore, ξ provides information about the tail behavior of the extreme value distributions. From Figure 4.9, it can be seen that the GEV distributions are characterized by an upper bound ($\xi_{GEV} < 0$) in the majority of the North Sea domain. Exceptions, where $\xi_{GEV} > 0$, can only be found in the Channel and to the south of the English coast. As for the GP distribution, all values for ξ_{GP} are below zero, meaning that there is an upper bound for the GP distribution in all areas of the North Sea basin.

One of the main shortcomings of the performed HAC analysis is that non-stationarity of extreme H_{m0} is not accounted for. This causes clusters to have different temporal trends for $Hm0$, resulting in a loss of performance of extreme value modelling at cluster level. As a result, using the representative models to derive reliable design values for H_{m0} is difficult. In future works, it is recommended to account for non-stationarity in clustering by applying temporal trend information. Hereby finding clusters with similar trends. Potential parameters describing non-stationarity can be obtained through the Mann-Kendall test or Theil-Sen (b) information. Early findings in this study revealed that binary data is difficult to cluster, making clustering based on the MK test result problematic. However, the P-value (p_{MK}) could be suggested instead. De Leo et al. (2020) showed that there is a strong anti-correlation between p_{MK} and b. However, this still leads to the challenges of choosing an appropriate step size between different p-values (See Section 7.2.2).

7.2.2. Choice of Clustering Method
This thesis applied Hierarchical Agglomerative Clustering (HAC) as the elected clustering method. This was mainly because HAC is straightforward and the dendrogram is an insightful tool to understand the cluster boundaries. However, whether HAC is the optimal clustering strategy in light of this research is debatable.

Limited Information
In HAC analysis, all grid points are characterized by a set of wave parameters that must be represented by a single, characteristic value. This limits the amount of information about the parameter that is presented to the HAC algorithm. Moreover, the number of parameters included in HAC analysis cannot be too large as it will lead to reduction of the clustering quality, further limiting the information that can be presented to the HAC algorithm. The limitation of information causes grid points in different geographical locations to be assigned to the same cluster, even when there is enough reason to believe they do not belong to the same cluster. This is because in HAC analysis, the generated output is always a representation of the input variables. If the input values of H_{m0}, T_p, θ_1 and θ_2 are similar in different locations, they are assigned to the same cluster. This requires additional assessment of the clustering output including splitting and merging of clusters based on the statistical properties of the input wave parameters at the grid points to arrive at the best clustering solution.

Small Dissimilarities between Grid Points
Another limitation that HAC clustering poses is that the HAC algorithm calculates the pairwise dissimilarities between grid points in order to determine cluster formation. However, $H_{m0,P99.5}$ and T_p, show little spatial variation between neighboring grid points, which results in the chaining effect (see Section 3.3.1.2 for more information). To overcome the chaining effect, the values for H_{m0} and T_p are rounded before clustering, thereby increasing the dissimilarities. However, rounding the values presents additional challenges because the right rounding value must be chosen, which is very subjective. In this research, we decided to round the values for H_{m0} and T_p down to the nearest integer. This causes the intra-cluster range for H_{m0} and T_p to be in the order of 1 meter and 1 second respectively, resulting in large uncertainties. This large uncertainty poses problem whilst deriving extreme value models at cluster level because the representative extreme value model for H_{m0} potentially fails to capture information about the most extreme intra-cluster return levels for H_{m0}.

A straightforward solution to overcome this problem is to round the values to smaller step sizes, thereby limiting the intra-cluster variation. This can be especially beneficial to identify more clusters in coastal regions, where extreme H_{m0} will reduce rapidly due to bottom effects. However, because there is always a trade-off between the number of clusters and the boundaries that you want the clusters to represent, decreasing the step-size will require an increased number of clusters. Otherwise not all boundaries for H_{m0} and/or T_P can adequately be represented. Potentially, one can remove T_p form the HAC-input whilst decreasing the rounding steps of H_{m0}, therefore keeping the total number of clusters the same, but that comes at the loss of information about T_p, increasing the uncertainty surrounding T_p if one were to potentially derive extreme design values for T_p on the cluster level.

Number of clusters
Another problem with the HAC analysis is the large number of clusters ($\mathcal{K}$) required to adequately represent the boundaries between the input values of the wave parameters. If $\mathcal{K}$ is too high, the clustering analysis does not meet one of the goals of clustering analysis, which is the reduction of data volume required to study the extreme wave climate in the North Sea domain. A maximum possible number of clusters ($\mathcal{K}_{max} = 50$) was imposed to limit the number of clusters evaluated by the HAC-algorithm. To identify $\mathcal{K}$ that yields the best clustering results, the diagnostic tests Silhouette Coefficient (SC), Calinski-Harabasz score (CH) and the Davies-Bouldin ratio (DB) were evaluated. A closer inspection of Figure 5.3 shows that according to SC, CH and DB, the ideal number of clusters is equal to $\mathcal{K}_{max} = 50$. Potentially, a higher value for $\mathcal{K}_{max}$ is required to find the optimal number of clusters. Choosing $\mathcal{K} = 50$ resulted in numerous grid points being assigned to the same cluster, despite the fact that the HAC input values for these grid-points are different. Figure 5.4 shows that several clusters near the Dutch Coast and English Channel exist where the HAC algorithm did not adequately find the boundaries between input values for $H_{m0,99.5}$ and T_P. This is a reason to suggest that the value for $\mathcal{K}_{max}$ should be increased. Furthermore, after the clustering output was further developed, the North Sea was separated into 63 clusters, exceeding $\mathcal{K}_{max} = 50$. This is further reason to suggest that the considered value for $\mathcal{K}_{max}$ is too low.

Figure 7.1 illustrates the effects of increasing $\mathcal{K}$ on the clustering output. The same HAC input values are used, i.e. $H_{m0,99.5}$, T_P, θ_1 and θ_2 and Ward linkage is used as a linkage mechanism. The value for

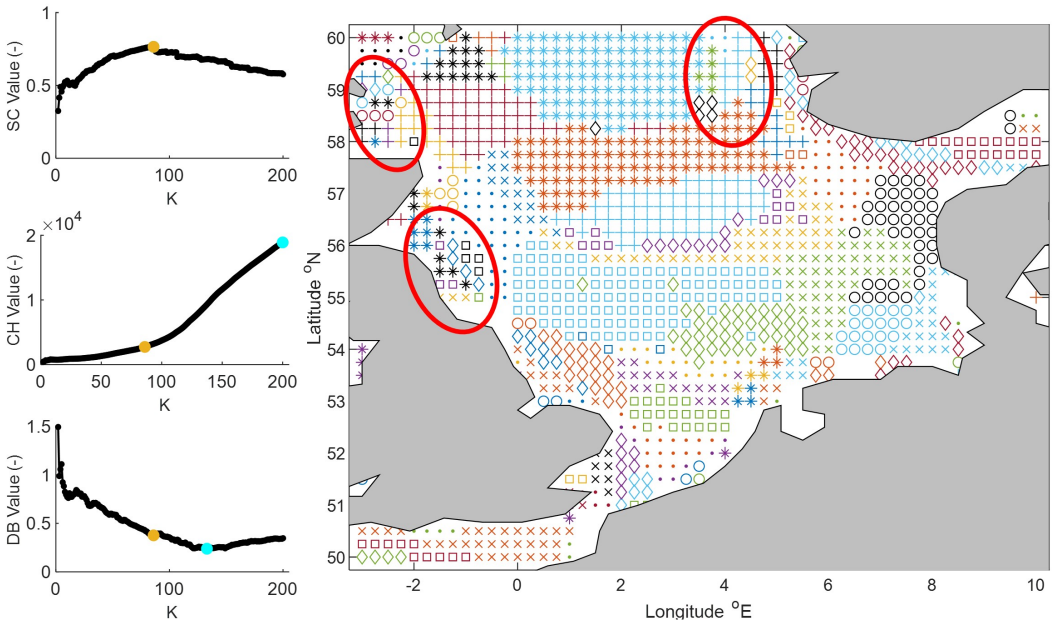


Figure 7.1: Effects of increasing $\mathcal{K}$ on clustering results. Left shows plots for SC, CH and DB for increased values of $\mathcal{K}$. Yellow circle indicates selected value of || to generate clustering output. Cyan indicates optimal number of clusters as indicated by SC, CH and DB. Right shows clustering output for increased value of $\mathcal{K}$.

$\mathcal{K}_{max}$ is increased to 200. Figure 7.1 shows that $\mathcal{K}_{max}$ must be increased to determine $\mathcal{K}$ that yields the optimal clustering output. For SC and DB, the ideal values for $\mathcal{K}$ are 91 and 115 respectively. The optimal number of clusters according to CH remains equal to $\mathcal{K}_{max}$. This can be attributed to the similar nature of Ward linkage and CH. Ward linkage aims to derive clusters clusters with the minimum Within Cluster Sum of Squares (WCSS, see Section 3.3.3.4). Similarly, CH assesses the quality of the clustering output by evaluating the intra-cluster and inter-cluster variances (see Section 3.3.4), thereby always resulting in a preference in higher values for $\mathcal{K}$ for Ward linkage.

For simplicity reasons, the clustering output in Figure 7.1 is based on SC only, meaning that $\mathcal{K}$ is equal to 91. A careful analysis of Figure 7.1 shows that raising $\mathcal{K}$ primarily results in more cluster boundaries in coastal regions, e.g. near the British coast and Norwegian Coast. However, the new cluster boundaries are primarily based on the input values for θ_1 and θ_2 because of the small dissimilarities that exist between the values. Further increasing $\mathcal{K}$ does not help to identify clusters based on geographical locations because they are not included as input parameters. Also, cluster boundaries based on $H_{m0,99.5}$ T_p are not identified because of the coarse pre-processing of these values and no more "categories" can be identified. Applying $\mathcal{K} = 115$ (optimal $\mathcal{K}$ according to DB), cluster boundaries based on θ_1 are θ_2 is further enhanced. In engineering practice, the dominant wave direction are typically defined as a range when performing extreme value analysis. Therefore, clusters containing very precise information about the dominant wave directions are not required. Based on this, it was decided that $\mathcal{K}_{max} = 50$ is appropriate for this research.

ERA5 Resolution

It could be argued that the spatial resolution of the ERA5 Reanalysis dataset too coarse for the HAC algorithm to be properly functional. The spatial resolution is 0.25 degrees (approximately 30 km) in longitude and latitude resulting in grid points that must represent a surface area of $900\ km^2$. Over a surface area that big, waves can undergo substantial development meaning that extreme waves will show different characteristics between neighboring grid cells. As a result, the obtained clusters have relatively small number of grid points, especially in coastal areas. The applicability of clustering analysis increases for datasets with higher spatial resolutions because of the relative reduction in data volume (i.e. one time-series per cluster) increases compared to small resolution datasets.

Suggestions for Other Clustering Strategies
In short, whether HAC is the optimal clustering method for storm wave climate is debatable. The limited information that can be presented to the HAC algorithm and small spatial dissimilarities form challenges in the clustering analysis. Furthermore, the optimal number of clusters yielded by the algorithm exceeds the number that is deemed reasonable given the objective within this research. Given the benefits that a clustering analysis provides, ideally, a different clustering method is studied to determine whether better results can be obtained. The selected clustering approach must enable more information about the extreme wave characteristics to be included. A potential solution can be found in Goharnejad et al. (2022) who applied self-organizing maps (SOM) to identify clusters with similar extreme wave climates in the Northwest Atlantic ocean. In Goharnejad et al. (2022), the same wave parameters (H_{m0}, T_p and θ) are used as input, but the full time-series are considered at each grid point. The wave parameters are separated into a set of bins, and the percentage of occurrence of each value within each bin is subsequently computed at each grid point. Geo-locations serve as additional input to ensure that extreme wave climates in different locations are not clustered together. The SOM algorithm subsequently uses the probabilities of occurrence to identify clusters that have similar percentages in terms of H_{m0}, T_p and θ.

Another potential clustering solution is presented in Sartini et al. (2018). They performed a clustering analysis using regional frequency analysis (RFA) to identify regions that share similar storm footprints. Both the magnitude of the extreme wave events as well as the time of occurrence are used to calculate the probability that different grid-points are affected the same extreme wave events. This probability serves as dissimilarity information while clustering regions that are often impacted by the same storms. This clustering method accounts for the spatial difference between the occurrence of extreme wave events, however does not define cluster boundaries based on different magnitudes of extreme H_{m0} and T_P, resulting in high intra-cluster variances.

Whether a successful clustering analysis can be performed in light of the goals of this research (i.e. non-stationary extreme value analysis at cluster level) must be investigated further in the future. For the clustering output to be deemed applicable, the clustering approach must satisfy the following requirements:

- The clustering output must be determined based on complete time-series for H_{m0}, T_P and θ;
- The clustering analysis must be able to differentiate between extreme wave events originating from different storm events. Either by accounting for the time of occurrence of extreme wave events or by including geo-locations of the grid-points in the clustering analysis;
- The resulting clusters must account for non-stationarity in extreme wave climate.

7.3. Accounting for Non-Stationarity in HAC Analysis

This section presents a short exploration of an updated clustering strategy to account for non-stationarity of the significant wave height (H_{m0}). Accounting for non-stationarity is expected improve the performance of non-stationary extreme value modelling at cluster level. The first method involves detrending of extreme H_{m0} data prior to clustering. The second method involves adding temporal trend information as an additional input parameter.

7.3.1. Detrending Before clustering

Detrending causes all H_{m0} observations to be centered around zero. Therefore, using the extreme wave threshold to represent H_{m0} is no longer applicable. Detrending prior to clustering requires another representation of H_{m0}. Representative values for H_{m0} are determined based on detrending Annual Maxima (AM) H_{m0} observations in a similar manners as is done whilst performing extreme value analysis at cluster level (see Section 7.2). The maximum observation after detrending (H^*_{m0}) is selected to represent H_{m0} at the grid points. Figure 7.2 shows the updated clustering output. Furthermore, other input values are unchanged, the selected linkage mechanism is Ward and the value for $\mathcal{K}$ is 50. Figure 7.2 shows that the cluster boundaries become poorly defined in several regions throughout the research domain. Grid points are assigned to different clusters on an alternating basis. This increases the ambiguity regarding the choice of a cluster for extreme value analysis and required

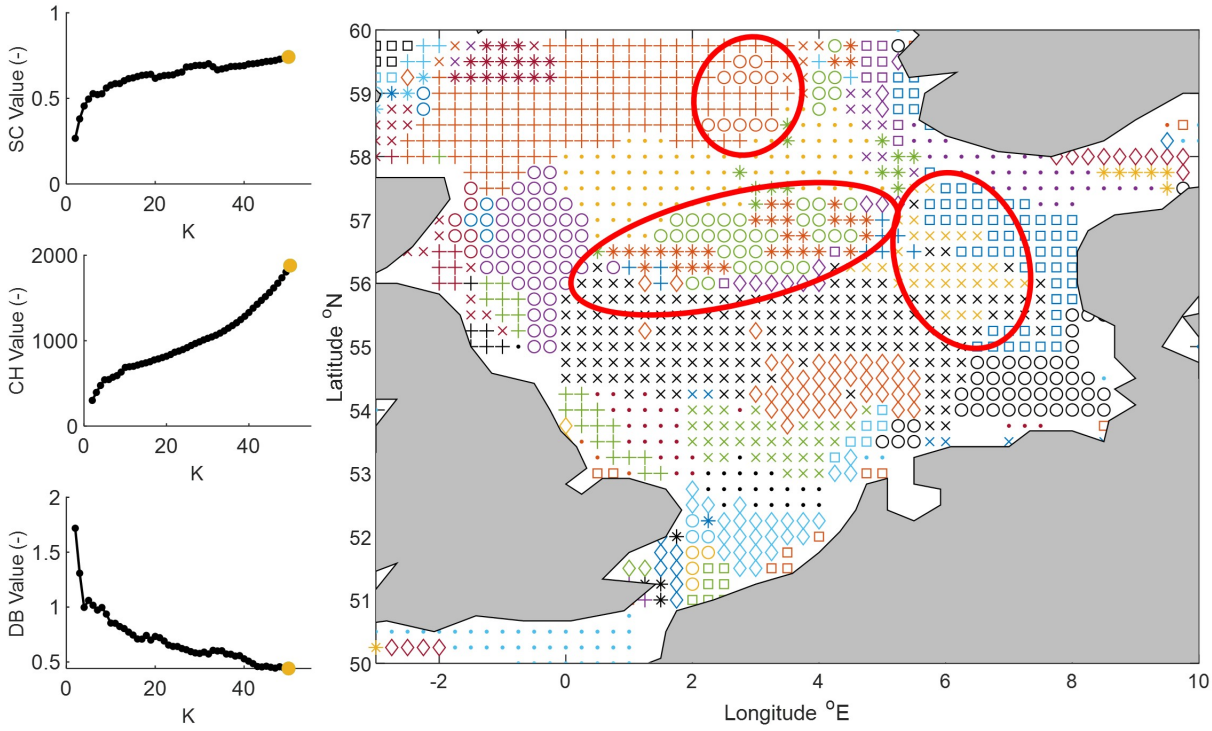


Figure 7.2: Illustration showing the effects of detrending H_{m0} prior to clustering. Left figures show the results of SC, CH and DB. Right shows the resulting clustering map for $\mathcal{K} = 50$. All other input variables are unchanged.

additional assessment of the clustering output to see if the clusters can be merged. Furthermore, the region corresponding to cluster 59 is still assigned to the same cluster. This indicates that the problem surrounding heterogeneous non-stationarity in clusters remains unsolved.

The clustering output in Figure 7.2 is a direct result of using H^*_{m0} as input. The values of the time-series for H^*_{m0} are all centered around zero as a result of detrending. This means that the clustering output is based on the range in extreme H_{m0} instead of the relative magnitude of H_{m0} between different grid points. Preferably, detrending of H_{m0} is performed in such a way that the relative magnitude for H_{m0} between grid points remains, so that the clustering analysis provides information about the magnitude of extreme H_{m0} in different regions of the research domain.

7.3.2. Trend Information as Input Parameter

Instead of detrending H_{m0}, it is potentially more useful is to add temporal trend information as an additional input parameter. This can for example be done by adding the Theil Sen estimator (b), which describes the temporal slope of H_{m0} observations. Figure 7.3 shows the clustering results when b is added as an additional input parameter. All other input parameters are unchanged. Ward is selected as linkage mechanism and the value for $\mathcal{K}$ is equal to 50. Figure 7.3 shows that the region corresponding to cluster 59 is now split up in two parts. The region in the West, where the downward trend was observed, is now split from the rest of the cluster. However, this means that Eastern part of this region includes a part with an upward trend and no trend respectively. Furthermore, adding b as clustering input brings additional difficulties:

- More input parameters requires an increase of $\mathcal{K}_{max}$ so that the clustering algorithm can adequately identify cluster boundaries based on the input values.
- The values of b exhibit small dissimilarities between spatially neighbouring grid-points. This potentially results in the chaining problem. Subjective pre-processing is required to prevent chaining of grid points.
- The value of b depends on the length of the observation period, thereby bringing additional ambiguity to the clustering analysis.

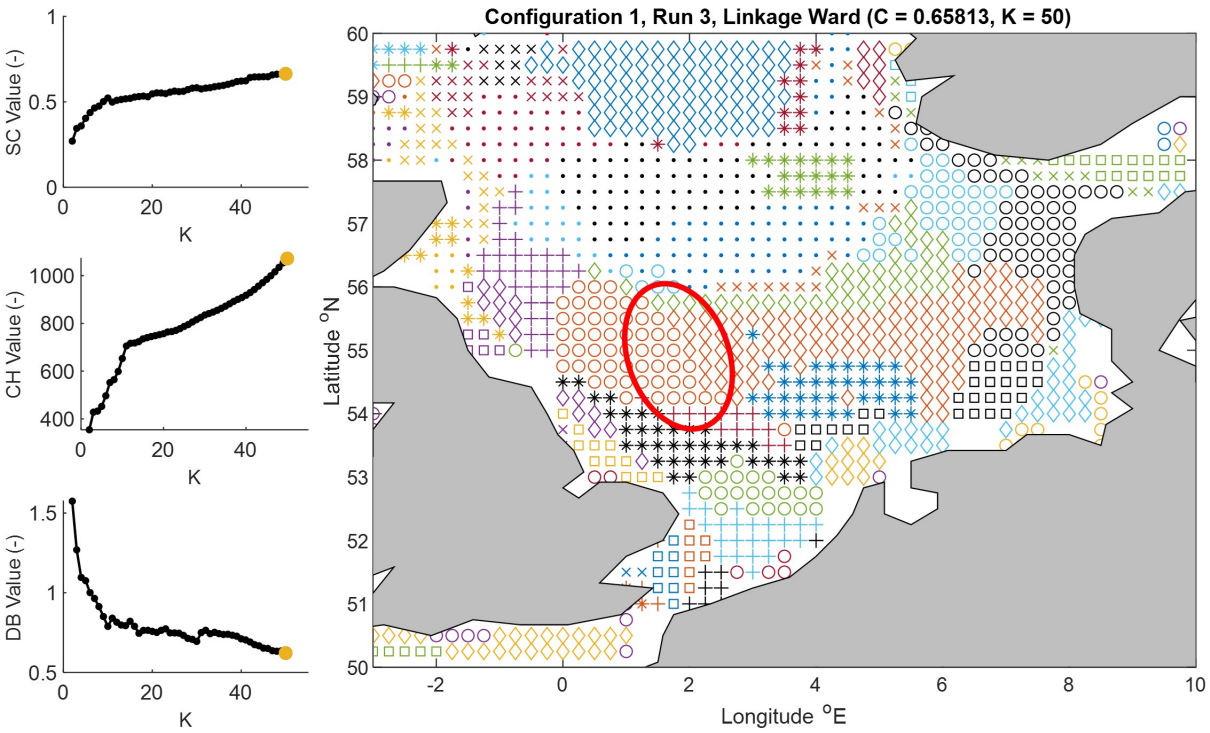


Figure 7.3: Illustration showing the effects including b as HAC input parameter on the clustering output. Left figures show the results of SC, CH and DB. Right shows the resulting clustering map for $\mathcal{K} = 50$. All other input variables are unchanged.

Based on the findings in this section, adding b as input parameter shows more promise than detrending H_{m0} data prior to clustering. However, a decision regarding the optimal strategy to account for non-stationarity in extreme wave parameters whilst performing clustering analysis cannot be made and additional research is required.

7.4. Temporal Trend Analysis

In Chapter 4, the temporal trends of the extreme wave climate were performed to study whether significant trends exists in the extreme wave climate in the North Sea domain. Not only does this give important information regarding potential effects of climate change but also where non-stationarity potentially can be applied in the extreme value models of the extreme wave climate. Since the extreme waves in the North Sea can be characterized by wind waves, temporal trend analysis of the extreme wind speeds has been performed to study whether they can explain the observed trends in extreme H_{m0}. This section covers the limitations and assumptions made for the temporal trend analysis and the implications they have on the results. Also, suggestions will be provided on how the temporal trend analysis can be improved in further research.

7.4.1. Limitations and Suggested Improvements

Fixed Thresholds in POT Analysis

The temporal trend analysis for extreme H_{m0} events selected using POT analysis is further influenced by the application of the threshold (u). This research considers fixed values for u. The Mann Kendall (MK) and Theil-Sen (TS) tests applied to study the non-stationarity only consider the magnitude of extreme H_{m0} exceeding u. This means that extreme H_{m0} events included in the non-stationarity analysis have a minimum value for H_{m0} equal to u, making it more difficult to detect and quantify a statistically significant trend, especially for longer time scales. The presence of long term trends in extreme H_{m0} is potentially obscured if H_{m0} events no longer exceed the threshold. This is illustrated in Figure 7.4. This illustration shows a cyclical signal in the data. However, the MK and TS tests are only performed over the data selected during the POT analysis, meaning that no statistically significant trend is selected in the data.

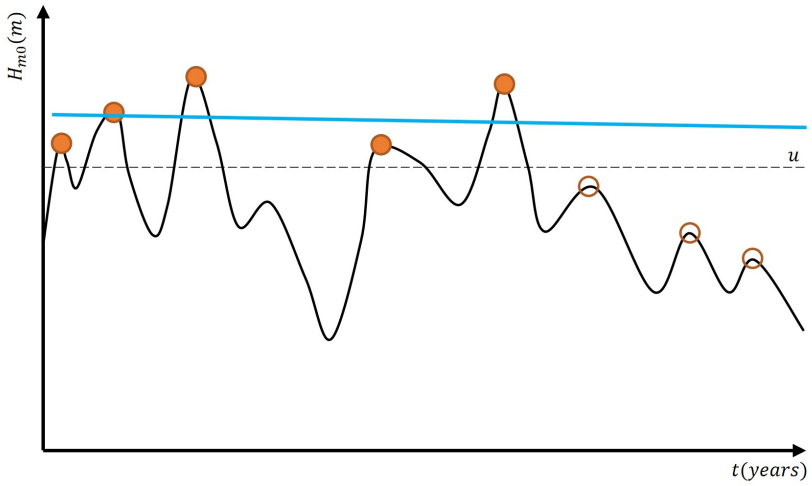


Figure 7.4: Illustration showing the limitations posed by Peak over Threshold for temporal trend analysis. Filled circles indicate extreme events considered in the Analysis, empty circles represent ”extreme events” not considered in the trend analysis

Full directional spectrum

A last limitation of the temporal trend analysis as performed in this research, is that the full directional wave spectrum was considered for the temporal trend analysis in the complete North Sea basin. This means that the selected extreme events do not necessarily have a homogeneous origin, especially at grid points that are characterized by multiple dominant wave directions. A short study for cluster 8 (Dutch coast) showed that accounting for the dominant wave direction results in the detection of temporal trends in regions were no trends were detected when considering the full directional spectrum. Furthermore, opposing trends were observed in cluster 8 for the different dominant wave directions. This shows that accounting for the dominant wave directions is essential to get a comprehensive understanding of the temporal trends in extreme wave climate at the cluster level.

The assumptions and limitations described above have the following effects on the temporal trend analysis. The regions where significant temporal trends were detected using POT selection are smaller than regions detected using AM selection, and also the magnitude of the trend (b) smaller than for the trends detected using annual maxima. Apart from the magnitude of H_{m0}, also the annual number of threshold exceedances (N_{exc}) and exceedance duration (T_{dur}) were analyzed in the temporal trend analysis. The observed trends for N_{exc} especially show results that are in accordance with the detected trends for AM H_{m0}, meaning that N_{exc} is a good indicator to analyze trends in the intensity of the extreme wave climate, even when the tests for H_{m0} intensity show very small regions with a trend.

Suggested improvements

Improvements can be made regarding the temporal trend analysis improving the quality of this research as a whole. The temporal trend analysis using POT analysis can potentially be improved by applying a time-varying threshold so that values for H_{m0} are included in the temporal trend analysis that previously were below the considered threshold. Another method to improve the temporal trend analysis for POT is by considering the threshold value instead of the magnitude of the exceedances. If significant temporal trends exists for extreme H_{m0}, similar trends can also be observed for the threshold value. Taking the threshold value also means only one value per year needs to be considered, partially removing the problem surrounding the seasonality (De Leo et al., 2020).

7.4.2. Effects of Selected Temporal Horizon

One of the major discoveries in chapter 4 is that the direction and magnitude of the detected trends vary based on the considered temporal horizon. In this research, two temporal horizons are analyzed, i.e. a short term horizon (1990-2020) and a long term horizon (1950-2020). The fact that observed trends vary is reason to believe that the non-stationarity of extreme H_{m0} in the North Sea is not monotonic but can be described by oscillatory cycles. It is therefore very difficult to make future projections about

the temporal development of extreme H_{m0} to account for in design and risk assessment of hydraulic infrastructure. To make better projection about the long-term behavior of the extreme wave climate, a better understanding of the non-stationarity of the forcing mechanisms of wave generation (i.e. wind storms) is required.

It was also discovered that the temporal trends in the extreme wind speeds show similar spatial and temporal distributions as the trends for extreme H_{m0}. The cyclic behavior of the extreme H_{m0} events can potentially be explained by a cyclic behavior in the physical drivers of the storm climate. The storm climate in Northwestern Europe is largely influenced by the North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) (see Chapter 2.1). A large scale study into the effects NAO and AO on wave generation is not performed in this research. However, Freitas et al. (2022) show that there are large correlations between the AO index and extreme wind and wave climate in the North Atlantic Ocean near North Western Europe, especially in winter months. Freitas et al. (2022) state that this correlation remains until the end of the 21st century and that AO+ events are becoming less frequent, but increase in magnitude. This is reason to believe that the peak magnitudes of extreme wave events will increase towards the end of the century. A more detailed study of the influence of the NAO and AO indices in the North Sea domain specifically, potentially provides a better understanding about the non-stationarity of the wave climate and help make better decisions regarding the modelling of extreme wave events in the future.

One of the identified limitations that surrounding the applicability of NEVA in engineering practice is that this research does not consider future projections of extreme wave conditions. Future projections provide crucial information surrounding the future non-stationarity of the extreme wave climate. Studies such as Bonaduce et al. (2019) and Grabemann et al. (2015) combined general circulation models (GCMs) and wave forcing models (Wave Watch III, WW-III) to project the future wave climate in the North Sea and Baltic Sea for several emission scenarios (RCPs). Grabemann et al. (2015) discovered that towards the end of the 21st century, a downward trend in extreme H_{m0} in Western regions of the North Sea and an upward trend in the Eastern regions can be projected. This matches observations for the long-term temporal trend analysis in this report. However, Bonaduce et al. (2019) only projected a downward trend towards the end of the 21st century for RCP 8.5 scenarios. These findings further reinforce the uncertainty surrounding the future projections of the extreme wave climate making decisions about extreme wave modelling difficult. A comprehensive study of the extreme wave climate including future projection is therefore required to improve decision making regarding the applicability non-stationary models for the design and risk assessment of hydraulic infrastructure.

7.5. Extreme Value Analysis

In Section 5.2, we evaluated how representative extreme value models can be derived that describe extreme H_{m0} values at cluster level. Five different techniques to obtain representative time-series for extreme H_{m0} were considered that were tested on three different clusters. To account for temporal variations of extreme H_{m0}, a non-stationary component was added to the location parameter of the extreme value model that considers the intra-cluster trend information in extreme H_{m0}. This section reflects on the assumptions and limitations and their effect on the extreme value modelling at cluster level. Furthermore, the benefits of applying Bayesian Inference (BI) in light of this research over Maximum Likelihood Estimation (MLE) are discussed.

Differences with Conventional Extreme Value Models

The first discussion point is the considered method of including non-stationarity in the GEV-model, which differs from conventional modelling. In this research, the intra-cluster extreme H_{m0} data is detrended before the GEV parameters (θ) are inferred using Bayesian Inference ($\theta = \{\mu^*, \sigma^*, \xi^*\}$). Subsequently, non-stationarity is included in the GEV model by adding the TS parameters (b and b_0) to μ^*. The location parameter of the extreme value model becomes time-dependent and the GEV distributiono parameters subsequently read: $\theta = \{\mu^* + b \cdot t + b_0, \sigma^*, \xi^*\}$).

This practice was elected because this research considers the extreme H_{m0} values at cluster level. The intra-cluster range of extreme H_{m0} is quite large as a result of pre-processing of the input parameters in the Hierarchical Clustering analysis (see Section 7.2.2). Three methods to derive representative

time-series involved aggregation techniques to see how grid-point information can be pooled to obtain representative time-series for extreme H_{m0} at cluster level. For these aggregation methods to work properly, the H_{m0} data at the different grid points must be made homogeneous, which is achieved by detrending the extreme H_{m0} data.

Conventional practice would account for non-stationarity by inferring the two parameters representing the location parameter (μ_1 and μ_0) directly using BI such that the GEV parameters are described by $\theta = \{\mu_1 \cdot t + \mu_0, \sigma, \xi\}$. As can be seen, the major difference between the method presented in this research and the conventional method lies in the fact that different location (μ) parameters will be inferred. μ is particularly of interest as it determines where the GEV distribution is located with respect to H_{m0}. Different definitions of μ may therefore lead to different extreme H_{m0} return levels. Studying the differences between the values for μ is left outside the scope of this research.

Bayesian Inference and Maximum Likelihood Estimation
This research inferred the parameters of the non-stationary GEV distribution using Bayesian Inference (BI). Apart from BI, Maximum Likelihood Estimation (MLE) is often employed to infer distribution parameters. BI brings advantages over MLE in light of this research. First of all, MLE only employs observed data to infer distribution parameters. Using MLE in this research means that the GEV parameters (μ^*, σ^* and ξ^*) would be inferred based on the representative time-series for H^*_{m0} only. The use of priors ($P(theta)$ in BI enables for the construction of informative priors using extreme value information from grid points with similar extreme wave characteristics. This provides more accurate models describing extreme H_{m0} conditions (Antonini et al., 2019; Rickets, 2021).

Another advantage of BI over MLE lies in its ability to quantify uncertainty. In the Frequentist approach of MLE, the GEV parameters are assigned a fixed value. Moreover, no probabilities can be assigned to describe the uncertainty of the GEV parameters (Bickel & Lehmann, 2012). This is described in more detail in Section 3.4.2.1. Although methods exist to describe the uncertainty for MLE models, such as Fisher's approach (Ly et al., 2017), uncertainty quantification of the GEV parameters using BI in more straightforward. By examining the posterior distributions of the GEV parameters confidence intervals can directly be constructed (Cheng et al., 2014). This becomes especially useful when the number of model parameters becomes large. For example, when non-stationary location parameters are inferred directly, or when scale and location parameters are also assumed to be non-stationary. Bayesian Inference computes the posterior distribution for all considered model parameters. Subsequently, uncertainty bounds of estimated return levels can be obtained by taking the uncertainty in all model parameters into account.

Accounting for Non-Stationarity
The major consideration in the computation of the representative GEV models is how to account for potential non-stationarity. In this research, we applied the intra-cluster trend information about b and b_0. This becomes challenging when the intra-cluster values for b and b_0 vary in magnitude and direction. This is the direct result of not accounting for non-stationarity during the Hierarchical Clustering analysis. The H_{m0} data from the clustering centroid showed promising results for its ability to infer representative extreme value models for clusters 12 and 20, where similar intra-cluster values for b and b_0 can be observed. However, the same approach failed to capture the most extreme return levels for cluster 59. (see Figure 5.20). In the temporal trend analysis, opposing trends in extreme H_{m0} can be observed between 1950 and 2020. As a result, the grid point that yielded the clustering centroid was located in a zone with negative values for b, while the majority of the grid points showed either no trends or positive values for b.

Another assumptions made in this part of the research is that the non-stationarity in the GEV model can be described by a linear trend in time. The temporal trend analysis (Section 4.2.2) showed that the direction and magnitude of the temporal trends differ depending on the considered temporal horizon. This implies that temporal trends in extreme wave climate are most likely not linear, but show long term oscillations. The oscillations in extreme H_{m0} make the applicability of non-stationary extreme value models for the design and risk assessment of hydraulic infrastructure debatable. Without a thorough understanding of the non-stationarity of the extreme wave climate, including the physical drivers behind

wave generations, the success of using non-stationary models to infer design values for H_{m0} cannot be guaranteed.

Approximated Stationary Distributions
The inferred non-stationary models have a location parameter that includes a temporal covariate ($\mu = \mu^* + b_0 + b \cdot t$). A methodology for approximating extreme value distributions under the assumption of stationarity is explored. In approximated stationary modelling (AST), the GEV parameters corresponding to the median of the observation period (i.e. 1985, $t = 35$) are used to approximate H_{m0} return levels under stationary conditions.

The key assumption behind AST modeling is that under stationary conditions, the median value of the linear trend in effective return levels over the observation period can be employed to derive design values. A short evaluation of the performance of AST modelling is presented in Section 6.1.3. It was concluded that AST modelling provided similar results as stationary modelling for the 50-year return period. However, for the 1000-year return period, the difference between the H_{m0} design values between the two models was in the order of $0.5\,m$. This is a significant difference for H_{m0} that must be accounted for in design and risk assessment of hydraulic infrastructure. This means that the applicability of AST modelling is debatable. Further potential limitations affecting the applicability of AST modelling include:

- Outliers in the set of extreme observations influence the derivation of the GEV parameters. This turn influencing the effective return levels. A critical assessment of the derived GEV parameters for stationary modelling and AST modelling is not performed in this research;
- Furthermore, the effective return values derived using AST modelling depend on b and b_0. The values for b and b_0 in turn, depend on the considered temporal horizon of the observation period. AST modelling therefore still depends on a non-stationary component.

The fact that these limitations are not properly understood, combined with limited testing with mixed results, makes ant conclusions regarding the applicability of AST modelling impossible. A critical analysis of AST modelling is required to study its applicability. In the meantime, it is advised to use conventional stationary models if design values under stationary conditions are desired.

Improving Clustering Analysis
What becomes apparent is that the limitations and choices related to the clustering analysis affect the applicability of extreme value models to describe H_{m0} at cluster level. The fact that clusters with heterogeneous temporal trends are obtained, makes deriving applicable models particularly complex. This further supports the suggestion that an improved clustering analysis is required to obtain more meaningful results regarding the applicability of non-stationary models at cluster level. Preferably, the improved clustering analysis analysis is able to include more detailed information about the extreme wave characteristics (see section 7.2.2) as well as account for non-stationarity of the extreme wave climate. This reduces the ambiguity surrounding the selection of the appropriate values for b and b_0 to be included in the non-stationary models.

Limited Number of Clusters
Extreme value models derived using extreme H_{m0} observations from the clustering centroid showed the most promising results for describing extreme H_{m0} at cluster level, in both stationary and non-stationary cases. However, only three clusters were analyzed to study the applicability of this methodology. Moreover, cluster 59 already showed problems as a result of the complex non-stationarity at the grid points. This severely limits any conclusions regarding the success of this methodology. An improved clustering analysis and testing this method on multiple clusters is required to verify whether the extreme H_{m0} data from the clustering centroid is indeed the best method to derive non-stationary models describing the extreme wave climate on a cluster level.

7.6. Further Research Limitations

Apart from the limitations discussed above, other limitations need to be discussed. This includes the exclusion of other wave parameters, such as the peak period and the dominant wave direction. Furthermore, the role of water level on the extreme wave climate must be addressed.

7.6.1. Wave Periods and Dominant Wave Direction

This research only considers the significant wave height (H_{m0}) to study non-stationarity in the extreme wave climate and to assess the applicability of non-stationary extreme value models at cluster level. Other wave parameters used in the design and risk assessment of hydraulic infrastructure, including mean wave period ($T_{-m1,0}$) the peak wave period (T_P) and dominant wave direction (θ) are not considered in this research. Both T_P and θ play an important role in the design and risk assessment of hydraulic infrastructure. T_P describes the waves that contain the most energy in the wave field. θ is especially important for coastal infrastructure as it determines the incoming wave angle that hydraulic infrastructure is subjected to. Temporal changes for $T_{-m1,0}$, T_P and θ mean that the hydraulic boundary conditions change over time. This means that it is crucial to have an understanding about the non-stationary of $T_{-m1,0}$, T_P and θ. Further research should therefore also focus on deriving models for $T_{-m1,0}$, T_P and θ at cluster level and account for non-stationarity of these parameters.

7.6.2. Water Level and Water Depth

A final remark is made about the role of water level in the extreme wave climate. The local water level is important in the context of extreme waves because deeper water can accommodate higher waves. This is especially important for the design and risk assessment of coastal infrastructure. Higher water levels and subsequent higher water depths reduce the effect of depth induced breaking. Depth induced breaking plays an important role in wave attenuation in coastal areas. Increased water levels allow higher waves to approach coastal structures, compromising their safety. In the North Sea domain, especially along the English, Dutch, German and Danish coast, a potential increase in extreme water level will be troublesome due to the relatively shallow water water depths that prevail in this area (see Figure 4.2).

As a result of climate change, sea level rise is expected to reach 1 meter before the end of the 21st century (KNMI, 2021). The role of sea level rise on extreme water levels and subsequently extreme wave climate therefore cannot be ignored. Furthermore, additional water level components including wave and wind set up play a more significant role in shallow water, further increasing water depths in nearshore areas (Jonkman et al., 2018). As a consequence, coastal infrastructure will be subjected to even more extreme wave events. It is therefore strongly recommended to perform research into the non-stationarity of extreme water levels and its effects on the extreme wave climate.

8

Conclusion

This chapter combines the key results from the research and the Discussion to arrive at final answers for the research questions. In order to answer the main research question, four sub-research questions were defined. The sub-research questions will be answered before their answers are used to answer to the main research question.

Sub-research question 1: What are the observed changes over time of the extreme H_{m0} events in the North Sea basin and what are the underlying causes?

In Chapter 4, the non-stationarity of the North Sea extreme wave climate was studied using a temporal trend analysis. Extreme H_{m0} events were selected Annual Maxima (AM) and Peak over Threshold (POT). Mann-Kendall (MK) tests were employed to detect the presence of statistically significant trends. The Theil-Sen (test) calculated the magnitude of statistically significant trends. Furthermore, two different temporal horizons are studied.

For the short temporal horizon (1990-2020), several regions in the North Sea basin were identified that strictly showed downward trends for extreme H_{m0}. The magnitude of these trends varied between $1-2 cm/yr$ for POT extremes and exceeded $3 cm/yr$ for AM extremes. Furthermore, the MK test for the exceedance frequency (N_{exc}) showed downward trends in regions in accordance with detected trends for AM H_{m0} extremes. These findings are reason to believe that the temporal horizon between 1990 and 2020 is dominated by decreasing intensity of extreme H_{m0}. The average threshold exceedance duration (T_{dur}) showed nearly no trends for the short temporal horizon. For the long temporal horizon (1950-2020), two different temporal trends were identified. The western North Sea is characterized by downward trends for extreme $Hm0$, whereas the eastern and northern region as well as a region South of britain are characterized by upward trends. The slope of the downward trends was in the order of $\leq 1\ cm/year$ for extreme H_{m0} values selected using both AM and POT. The slope of the upward trends showed values between $1-2\ cm/year$ for AM H_{m0} and $\leq 1\ cm/year$ for POT H_{m0}. The non-stationarity of N_{exc} showed similar spatial patterns as AM H_{m0}. These results verify that extreme H_{m0} events between 1950 and 2020 have become more intense in the Eastern parts of the North Sea whilst becoming less intense in Western regions.

Temporal trend analysis of extreme wind speeds (U_{10}) showed similar results as were found for extreme H_{m0}, i.e. downward trends for the short temporal horizon and a combination of upward trends (East) and downward trends (West) for the long temporal horizons. The spatial distribution of the temporal trends for U_{10} is very similar to the spatial distribution of temporal trends in extreme H_{m0} for both temporal horizons. The North Sea extreme wave climate is characterized by locally generated wind waves. Therefore, there is reason to believe that the temporal evolution of extreme H_{m0} can at least partially be explained, by similar temporal evolution of extreme U_{10}.

Further analysis at cluster level showed that accounting for dominant wave directions (θ) matters for trend detection. Within clusters 12, 20 and 59, more grid-points showed temporal trends when accounting for θ. Cluster 8 even showed that opposing trends may be detected for different θ. Accounting for θ is therefore important when accounting for non-stationarity in extreme H_{m0} during the design of hydraulic infrastructure.

Sub-research question 2: How can Hierarchical Agglomerative Clustering (HAC) be employed to identify regions in the North Sea that exhibit similar extreme wave characteristics?

During the Hiearchical Clustering (HAC) analysis, two wave parameter combinations were tested. Initially, it was found that a combination of the 99.5 percentile of the significant wave height ($H_{m0,P99.5}$), the corresponding peak period (T_p) and the dominant wave directions (θ_1 and θ_2) yielded the best clustering results. Equal weight was assigned to all input values, and the number of identified clusters is equal to $\mathcal{K} = 50$. The clustering output depends on the input values. Grid-points in different locations were assigned to the same cluster, despite there being no statistical evidence that the grid points in question share similar extreme wave characteristics, or that extreme wave events at these grid-points are the result of the same storm. Moreover, several grid-points in coastal regions were assigned to different clusters despite their being statistical evidence of the considered wave parameters to assign them to the same cluster. This required further development of the clustering output. In the end, 63 different clusters were identified, each with their own characteristic values for $H_{m0,P99.5}$, T_p, θ_1 and θ_2.

Clustering analysis shows potential for identifying regions with similar extreme wave characteristics. However, the combination of HAC analysis and the selected input parameters is not considered to be the optimal technique for this research. Limited information can be presented to the HAC algorithm, causing grid-points with different extreme wave characteristics to be clustered, because they share similar input values. Furthermore, small spatial dissimilarity between neighboring grid points for H_{m0} and T_p requires subjective pre-processing including a subjective selection of appropriate rounding values to prevent chaining of grid-points. As a results, clusters have large ranges for extreme H_{m0} and T_p. Finally, not accounting for non-stationarity of extreme H_{m0} means that the obtained clusters show heterogeneous non-stationarity for H_{m0}. This is troublesome when deriving representative (non-stationary) extreme value models describing extreme H_{m0} at cluster level.

Sub-research question 3: How can the derived clustering output and temporal variability be applied to derive representative non-stationary statistical models that describe extreme H_{m0} at cluster level?

Representative extreme value models are derived by detrending intra-cluster AM H_{m0} data (H^*_{m0}) before fitting the Generalized Extreme Value (GEV) parameters (μ^*, σ^* and ξ^*) using Bayesian Inference (BI). Potential non-stationarity is accounted for by adding the intra-cluster trend information (b and b_0) to μ^*, so that the GEV parameter read $\theta = \{\mu^* + (b \cdot t + b_0), \sigma^*, \xi^*\}$. Five techniques to obtain representative time-series for H_{m0} were evaluated in clusters 12, 20 and 59. The extreme value models obtained using the AM H_{m0} observations from the clustering centroid showed the most promising performance for describing extreme H_{m0} at the cluster level. The confidence bounds of the extreme value model capture the most extreme H_{m0} return levels of the intra-cluster grid points when accounting for non-stationarity without severe overestimation of the extreme wave conditions.

Accounting for non-stationarity relies on the selection of appropriate values for b and b_0. Problems arise in cluster 59, which exhibits heterogeneous temporal trends. The extreme value model fails to accurately represent intra-cluster extreme H_{m0} statistics because there is a difference between the non-stationarity included in the extreme value model at cluster level and the non-stationarity at the individual grid-points. Only when all intra-cluster grid points are characterized by similar trends, the extreme value models at cluster level might be preferred over conventional modelling. It should be emphasized that

the number of clusters studied was insufficient to draw definitive conclusions about the applicability of representative models at cluster level. It is strongly advised to improve the clustering analysis to arrive at more conclusive answers. The clustering output must account for potential non-stationarity of extreme H_{m0} and the number of tested clusters must be increased.

Sub-research question 4: Do representative extreme value models at cluster level form a useful alternative compared to stationary modelling at grid point-level in the design and risk assessment of hydraulic infrastructure in light of a changing climate?

A practical example is presented to assess the application of non-stationary models at cluster level in light of climate change. The temporal evolution of effective extreme H_{m0} return levels is analyzed for typical return periods for hydraulic infrastructure. It was found that non-stationary models may be preferred over conventional stationary modelling in clusters fully characterized by an upward trend for extreme H_{m0}. Using non-stationary modelling could be essential when accounting for increasing extreme H_{m0} conditions to which hydraulic infrastructure may be subjected due to climate change. In clusters where no temporal trend for extreme H_{m0} is present or in clusters characterized by heterogeneous temporal trends, non-stationary modelling did not provide benefits over stationary modelling.

Even when a cluster is characterized by an upward trend for extreme H_{m0}, the application of non-stationary models is not straightforward. The observed temporal trends for extreme H_{m0} depend on the considered temporal horizon. It is believed that the non-stationarity is better described by a cyclic rather than a monotonic trend. Without understanding the underlying cause of the non-stationarity and without future projections of the variability of extreme H_{m0}, it is not possible to place the observed temporal trends in a broader perspective regarding climate change. This limits the applicability of non-stationary extreme value models at cluster level for the design and risk assessment of hydraulic infrastructure in light of climate change.

How can clustering analysis and non-stationary extreme value models at cluster level be used to characterize the statistics of extreme H_{m0} events in the North Sea and when do non-stationary extreme value models form a practical alternative compared to traditional stationary models for the design and risk assessment of hydraulic infrastructure?

The potential for a design methodology involving a clustering analysis and non-stationary extreme value analysis is presented in this research. However, the hierarchical clustering (HAC) used in this study does not produce the required output for this methodology to be implemented successfully. Promising results are obtained for using extreme H_{m0} data of the clustering centroid to compute non-stationary extreme value models for H_{m0} at cluster level. Despite its potential to represent extreme wave characteristics at cluster level, caution is required when applying this methodology in the design and risk assessment of hydraulic infrastructure. Not accounting for non-stationarity during clustering reduces the performance of non-stationary extreme value models at cluster level. The non-stationary extreme value models at cluster level only form a potential alternative to conventional models when all intra-cluster grid-points exhibit similar temporal non-stationarity. Moreover, the intra-cluster grid points must be characterized by increasing extreme wave conditions over time. Whether this methodology can be applied to successfully project future extreme H_{m0} events in light of a changing climate is debatable. The non-stationarity of the extreme wave climate and the underlying cause of this non-stationarity are not properly understood. Without a better understanding of the non-stationarity and without future projections of the variability of extreme H_{m0}, the applicability of clustering and subsequent non-stationary extreme value modelling at cluster level in light of climate change is limited.

Bibliography

Allen, S., Barros, V., Burton, I., Campbell-Lendrem, D., Cardona, O., Cutter, S., Dube, O. P., Ebi, K., Field, C., Handmer, J., Lal, P., Lavell, A., Mach, K., Mastrandrea, M., Mcbean, G., Mechler, R., Mitchell, T., Nicholls, N., O Brien, K., & Wilbanks, T. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://doi.org/10.13140/2.1.3117.9529

Antonini, A., Raby, A., Brownjohn, J. M. W., Pappas, A., & D'Ayala, D. (2019). Survivability assessment of fastnet lighthouse. *Coastal Engineering*, *150*, 18–38. https://doi.org/10.1016/j.coastaleng.2019.03.007

Arns, A., Wahl, T., Haigh, I. D., Jensen, J., & Pattiaratchi, C. (2013). Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, *81*, 51–66. https://doi.org/10.1016/j.coastaleng.2013.07.003

Baldwin, M. P. (2001). Annular modes in global daily surface pressure. *Geophysical Research Letters*, *28*(21), 4115–4118. https://doi.org/https://doi.org/10.1029/2001GL013564

Bickel, P. J., & Lehmann, E. L. (2012). Frequentist Interpretation of Probability. *Selected works of e. l. lehmann* (pp. 1083–1085). Springer US. https://doi.org/10.1007/978-1-4614-1412-4{_}92

Bonaduce, A., Staneva, J., Behrens, A., Bidlot, J.-R., & Wilcke, R. A. I. (2019). Wave Climate Change in the North Sea and Baltic Sea. *Journal of Marine Science and Engineering*, *7*(6), 166. https://doi.org/10.3390/jmse7060166

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Camus, P., Cofiño, A. S., Mendez, F. J., & Medina, R. (2011). Multivariate wave climate using self-organizing maps. *Journal of Atmospheric and Oceanic Technology*, *28*(11), 1554–1568. https://doi.org/10.1175/JTECH-D-11-00027.1

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, *127*(2), 353–369. https://doi.org/10.1007/s10584-014-1254-5

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Cunnane, C. (1979). A note on the Poisson assumption in partial duration series models. *Water Resources Research*, *15*(2), 489–494. https://doi.org/https://doi.org/10.1029/WR015i002p00489

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

De Leo, F., Besio, G., Briganti, R., & Vanem, E. (2021). Non-stationary extreme value analysis of sea states based on linear trends. Analysis of annual maxima series of significant wave height and peak period in the Mediterannean Sea. *Coastal Engineering*, *167*. https://doi.org/10.1016/j.coastaleng.2021.103896

De Leo, F., De Leo, A., Besio, G., & Briganti, R. (2020). Detection and quantification of trends in time series of significant wave heights: An application in the Mediterranean Sea. *Ocean Engineering*, *202*. https://doi.org/10.1016/j.oceaneng.2020.107155

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer. https://books.google.nl/books?id=XLUMIlombgQC

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470977811

Fern Tay, H. (2021). A Little-Known Trick in Hierarchical Clustering: Weights.

Ferro, C. A. T., & Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B : Statistical Methodology*, *65*(2), 545–556. https://doi.org/10.1111/1467-9868.00401

Freitas, A., Bernardino, M., & Guedes Soares, C. (2022). The influence of the Arctic Oscillation on North Atlantic wind and wave climate by the end of the 21st century. *Ocean Engineering*, *246*, 110634. https://doi.org/10.1016/j.oceaneng.2022.110634

Goharnejad, H., Perrie, W., Toulany, B., Casey, M., & Zhang, M. (2022). Clustering of Climate Change Impacts on Ocean Waves in the Northwest Atlantic. *Journal of Atmospheric and Oceanic Technology*, *39*(2), 237–257. https://doi.org/10.1175/JTECH-D-21-0053.1

Grabemann, I., Groll, N., Möller, J., & Weisse, R. (2015). Climate change impact on North Sea wave conditions: a consistent analysis of ten projections. *Ocean Dynamics*, *65*(2), 255–267. https://doi.org/10.1007/s10236-014-0800-z

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press.

Huthnance, J. M. (1989). Physical Oceanography of the North Sea. *Ocean and Shorline Management*, *16*, 199–231.

Jiang, Z., Hu, W., Dong, W., Gao, Z., & Ren, Z. (2017). Structural reliability analysis of wind turbines: A review. https://doi.org/10.3390/en10122099

Jonkman, S. N., Jorissen, R. E., Schweckendiek, T., & Van Den Bos, J. P. (2018). *Flood defences Lecture notes CIE5314 3rd edition 2018* (tech. rep.). Delft University of Technology. Delft.

Karmpadakis, I., Swan, C., & Christou, M. (2020). Assessment of wave height distributions using an extensive field database. *Coastal Engineering*, *157*. https://doi.org/10.1016/j.coastaleng.2019.103630

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*, 1287–1304. www.elsevier.com/locate/advwatres

Kaufman, L., & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis* (L. Kaufman & P. J. Rousseeuw, Eds.). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316801

Kendall, M. G. (1955). *Rank Correlation Methods*. C. Griffin. https://books.google.nl/books?id=TJAuAAAAMAAJ

KNMI. (2021). *KNMI Klimaatsignaal '21 - Hoe het klimaat in Nederland snel verandert* (tech. rep.). Koninglijk Nederlands Meteorologisch Instituut. De Bilt.

Laface, V., Malara, G., Romolo, A., & Arena, F. (2016). Peak over threshold vis-à-vis equivalent triangular storm: Return value sensitivity to storm threshold. *Coastal Engineering*, *116*, 220–235. https://doi.org/10.1016/J.COASTALENG.2016.06.009

Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*, 103–117. www.elsevier.com/locate/jhydrol

Liang, B., Shao, Z., Li, H., Shao, M., & Lee, D. (2019). An automated threshold selection method based on the characteristic of extrapolated significant wave heights. *Coastal Engineering*, *144*, 22–32. https://doi.org/10.1016/J.COASTALENG.2018.12.001

Lowe, R. J., Cuttler, M. V., & Hansen, J. E. (2021). Climatic Drivers of Extreme Sea Level Events Along the Coastline of Western Australia. *Earth's Future*, *9*(4). https://doi.org/10.1029/2020EF001620

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*(7), 5469–5494. https://doi.org/10.1002/2016WR019676

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., & Wagenmakers, E.-J. (2017). A Tutorial on Fisher information. *Journal of Mathematical Psychology*, *80*, 40–55. https://doi.org/10.1016/j.jmp.2017.05.006

Maimom, O., & Rockach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer US. https://doi.org/10.1007/978-0-387-09823-4

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometria*, *13*(3), 245–259. https://www.jstor.org/stable/1907187

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Pean, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekci, O., Yu, R., & Zhou, B. (2021). *Climate change 2021: The Physical Science Basis. Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (tech. rep.). Cambridge University Press. Cambridge.

Matalas, N. C. (1997). Stochastic Hydrology in the Context of Climate Change. *Climatic Change*, *37*(1), 89–101. https://doi.org/10.1023/A:1005374000318

McCabe, G. J., Clark, M. P., & Serreze, M. C. (2001). Trends in Northern Hemisphere Surface Cyclone Frequency and Intensity. *Journal of Climate*, *14*(12), 2763–2768. https://doi.org/10.1175/1520-0442(2001)014<2763:TINHSC>2.0.CO;2

Méndez, F. J., Menéndez, M., Luceño, A., & Losada, I. J. (2006). Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research: Oceans*, *111*(7). https://doi.org/10.1029/2005JC003344

Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, *6*(17), 3299–3303. https://doi.org/10.19026/rjaset.6.3638

Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards: Stationarity is immortal! https://doi.org/10.1002/2014WR016092

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *WIREs Data Mining and Knowledge Discovery*, *2*(1), 86–97. https://doi.org/10.1002/widm.53

NOAA. (2022). North Atlantic Oscillation (NAO).

Pickands, J. (1975). *Statistical Inference Using Extreme Order Statistics* (tech. rep. No. 1).

Raby, A. C., Antonini, A., Pappas, A., Dassanayake, D., Brownjohn, J., & D'ayala, D. (2019). Wolf Rock lighthouse: past developments and future survivability under wave loading. *Philisophical Transactions of the Royal Society A*, *377*(2155), 1–27. https://doi.org/10.1098/not

Ragno, E., AghaKouchak, A., Cheng, L., & Sadegh, M. (2019). A generalized framework for process-informed nonstationary extreme value analysis. *Advances in Water Resources*, *130*, 270–282. https://doi.org/10.1016/j.advwatres.2019.06.007

Ragno, E., AghaKouchak, A., Love, C. A., Cheng, L., Vahedifard, F., & Lima, C. H. (2018). Quantifying Changes in Future Intensity-Duration-Frequency Curves Using Multimodel Ensemble Simulations. *Water Resources Research*, *54*(3), 1751–1764. https://doi.org/10.1002/2017WR021975

Rickets, L. (2021). Maximum Likelihood vs. Bayesian Estimation.

Roscoe, K., Caires, S., Diermanse, F., & Groeneweg, J. (2010). Extreme offshore wave statistics in the North Sea. *WIT Transactions on Ecology and the Environment*, *133*, 47–58. https://doi.org/10.2495/FRIAR100051

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sadegh, M., Ragno, E., & AghaKouchak, A. (2017). Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research*, *53*(6), 5166–5183. https://doi.org/10.1002/2016WR020242

Sartini, L., Mentaschi, L., & Besio, G. (2015). Comparing different extreme wave analysis models for wave climate assessment along the Italian coast. *Coastal Engineering*, *100*, 37–47. https://doi.org/10.1016/j.coastaleng.2015.03.006

Sartini, L., Weiss, J., Prevosto, M., Bulteau, T., Rohmer, J., & Maisondieu, C. (2018). Spatial analysis of extreme sea states affecting Atlantic France: a critical assessment of the RFA approach. *Ocean Modelling*, *130*, 48–65. https://doi.org/10.1016/j.ocemod.2018.07.008

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas science bulletin*, *38*, 1409–1438.

Sokal, R. R., & Rohlf, F. J. (1962). THE COMPARISON OF DENDROGRAMS BY OBJECTIVE METHODS. *TAXON*, *11*(2), 33–40. https://doi.org/https://doi.org/10.2307/1217208

Suarez-Alvarez, M. M., Pham, D. T., Prostov, M. Y., & Prostov, Y. I. (2012). Statistical approach to normalization of feature vectors and clustering of mixed datasets. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *468*(2145), 2630–2651. https://doi.org/10.1098/rspa.2011.0704

Theil, H. (1992). *A Rank-Invariant Method of Linear and Polynomial Regression Analysis*. Springer Netherlands. https://doi.org/10.1007/978-94-011-2546-8{_}20

Timmermans, M. (2021). *A PRACTICAL ASSESSMENT OF THE IMPACT OF USING MULTIVARIATE STATISTICAL MODELS IN THE DESIGN OF COASTAL INFRASTRUCTURE* (tech. rep.). Delft University of Technology. Delft.

Van Den Bos, J. P., & Verhagen, H. J. (2018). *Breakwater design Lecture notes CIE5308* (tech. rep.). Delft University of Technology. Delft.

Van Der Kooij, L. (2020). *Extreme value analysis of complex wave systems* (tech. rep.). Delft University of Technology. Delft.

Vergouwe, R. (2014). *The National Flood Risk Analysis for the Netherlands* (tech. rep.). Rijkswaterstaat. The Hague.

Viselli, A. M., Forristall, G. Z., Pearce, B. R., & Dagher, H. J. (2015). Estimation of extreme wave and wind design parameters for offshore wind turbines in the Gulf of Maine using a POT method. *Ocean Engineering*, *104*, 649–658. https://doi.org/10.1016/j.oceaneng.2015.04.086

Voorendt, M. Z., & Molenaar, W. F. (2019). *Manual Hydraulic Structures* (tech. rep.).

Wang, J., Liu, J., Wang, Y., Liao, Z., & Sun, P. (2021). Spatiotemporal variations and extreme value analysis of significant wave height in the South China Sea based on 71-year long ERA5 wave reanalysis. *Applied Ocean Research*, *113*. https://doi.org/10.1016/j.apor.2021.102750

Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Weiss, J., Bernardara, P., & Benoit, M. (2014). Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights. *Journal of Geophysical Research: Oceans*, *119*(5), 2906–2922. https://doi.org/10.1002/2013JC009668

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

Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, Junjie Wu, & Sen Wu. (2013). Understanding and Enhancement of Internal Clustering Validation Measures. *IEEE Transactions on Cybernetics*, *43*(3), 982–994. https://doi.org/10.1109/TSMCB.2012.2220543

A

Principles of Ocean Waves

Extreme sea states refer to wind-generated gravity waves (hereafter wind waves) on the water surface. Wind waves result from the wind blowing over the water surface, where friction between the wind and water in combination with the uneven distribution of the wind field results in ripples (Holthuijsen, 2007). Continuous wind over a long fetch causes wave-induced wind-pressure variations. The effect of this is that the wind-pressure is maximum on the windward side of the wave and a minimum on the leeward side of the wave. This causes the wave to be "pushed down" on the windward side, and is "pulled" upward on the leeward side. This coupling between wind pressure and wave motion transfers energy into the waves and the initially small ripples grow into waves (Holthuijsen, 2007).

Ocean waves are typically subdivided in wind generated waves and swell waves. Swell waves (often referred to as surface gravity waves) propagate along the water surface under the predominating influence of gravity. Swell waves have their origin as wind waves in distant wind fields. But due to having a longer wave period and subsequently a longer wave length than other waves generated in the wind field, swell waves have higher traveling velocity resulting in wave dispersion (Holthuijsen, 2007). As a consequence, these swell waves travel faster, which results in wave field consisting of waves with long periods and low steepness.

To complicate matters further, a sea state typically consists of a combination of wind wave and swell systems, sometimes coming from different directions and with different spectral shapes (Van Der Kooij, 2020). In the North Sea, waves generated by a local wind field prevail (Weisse et al., 2012), but these waves may be superimposed by swell waves. It is therefore critical to analyse the wave climate and determine the wave classes present in the system. A typical solution is to study the wave spectra (see section A), where wind waves occur for lower frequencies than swell waves (Holthuijsen, 2007). Another approach is by plotting H/T diagram to study the relationship between the significant wave height and the peak wave period. The steepness of the waves s can be calculated by $s = H/L_p$ where L_p is the wavelength that can be calculated using the peak wave period (Van Den Bos & Verhagen, 2018). Wave steepness values rarely exceed $s = 0.05$. These high values for wave steepness are often related to locally generated wind waves. Low values for wave steepness (values for s approaching 0.01) are related to swell waves from remote wave fields. Values lower than 0.01 are not often reported (Van Den Bos & Verhagen, 2018).

Spectral Wave Analysis

In practice, the behavior of waves is often described using the two-dimensional (2D) energy wave spectrum $E(f, \theta)$ (Holthuijsen, 2007). Typically, the wave energy spectrum is expressed as the product of its one-dimensional energy spectrum, $S(f)$, and a Directional Spreading Function $D(f, \theta)$ as follows (Holthuijsen, 2007):

$$E(f,\theta) = S(f) \cdot D(f,\theta) = E(f,\theta) = \lim_{\Delta f \to 0} \lim_{\Delta \theta \to 0} \frac{1}{\Delta f \Delta \theta} E\left\{\frac{1}{2}\underline{a}^2\right\} \tag{A.1}$$

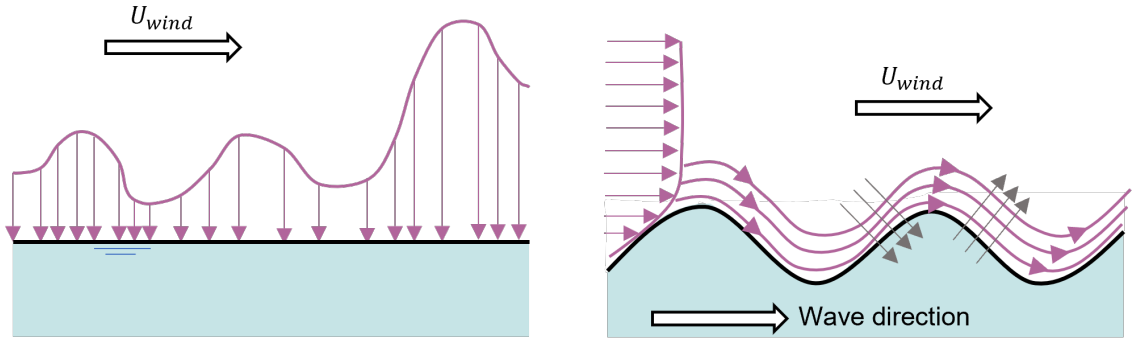


Figure A.1: Principles behind wave generation. Left figures shows Philips mechanism where disturbances in the wind field cause ripples in the water surface. Right figure shows Miles mechanisms for the further generation of waves (adopted from Holthuijsen (2007).

The wave energy spectrum can now be defined as the energy for waves traveling in a particular directional bandwidth as follows:

$$S(f) = \int E(f,\theta)d\theta \tag{A.2}$$

The shape of the resulting energy curve has been subjective research in coastal and oceanic engineering. The shape of the spectrum is the result of the forcing mechanism causing wave generation. The most common spectral shape in engineering practice in the North Sea is the JONSWAP shape ((Holthuijsen, 2007). JONSWAP has been established after careful analysis of observed wave conditions in the North Sea. However, even in the North Sea, JONSWAP is not even the most prevailing spectral shape in all regions of the North Sea domain (Karmpadakis et al., 2020). Integrating the energy density function (Equation A.2) over the whole interval of f leads by definition to the total energy contained in the wave field of the considered wave propagation direction.

$$E_{tot} = \int_0^\infty S\,(f)\,df \tag{A.3}$$

The information obtained by the energy spectrum can be used to derive characteristic wave parameters corresponding to the sea state. The properties of the wave parameters are provided by their spectral moment. To this end, Equation A.3 can be generalized into the following equation:

$$m_n = \iint f^n E(f,\theta)df d\theta \tag{A.4}$$

Wave Height and Period

Using the spectral moments obtained using Equation A.4, we can now determine the relevant wave parameters for this research. The significant wave height as derived from the wave energy spectrum is defined as follows:

$$H_{m0} = 4 \cdot \sqrt{m_0} \tag{A.5}$$

Here, m_0 is the zeroth spectral moment. This definition must not be confused with other definitions for the significant wave height, such as the mean of the highest one-third of the wave height (H_s). In deep waters, both H_{m0} and H_s should yield the same results, but in intermediate and shallow water depth, the resulting values for significant wave height could be very different (Van Den Bos & Verhagen, 2018). Apart from the significant wave height, the wave periods form an important parameter in the design and risk assessment of civil infrastructure. The mean wave period employed in this research is the energy mean wave period and is defined as follows:

$$T_{m-1,0} = \frac{m_0}{m_1} \tag{A.6}$$

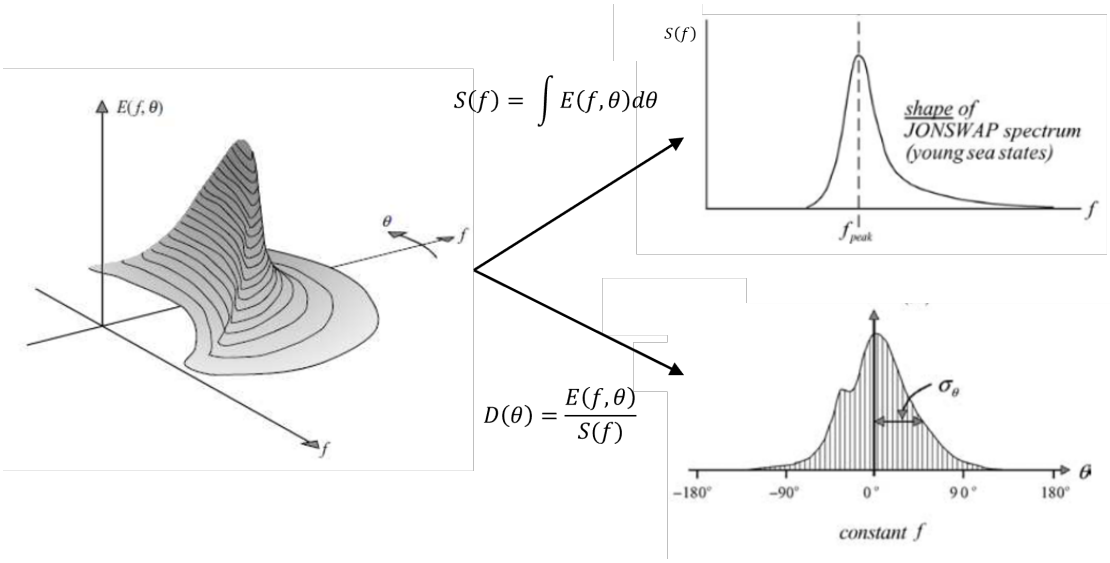


Figure A.2: Overview of spectral wave analysis and the required transformations to arrive at the wave energy spectrum and the directional spectrum. Left figure shows the two dimensional wave spectrum. The top figure on the right presents the wave energy spectrum and the bottom right figures shows the directional spectrum. Figures adopted from Holthuijsen (2007).

Research has shown that the mean wave energy period is a more robust value to describe the long-term stability of breakwaters and flood defense structures (Van Den Bos & Verhagen, 2018), and is therefore often used in design formulas for breakwaters, dikes and dunes. The peak wave period is the period of the wave that contains the most energy (Antonini et al., 2019) and is therefore defined by the peak of the wave spectrum $S(f)$. In mathematical terms, the peak period is expressed as:

$$T_P = \frac{1}{f_p} \tag{A.7}$$

Wave Direction

Apart from the wave height and period, the wave direction must be considered. Especially coastal engineering structures, the wave direction is important as it determines the angle of incoming waves. The directional spectrum of waves is generally described by the following equation:

$$D\left(\theta\right) = \frac{E\left(f,\theta\right)}{S(f)} \tag{A.8}$$

In general terms, this expression can be described as the ratio of the total energy and the energy in a particular direction. The directional spectrum can be visualized by a spectral shape that gives the wave energy in a particular direction theta (see Figure A.2). The mean wave direction is the mean over all frequencies and directions of the two-dimensional wave spectrum (Holthuijsen, 2007). Apart from the mean wave direction, the directional spreading is an important parameter, as it describes the uniformity of the wave field. However, the directional spreading of the waves is not considered.

Nearshore Waves

Apart from the wind speed, duration and fetch, the local water depth is important for the development and evolution of waves. For the analysis of the extreme wave climate in the design and risk assessment of hydraulic infrastructure, the offshore wave climate is preferred. Offshore, deep water conditions typically prevail and wave fields develop and evolve without bottom interaction, thereby avoiding shallow water and local bathymetry effects. Wave data in nearshore locations are considered to be less reliable. The bathymetry can have significant effects on the wave climate, which can result in shoaling, refraction and in very shallow conditions even wave breaking. Moreover, the presence of coastal structures potentially influences the nearshore wave climate as coastal structure result in reflecting waves. Apart from the local bathymetry affecting the wave climate, waves in turn can also have an effect on the

local bathymetry. This causes the wave climate to be inconsistent over time. For these reasons, only offshore wave information is used in the design and risk assessment of hydraulic structures, and this research only considers waves that are at least in intermediate water depth.

B

Example of Hierarchical clustering

To illustrate how agglomerative hierarchical clustering works, this chapter provides a simple example. Consider a dataset consisting of two different variables (X and Y) at 6 locations. We are interested in clustering the locations based on the observations for X and Y. We know the values for X and Y at each of the six locations. We can then plot the values for X and Y on a scatter plot (see figure B.1).

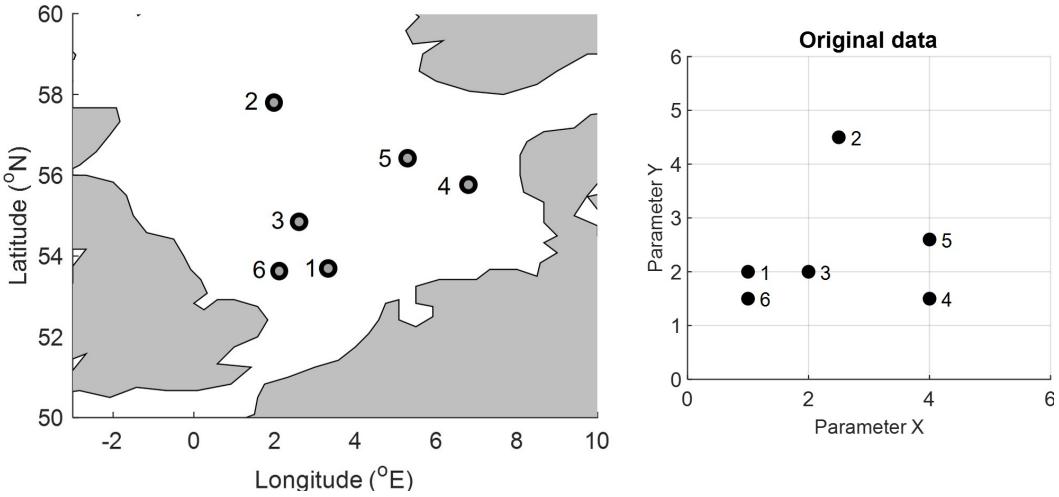


Figure B.1: Example data used in this chapter to present Hierarchical clustering.

The first step is to standardize the input values. Without normalization, greater importance is given to parameters with greater magnitude. Standardization of the input values can be achieved by means of z-score standardization:

$$X_i^* = \frac{X_i - \mu_X}{\sigma_X} \tag{B.1}$$

Because the input values are standardized, it is possible to assign weights to the different parameters. Weight may be assigned to emphasize particular parameters in the clustering that are thought more significant or to add some additional steering to obtain the desired clustering output. In this example, equal weight for both parameters is considered.

The next step is to calculate the pairwise dissimilarity between the grid points. For a dataset containing N grid points, there exist $N*(N-1)/2$ pairs for which the dissimilarity can be calculated. The Euclidean distance between two grid points (i and j) can be calculated as follows:

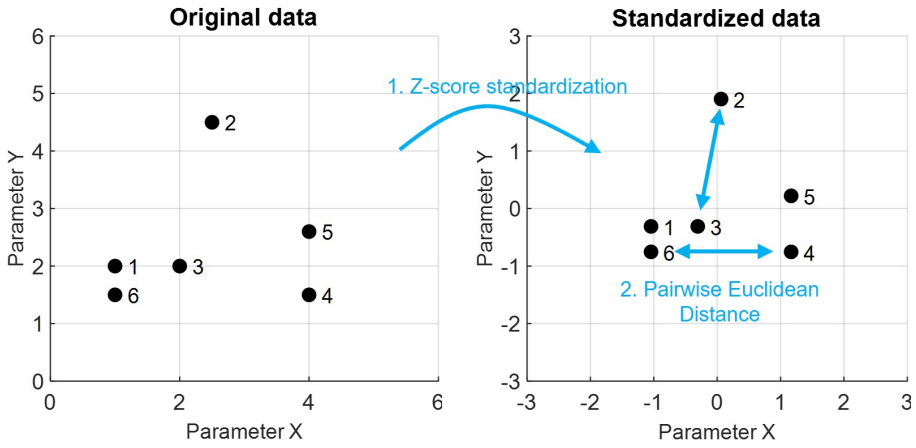


Figure B.2: Z-score standardization (see Equation 3.8) of example dataset.

$$d(i,j) = \|i - j\|_2 = \sqrt{\sum_{k=1}^{M} (i_k - j_k)^2} \tag{B.2}$$

Once the dissimilarity between all grid points in the dataset has been computed, it must be determined how grid points should be grouped into clusters. This can be done according several linkage mechanisms. The linkage function links pairs of objects that are closest together into binary clusters (clusters made up of two objects). Subsequently, the linkage function links the newly formed clusters to create bigger clusters until all grid points are linked together in a hierarchical tree. In this thesis, Single, Complete, Average and Ward linkage mechanisms are considered. A detailed overview of the different linkage mechanisms is presented in Section 3.3.3.

The hierarchical linkage is best understood when viewed graphically using dendrograms. In a dendrogram, the horizontal axis represents the objects included in HAC. The links between objects are represented by upside-down U-shaped lines. The height of the U-shapes represents the dissimilarity between the objects (grid points or clusters) as calculated by the selected linkage mechanism (T). This is also referred to as the cophenetic dissimilarity (Sokal & Rohlf, 1962). The resulting dendrograms for the four linkage mechanisms considered in this research have been plotted in Figure B.4.

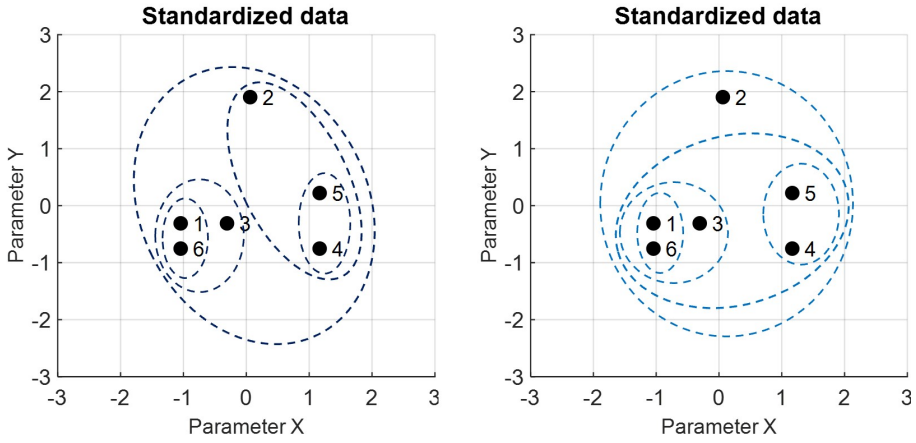


Figure B.3: Established clusters for our example data showing the hierarchy of linkage. Left figure shows the identified clusters for Complete, Single and Average Linkage. Right figure shows identified clusters for Ward linkage.

It must be verified that the cophenetic dissimilarities in the dendrogram accurately reflect the dissimilarities in the original dataset. One measure to study the performance of the linkage process is to compare the cophenetic dissimilarities (T) between grid points with the original pairwise dissimilarities

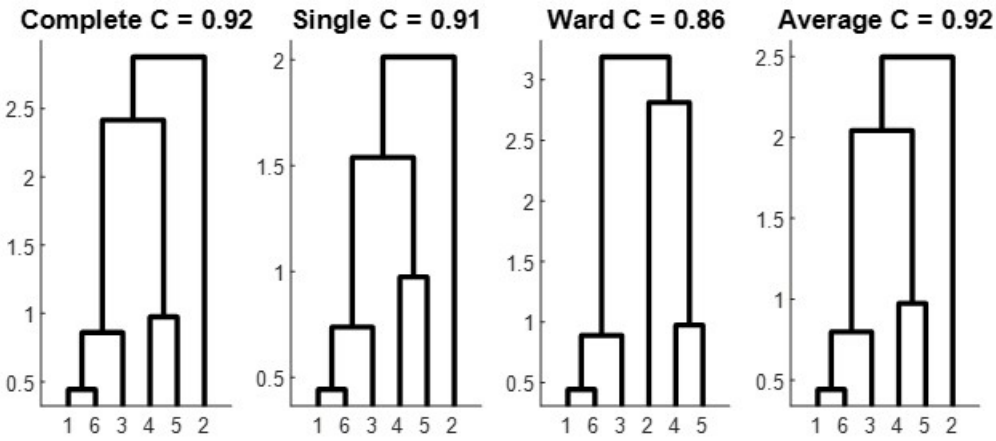


Figure B.4: Examples of the established dendrograms for the example dataset.

(d), using the cophenetic correlation coefficient (C, see Equation 3.16). The cophenetic correlation can be calculated using the following equation:

$$C = \frac{\sum_{i<j} \left(d(i,j) - \overline{d}\right) \left(T(i,j) - \overline{T}\right)}{\sqrt{\sum_{i<j} \left(d(i,j) - \overline{d}\right)^2 \sum_{i<j} \left(T(i,j) - \overline{T}\right)^2}} \tag{B.3}$$

In equation 3.16, $d_{i,j}$ is the pairwise Euclidean distance between grid points, and $\overline{d}$ is the mean Euclidean distance for all Euclidean distances. Let $T(i,j)$ be the cophenetic dissimilarity between objects i and j and $\overline{T}$ be the average cophenetic distance of all cophenetic correlations. The closer the value of the cophenetic correlation coefficient (C) is to 1, the more accurately the clustering solution reflects your data. The cophenetic correlation can be used to select the linkage mechanism that yields the best representation of the original dataset. This linkage mechanism will be selected for further analysis. For the presented example, this is Complete linkage.

The final step in HAC is to select the appropriate cut-off points that determines the number of clusters ($\mathcal{K}$) that adequately describes the extreme wave climate in the North Sea. In this research, three internal evaluation metrics are used to find the optimal value for $\mathcal{K}$. This includes the Silhouette Coefficient (SC), the Calinski-Harabasz (CH) score and the Davies-Bouldin (DB) ratio. A complete description of the three internal evaluation metrics can be found in section 3.3.4, and is not presented here. For the three internal evaluation metrics, the following rules apply:

- The optimal number of clusters as indicated by the Silhouette Coefficient (SC) is the value for $\mathcal{K}$ that yields the highest value for SC;
- The optimal number of clusters as indicated by the Calinski-Harabasz (CH) score is the value for $\mathcal{K}$ that yields the highest value for CH;
- The optimal number of clusters as indicated by the Davies-Bouldin (DB) ratio is the value for $\mathcal{K}$ that yields the lowest value for DB;

Based on the scores for SC, CH and DB, the ideal number of clusters to partition our example dataset must be determined. For CH and DB, this is relatively straightforward, and the optimal score for both metrics yields $\mathcal{K} = 5$ clusters. For SC, it can be observed that the ideal number of clusters would be $\mathcal{K} = 4$. However, it can also be observed that $\mathcal{K} = 4$ and $\mathcal{K} = 5$, yield an approximately equal score for SC. Therefore, it is decided to partition our example dataset into 5 clusters. The final cluster solution can be observed in Figure B.6.

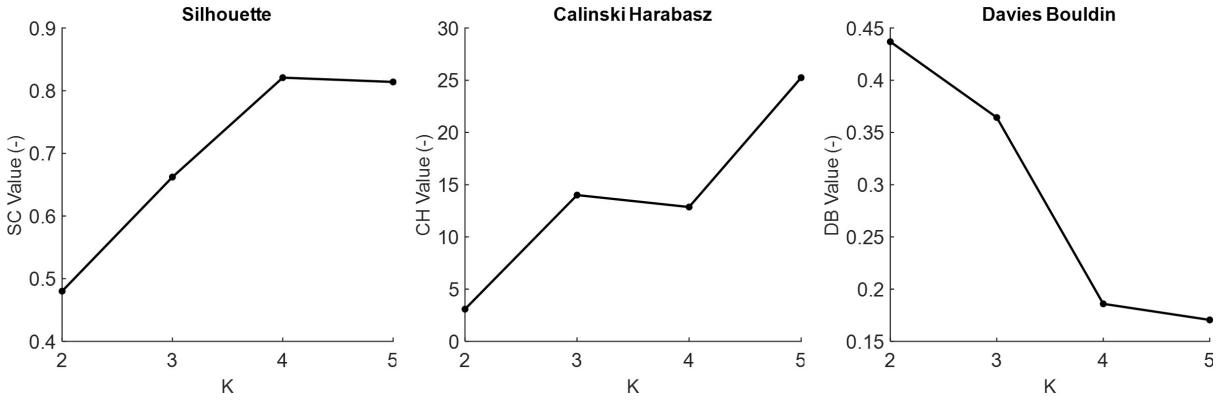


Figure B.5: Plots for Silhouette coefficient (SC), Calinski-Harabasz score (CH) and Davies-Bouldin ratio (DB) for our example data.

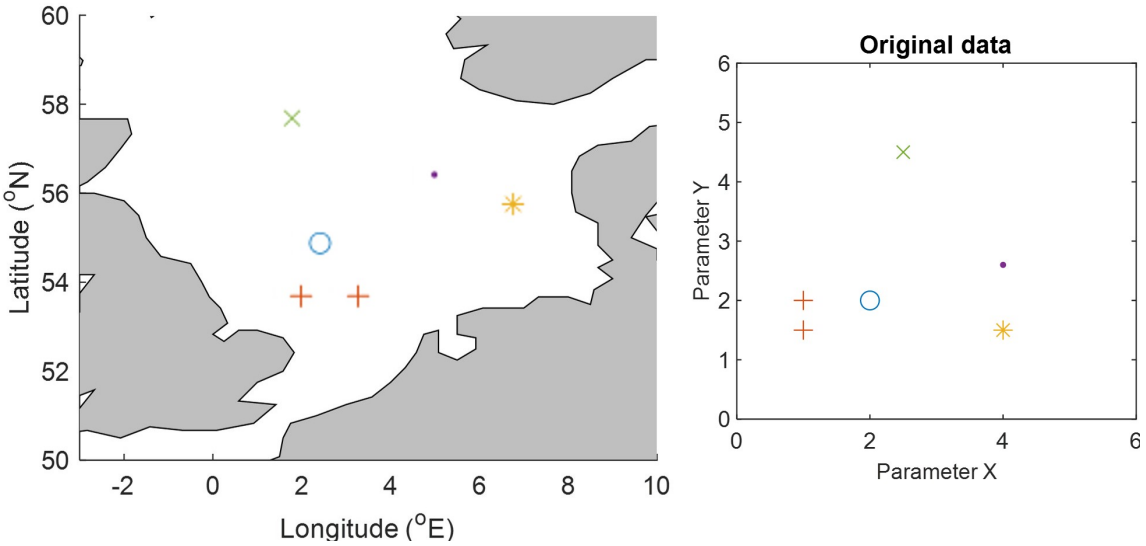


Figure B.6: Final Cluster solution of our example dataset. The selected linkage mechanism is Average linkage based on C. The scores for SC, CH and DB indicate the $\mathcal{K} = 5$ is the optimal number of clusters.

C

Supplementary Material Markov Chain Monte Carlo

ProNEVA applies a hybrid-evolution Markov Chain Monte Carlo (MCMC) approach to calculate the posterior distribution parameters (Ragno et al., 2019). MCMC algorithms are used to draw random samples from high dimensional, complex distributions (Sadegh et al., 2017). The goal of MCMC is to reach an equilibrium state that represents the target distribution (i.e. the posterior). MCMC converges closer to the true distribution as more samples are included in the MCMC. Three important components make up the MCMC algorithm (see also figure C.1):

1. Monte Carlo: a Method to generate random sample from a distribution. More samples are expected in an area with higher probability of occurrence, and less samples are expected in areas with a lower probability of occurrence;
2. Markov Chains: Sequence (Chain) of samples where each sample (i) is only dependent on the previous sample ($i-1$) in the sequence. This results in a "random walk" through the parameter space;
3. Metropolis Hastings: Acceptance algorithm to determine whether the next sample in sequence (θ_i) are to be rejected or not.

A general outline of the MCMC process is presented in Figure C.1. The MCMC process starts by drawing an initial random sample for θ, i.e. θ_0. For θ_0, the prior and likelihood probabilities can be calculated, and subsequently, the posterior probability. The next step is to randomly draw the next sample, i.e. θ_1. It is important that θ_1 lies in the vicinity of θ_0. Also, for θ_1, the prior and likelihood probabilities and subsequently, the posterior probability are calculated. Now, it must be tested whether to accept or reject θ_1. We only want to accept θ_1 if it is better than θ_0. If θ_1 is accepted, it gets added to the posterior distribution, otherwise, θ_0 is added to the posterior distribution again. The final value (θ_0 or θ_1) is then used to find a new proposal (i.e. θ_2) in the next iteration. The steps above are repeated for all N iterations of the MCMC.

The Metropolis Hastings ratio is applied to decide whether to accept or reject the value for θ_i. Considering the posterior probabilities θ_i and θ_{i-1} in which θ_{i-1} stands for the previous iteration, the acceptance ratio is defined as follows:

$$AR = \frac{P\left(\theta_i \mid Y_{obs}\right)}{P\left(\theta_{i-1} \mid Y_{obs}\right)} = \frac{P\left(Y_{obs} \mid \theta_i\right)}{P(Y_{obs} \mid \theta_{i-1})} \frac{P\left(\theta_i\right)}{P(\theta_{i-1})} \tag{C.1}$$

or each iteration of the MCMC, the acceptance ratio is tested according to:

$$\rho = min\left(1, AR\right) \tag{C.2}$$

If $\rho = 1$ (i.e. the acceptance ratio is larger than 1), the proposed θ_i has a higher posterior probability than θ_{i-1}, and θ_i is accepted. If $\rho < 1$, θ_i is not automatically rejected. Instead, a random value r is

drawn from a uniform distribution ($r \sim \mathcal{U}(0,1)$). If $r \leq \rho$, θ_i is accepted, otherwise it is discarded. The final value for theta (θ_i or θ_{i-1}) is used to find a new proposal for θ in the next iteration.

MCMCs are initialized randomly and must converge towards the correct value. This often requires many samples. When plotting our results and posterior distribution, it is not effective to include these early samples before the model has converged. A so-called Burn-In period is implemented that excludes the early samples for theta in the final results.

ProNeva applies a hybrid-evolution MCMC (HE-MCMC) that was proposed by Sadegh et al. (2017). Although a detailed mathematical explanation of HE-MCMC lies outside the scope of this research, the most important concepts are quickly explained below. For more detailed explanation of HE-MCMC algorithm, the reader is referred to Ragno et al. (2019) and Sadegh et al. (2017):

- Intelligent starting point: Rather than starting with one initial sample for theta, HE-MCMC draws samples for the entire prior space, that are randomly distributed over the chains. Within each chain, the sample with the highest $p\left(Y_{obs} \mid \theta\right)$ is elected as a starting point. This means that each chain has a starting point with the highest chance of success and increases the convergence rate of the MCMC process.
- Multiple Chains and Differential Evolution: Multiple chains are run in parallel in order to improve the accuracy of the proposals. The chains learn from each other rather than running all the chains independently. Differential evolution solves an important problem that random sampling in MCMC poses, namely the scale problem. The moving distance between successive proposals can be controlled so that the HE-MCMC searches the parameter domain more efficiently.
- Snooker update: An Adaptive Direction Sampling technique (ADS). The prior samples are utilized to define the sampling direction for the subsequent iterations, with the sampling direction tending to go over areas of higher density. This enhances the likelihood of discovering more successful proposals, resulting in faster convergence and a more accurate posterior distribution.
- Adaptive Metropolis (AM): The general idea is to generate proposal samples with already obtained knowledge about the target distribution.
- Gelman Rubin: A diagnostic test to assess the convergence of the chains. The Gelman-Rubin statistic analyzes the difference between the multiple Markov chains. The convergence is assessed by comparing the estimated inter-chains and intra-chain variances for each modelled theta. Large differences between these variances indicate non-convergence.

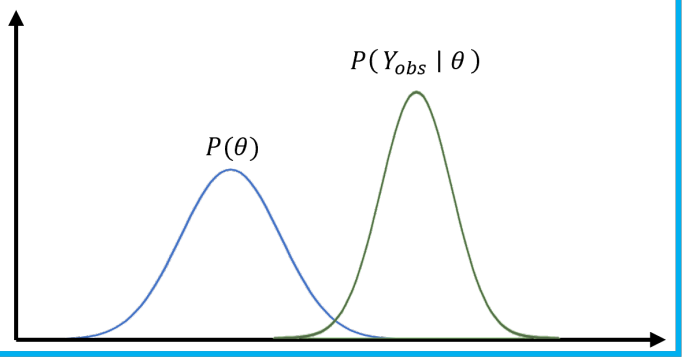
Input for Bayesian Inference and MCMC

1. Prior information

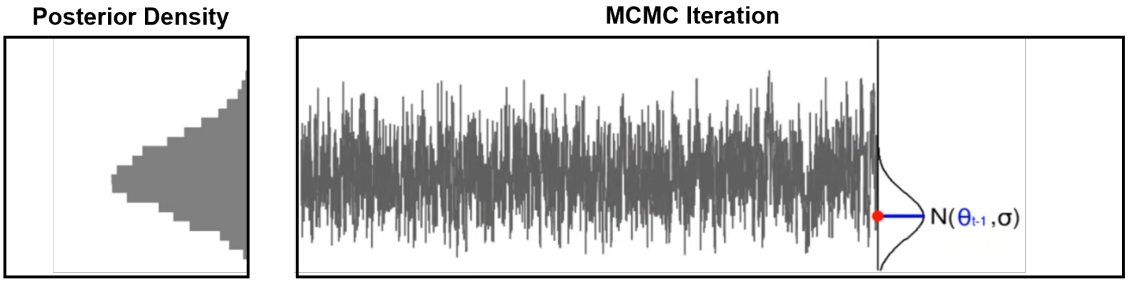
$$\theta = (\mu_0, \mu_1, \sigma, \xi)$$
$$P(\theta)$$

2. Set of observations at N time points

$$Y_{obs} = y_1, \dots, y_N$$
$$P(Y_{obs} \mid \theta) = \prod_{i=1}^{N} P(y_i \mid \theta)$$



Bayesian Inference Markov Chain Monte Carlo



Step 1: Draw Sample θ_i in the vicinity of θ_{i-1} according to $\mathrm{N}(\theta_{i-1}, \sigma)$

Step 2: Calculate prior, likelihood and posterior probabilities for θ_i and θ_{i-1}

$$P(\theta \mid Y_{obs}) \propto P(\theta) \cdot P(Y_{obs} \mid \theta) = P(\theta) \cdot \prod_{i=1}^{N} P(y_i \mid \theta)$$

Step 3: Calculate the acceptance ratio AR and testing criteria ρ

$$\rho = \min(1, AR) = \min(1, \frac{P(Y_{obs} \mid \theta_i)}{P(Y_{obs} \mid \theta_{i-1})})$$

Step 4: Accept or reject θ_i based on the following criteria:

- If $\rho = 1$, accept θ_i
- If $\rho < 1$, draw a random value $r \sim U(0,1)$, if $\mathrm{r} \leq \rho$, accept θ_i , if $r > \rho$, reject θ_i, and use θ_{i-1} instead.

Step 5: Repeat step 1 to 4 N iterations

Output:

Updated beliefs about the distribution parameters of extreme value distributions

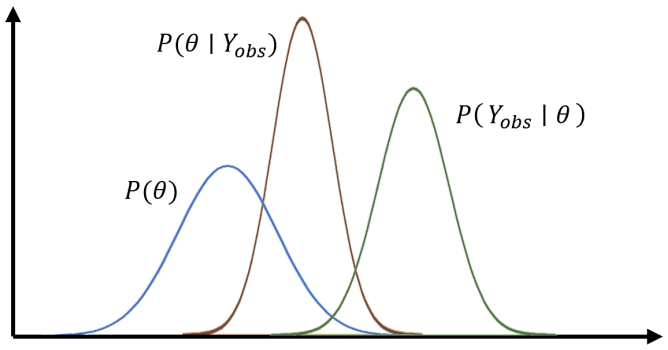


Figure C.1: Schematic overview of Bayesian Inference and Markov Chain Monte Carlo to update beliefs on the parameter distribution when new evidence is presented.

D

Supplementary Material North Sea Wave Climate

The ERA5 reanalysis dataset used in this research is the fifth generation of reanalysis datasets produced by the ECMWF and contains a highly detailed record of the global atmosphere, land surface and ocean waves spanning between 1950 and the near present (Hersbach et al., 2020). The ERA5 performance is evaluated and validated against in-situ observation data (provided by KNMI, Rijkswaterstaat) using several goodness-of-fit tests. This appendix describes the evaluation and validation of the H_{m0} time-series from the ERA5 Reanalysis used in this research.

Quality Control and Performance of ERA5 Datasets

For the validation and performance test of the ERA5 reanalysis data, a total of three goodness-of-fit metrics is applied. The linear correlation coefficient R is employed to measure the linear correlation between the reanalysis data and observations. It is the ratio between the covariance of the reanalysis data and the observation data and the product of their standard deviations and can be expressed using the following equation:

$$R = \frac{\sum_{i=1}^{N} \left[\left(C_i - \bar{C} \right) \left(O_i - \bar{O} \right) \right]}{\sqrt{\sum_{i=1}^{N} \left(C_i - \bar{C} \right)^2 \sum_{i=1}^{N} \left(O_i - \bar{O} \right)^2}} \tag{D.1}$$

Mean Bias Error (MBE, here referred to as $bias$) is the discrepancy between the reanalysis data and the observation data. Bias is used to estimate the average bias in the model and to find whether the ERA5 reanalysis data tends to over- or underestimate the data compared to the observations. A positive value for bias means that the reanalysis data tends to overestimate the data and vice versa. Bias can be calculated using:

$$\text{Bias } = \frac{1}{N} \sum_{i=1}^{N} \left(C_i - O_i \right) \tag{D.2}$$

Finally, the Root Mean Square Error (RMSE) is employed to reflect the standard deviation of the reanalysis data and can be calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(C_i - O_i \right)^2} \tag{D.3}$$

In this set of equation, N denotes the total number of data points included in the analysis, C_i denotes the ERA5 reanalysis value, $\bar{C}$ is the mean value of ERA5 Reanalysis, O_i denotes the in-situ observation value and $\bar{O}$ represents the mean value of all in-situ observations. The value for R lies between -1 (negative correlation) and 1 (positive correlation). Good performance of the dataset is indicated by a R

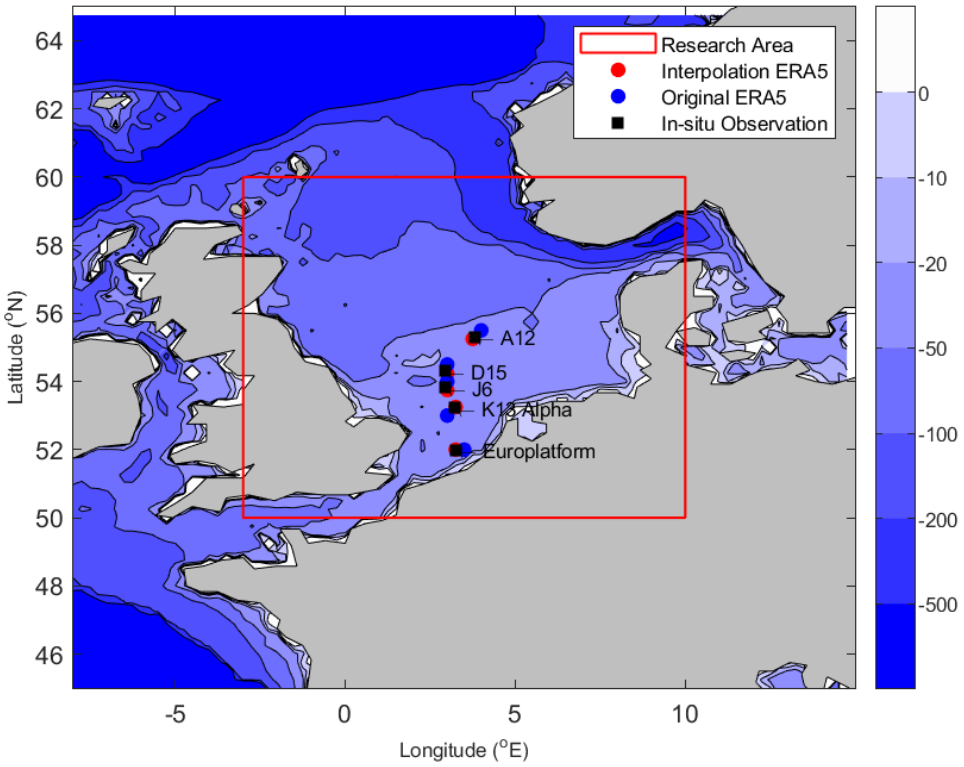


Figure D.1: Bathymetric map of North-West Europe showing the research area. The red box indicates the research area. The black squares denote the in-situ observation stations, red circles are the location where the interpolated data is extracted from and the blue circles give the original locations. Note that using interpolation, it is possible to get information about the offshore ocean data closer to the in-situ observation locations.

value close to 1. For the $bias$ and $RMSE$ holds that good performance is indicated by values that are close to 0.

Hydraulic Loading Parameters

The oceanic reanalysis data in ERA5 has a spatial resolution of 0.5 degrees in longitude and latitude and a temporal resolution of hourly data (Hersbach et al., 2020). A higher resolution can be obtained by means of interpolation (**wang2020**). In this study, the oceanic reanalysis data used for the study of spatio-temporal trend and extreme value analysis are linearly interpolated to obtain a spatial resolution of 0.25 degrees. This resolutio matches the spatial resolution of the atmospheric reanalysis. Both the interpolated data and the original dataset (i.e. spatial resolution of 0.5 degrees) are evaluated here to ensure that the interpolated data gives better results.

In-situ observations for water level and significant wave height is obtained from 5 measurement stations located at offshore boring platforms in the North Sea. The location of these boring platforms can be found in Figure D.1.

Significant Wave Height

The performance of the offshore wave data has been evaluated by means of testing the significant wave height (SWH). Other parameters related to wave boundary conditions, such as mean and peak wave period and wave direction have not been evaluated. It is assumed that their performance matches that of the significant wave height. The results for the goodness-of-fit tests for the SWH have been collected in Tables D.1 and D.2 (both the original and the interpolated ERA5 data). Scatter plots visualizing the performance of the ERA5 SWH can be seen in Figures D.2 and D.3. The colour of the scatter points indicates the density of the data point, with red indicating a high concentration of points.

Regarding the performance of the ERA5 SWH data, the following conclusions can be drawn:

- The scatterplots (Figures D.1 and D.2) show that ERA5 reanalysis data has good agreement with the in-situ observations, with the highest density of the scatter points on or very close to the identity line. For higher SWH values, the dispersion is a bit larger, but still within a reasonable limit (RMSE percentage is within 10% for higher observations).
- The bias shows very good agreement with the in-situ observations, with station J6, K13 and Europlatform showing a bias that is within 5 centimeters A12 and D15 tend to overestimate the SWH, but are still within reasonable values (order of 20 centimeters).
- Comparing the results for the original and interpolated datasets shows that it is beneficial to interpolate the data, as values for R, $bias$ and $RMSE$ for nearly all 5 station show better results. Only the bias at station J6 does not give better results. However, with a bias of $-0.037\ m$, the bias is still very much within acceptable limits.

Based on this considerations, it can be concluded that the (interpolated) ERA5 reanalysis data is suitable for the analysis of the offshore wave climate.

Table D.1: Statistics for in-situ SWH observations buoys for the original ERA5 data in the research domain. Contains location in longitude and latitude (values in brackets give the location of the original ERA5 datapoint), number of observations used for the validation of the interpolated ERA5 dataset (N), correlation coefficient (R), bias and root mean square error (RMSE).

Station	Latitude	Longitude	Duration	N	R	bias (m)	RMSE (m)
A12	55.30 (55.5)	3.80 (4.00)	2009 -2018	84790	0.982	0.137	0.235
D15	54.31 (54.50)	2.93 (3.00)	2009 - 2017	67128	0.967	0.184	0.295
J6	53.82 (54.00)	2.94 (3.00)	2009 - 2018	82136	0.970	0.010	0.215
K13 Alpha	53.22 (53.00)	3.22 (3.00)	1990 - 2018	249814	0.966	-0.019	0.226
Eurplatform	51.98 (52.00)	3.26 (3.50)	1990 - 2018	250688	0.969	-0.046	0.192

Table D.2: Statistics for in-situ SWH observations buoys for the interpolated SWH data in the research domain. Contains location in longitude and latitude (values in brackets give the location of the interpolated ERA5 datapoint), number of observations used for the validation of the interpolated ERA5 dataset (N), correlation coefficient (R), bias and root mean square error (RMSE).

Station	Latitude	Longitude	Duration	N	R	bias (m)	RMSE (m)
A12	55.30 (55.25)	3.80 (3.75)	2009 -2018	84790	0.979	0.102	0.222
D15	54.31 (54.25)	2.93 (3.00)	2009 - 2017	67128	0.967	0.158	0.277
J6	53.82 (53.75)	2.94 (3.00)	2009 - 2018	82136	0.971	-0.037	0.212
K13 Alpha	53.22 (53.25)	3.22 (3.25)	1990 - 2018	249814	0.970	0.023	0.214
Eurplatform	51.98 (52.00)	3.26 (3.25)	1990 - 2018	250688	0.968	-0.012	0.193

Wind speed

The primary driver for waves to develop is wind (see Appendix A. Also, wind has a significant influence on the water level, in the form of storm surge and wind set up. The wind models used in this research are obtained through the ERA5 Reanalysis datasets. This dataset consists of wind speed data between 1950 and 2020, giving the same temporal horizon as the wave datasets. The wind data is available in two vectors, namely u_{10}, which is the wind speed in x direction (i.e. East is positive), and v_{10}, which is the wind speed in the y direction (i.e. North is positive). However, for the analysis of the wind, the absolute magnitude ($\overrightarrow{U_{10}}$) and the wind direction with respect to true North (ϕ) are preferred. Using, the following equations, these can easily be obtained.

$$|\overrightarrow{U_{10}}| = \sqrt{u_{10}^2 + v_{10}^2} \tag{D.4}$$

$$\phi = mod\left(180 + \frac{180}{\pi} atan2(v, u), 360\right) \tag{D.5}$$

A goodness of fit analysis was not performed for the wind datasets. The reasoning for this is that there were no offshore based observation measuring stations available. Rijkswaterstaat only provides long time scale measurements for hydrodynamic parameters such as waves and water level. The KNMI only provides extensive land based wind observation data. For land based observations, care should be taken when comparing wind parameter datasets with observations, because wind observations vary on small spatial and temporal scales and are affected by the local terrain (Hersbach et al., 2020). This means the error between the ERA5 reanalysis data and the land based observations is potentially quite significant. Considering that, in this study, we are not necessarily interested in the wind climate over land, but rather in the long term trends of the wind climate as a possible driver of a changing storm climate (water level and wave height) in the North Sea. Therefore an extensive goodness-of-fit study for the wind climate was deemed excessive for the goal of this study, and was not performed.

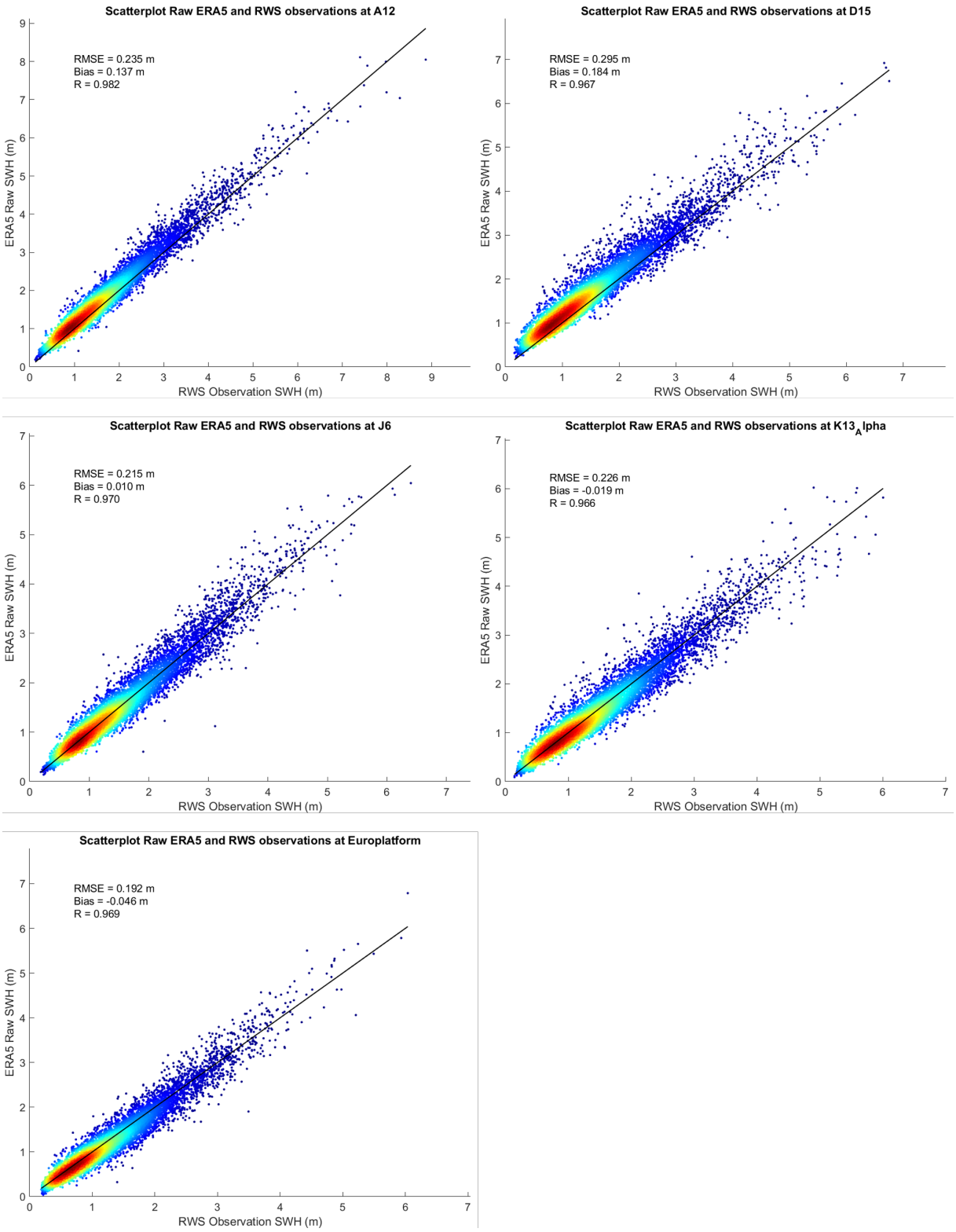


Figure D.2: Scatter plots showing the relation between the ERA5 H_{m0} and the observation H_{m0} for the original data.

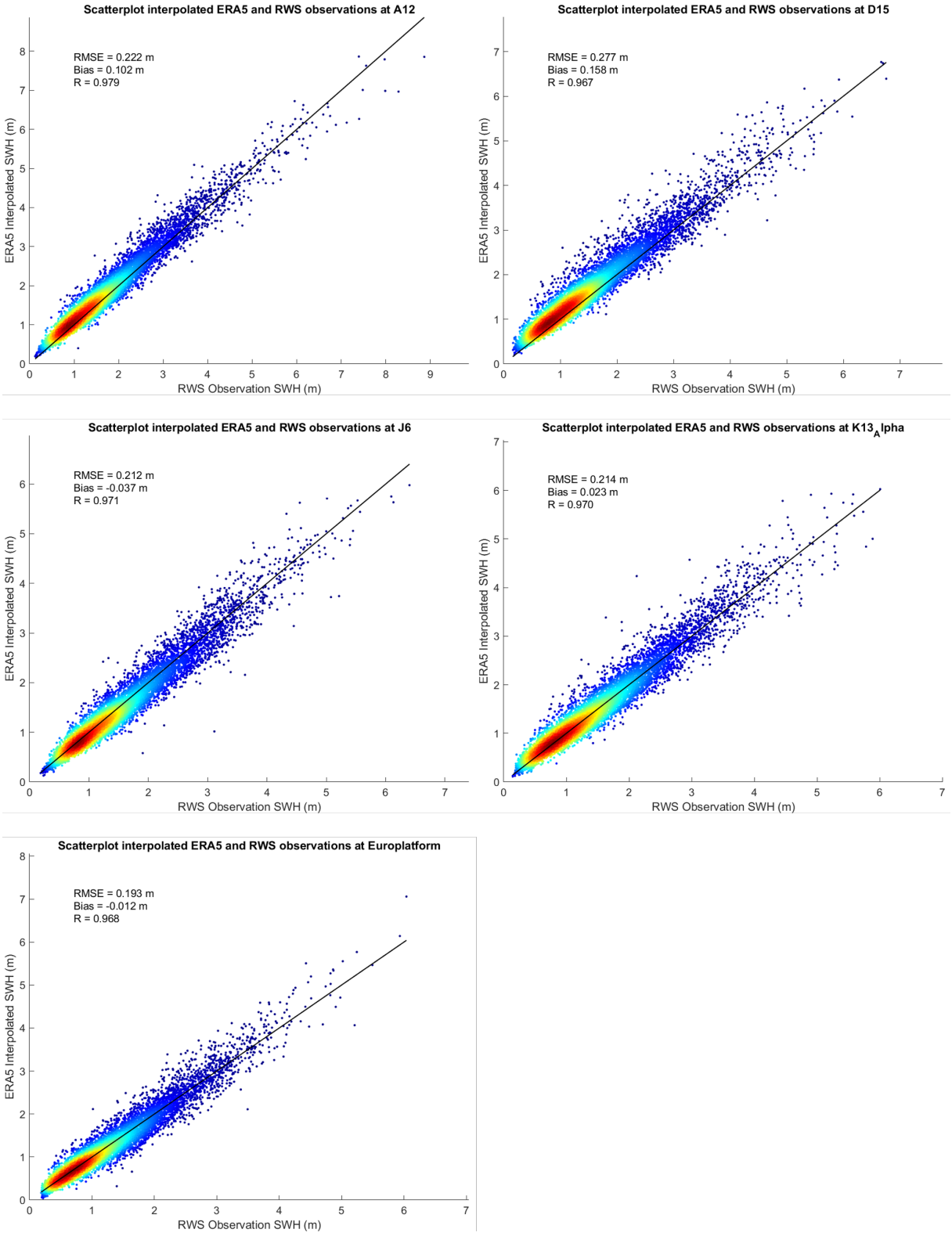


Figure D.3: Scatter plots showing the relation between the ERA5 H_{m0} and the observation H_{m0} for the interpolated data.

Wave Steepness Plots

In order to determine whether the extreme wave climate in the North Sea is dominated by wind or swell waves, steepness plots of the wave climate are made for 7 location in the North Sea domain (see Figure D.4) The steepness plots can be found in Figure D.5. From the steepness plots, it can be seen that the extreme waves (given in orange) all have a steepness between $s = 0.03$ and $s = 0.05$, which typically correspond to fresh wind waves. Based on this, it is decided that the extreme wave climate in the North Sea is dominated by wind waves rather than Swell waves.

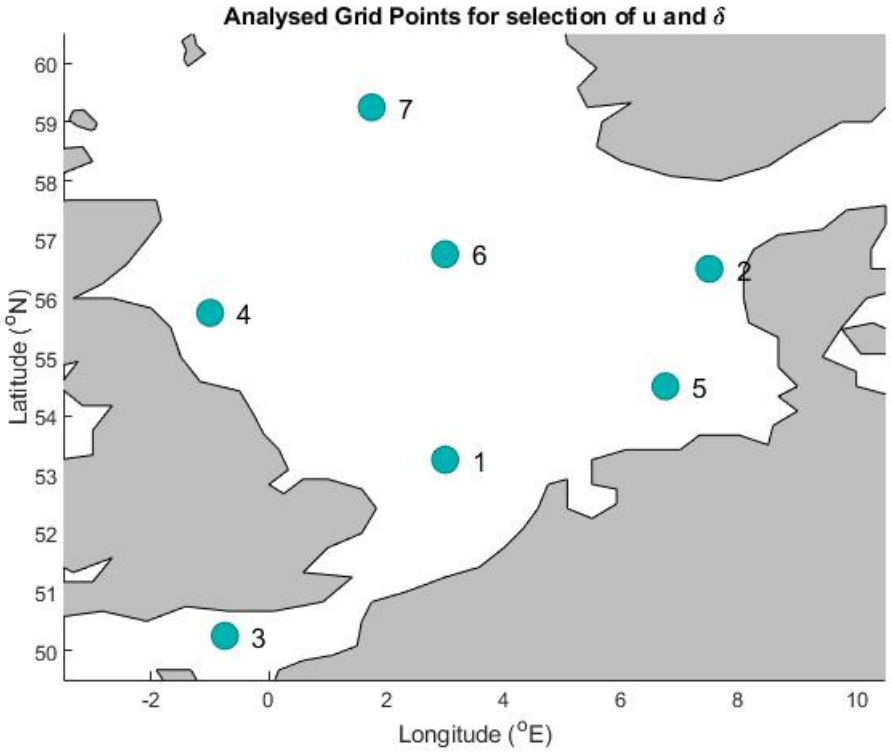


Figure D.4: Locations in the North Sea for which steepness plots have been derived.

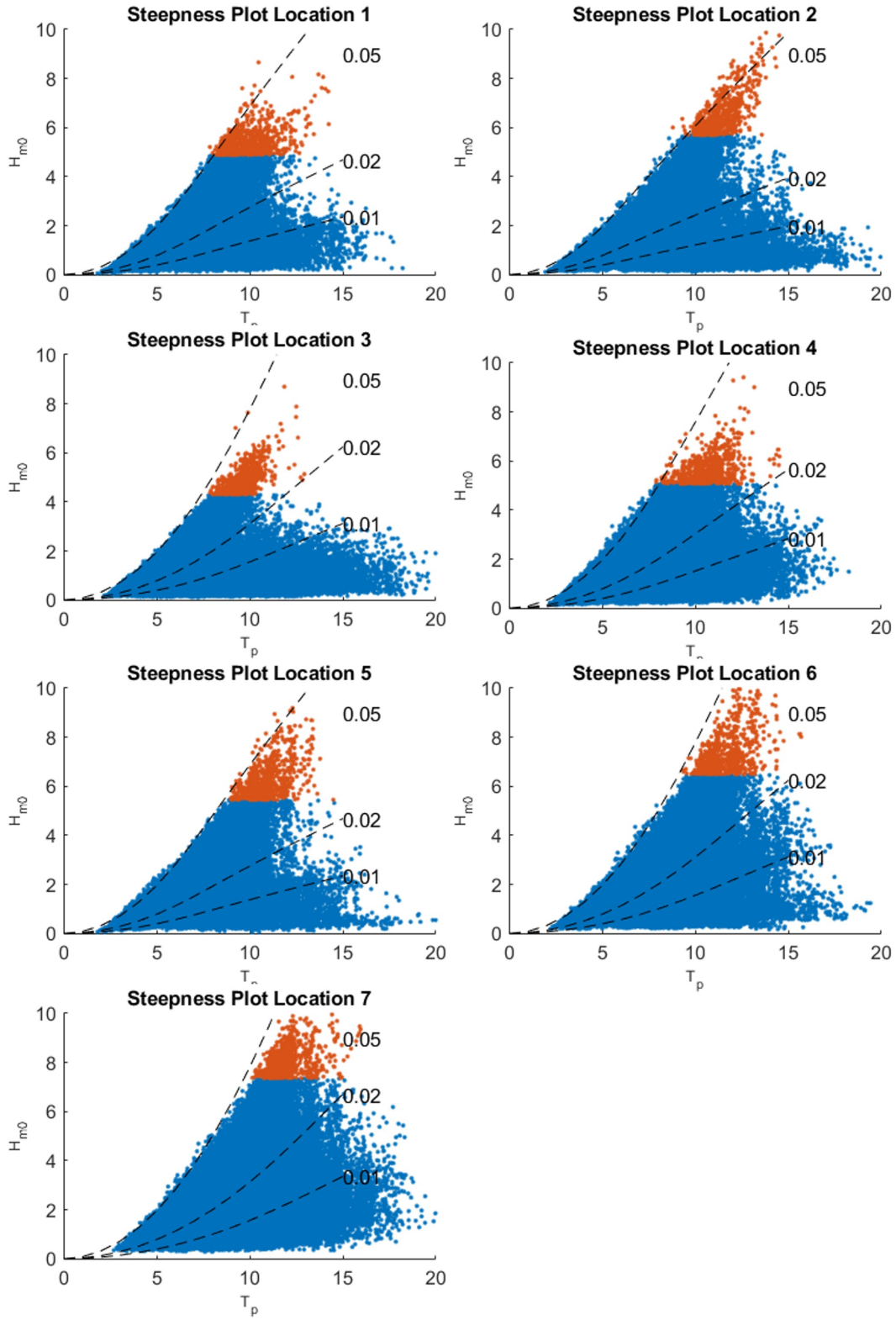


Figure D.5: Wave steepness plots for the 7 locations in the research domain

E

Supplementary Material: Peak over Threshold

The correct execution of a Peak over Threshold procedure requires the careful selection of an appropriate threshold (u) and declustering time lag (δ). The values for u and δ must be chosen such that the peaks exceeding u form a set of homogeneous and independent extreme events. In this research, Peak over Threshold is applied to define a storm threshold for the HAC-analysis in Chapter 5 and for the temporal trend analysis of the extreme wave climate in Chapter 4.

This Appendix first explains the diagnostic tests used to derive the appropriate values for u and δ. After that, we explain how u and δ were determined for the HAC analysis in Chapter 5 is explained and finally, the appropriate values for the intra-cluster spatio-temporal trend analysis of the extreme wave climate is explained.

E.1. Diagnostic Tests for Threshold Selection

The selection of the appropriate thresholds (u) and declustering times (δ) in this research is performed using the diagnostic tests of mean residual life, dispersion index and extremal index. Also, the the stability of the GP distribution parameters (Antonini et al., 2019; Coles, 2001; Davison & Smith, 1990) is studied. The sections below explain the diagnostic tests in more detail.

E.1.1. Mean Residual Life

A first attempt to find a suitable threshold level is by means of a mean residual life plot. In mean residual life plots, the mean observed excess over threshold u is plotted against u. For the GP distribution to be considered valid, the mean residual life plot should follow a straight line (Coles, 2001; Davison & Smith, 1990).

E.1.2. Extremal Index

The POT procedure and GP distribution are based on the assumption that the extremes above the threshold are independent. The extremal index (EI) can be used to validate whether the considered threshold and declustering time lag (δ) results in a set of independently distributed extremes (Ferro & Segers, 2003). The extremal index (EI), as proposed by Ferro and Segers (2003) evaluates the distribution of exceedance times of the sampled extremes and provides a measure of independence. The extremal index is defined as:

$$EI = \begin{cases} 0 & \text{if } max_i \{T_i\} \leq 1 \\ min \left\{ 1, 2 \left(\frac{\sum_i T_i}{n} \right)^2 / \left(\frac{\sum_i T_i^2}{n} \right) \right\} & \text{if } 2 \leq max_i \{T_i\} > 1 \\ min \left\{ 1, 2 \left(\frac{\sum_i (T_i - 1)}{n} \right)^2 / \frac{\sum_i (T_i - 1)(T_i - 2)}{n} \right\} & \text{otherwise} \end{cases} \tag{E.1}$$

In Equation E.1, T_i is the inter exceedance time defined as $T_i = S_i – S_{i-1}$. The extremal index is defined such that if $EI = 1$, the observations can be considered independent, and some degree of dependence is present for values $EI < 1$. Typically, observations are deemed sufficiently independent to be fit for extreme value analysis if $EI > 0.85$ (Davison & Smith, 1990).

E.1.3. Dispersion Index

Finally, the dispersion index (DI) can be used to verify the assumption that the selected extreme observations for threshold u and declustering time lag δ follow a Poisson process (Antonini et al., 2019; Timmermans, 2021). Let X be a random variable the follows a Poisson distribution with distribution parameter $\lambda > 0$. The probability function of X can be expressed as follows:

$$\mathrm{P}(\mathrm{X} = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \tag{E.2}$$

Where k is the number of occurrences (storm events). One of the key properties of the Poisson process is that the distribution parameter $lambda$ is equal to the expected value $E(x)$ and also the variance $Var(X)$ (Dekking et al., 2005). The dispersion index was proposed by Cunnane (1979) and can be defined as the ratio between the variance and expectation of the number of peaks:

$$DI = \frac{\sigma^2}{\mu}, \quad \text{for} \quad \mu \neq 0 \tag{E.3}$$

Therefore, the selected extreme values can be considered Poisson distributed if the Dispersion index is equal to 1. An acceptable peak separation should give a dispersion index close to 1 (Antonini et al., 2019; Cunnane, 1979). Often, the dispersion index comes with confidence bounds. These confidence bounds indicate the domain for which the Chi-squared test with null-hypothesis $H_0 : DI = 1$ cannot be rejected (Timmermans, 2021).

E.1.4. Stability of GPD Parameters

A complementary technique to the three diagnostic plots mentioned above is to fit the GP distribution to a range to a range of potential thresholds, and check the stability of the shape and scale parameters (Coles, 2001). The argument is as follows: according to extreme value theory, if the GP distribution is an appropriate model to for the observations above a certain threshold u_0, then for a higher threshold u, a GP distribution should also apply. Both GP distributions have identical shape parameters (Coles, 2001). However, the scale parameter requires reparameterization for a correct stability analysis (Coles, 2001). If σ_u is the scale parameter for some threshold $u > u_0$, it follows from extreme value theory that:

$$\sigma_u = \sigma_{u_0} + \xi\,(u - u_0) \tag{E.4}$$

Then, reparameterization of σ_u can be done as follows:

$$\sigma^* = \sigma_u - \xi u \tag{E.5}$$

Consequently, estimates of the $\sigma*$ and ξ should be constant for any value of $u > u_0$, if u_0 is a valid threshold of the observations to follow the GP distribution (Coles, 2001). See also figure E.2 for an example of stable regions of the shape and scale parameters of the GPD distributions.

E.2. Threshold Selection for HAC-Analysis

For the Hierarchical Agglomerative Clustering (HAC) analysis, the definition of a storm climate was required. Waves exceeding a particular threshold u would be considered storm waves. To make the HAC-analysis straightforward, we decided that the selected values for u and δ should be appropriate throughout the research domain. Potentially useful threshold values for the analysis of the extreme significant wave height have been studied in Arns et al. (2013), Méndez et al. (2006), Sartini et al. (2015), and Viselli et al. (2015). The definitions of the thresholds are listed in Table E.1. It was decided to analyse these thresholds in this procedure as well, with one additional threshold value namely the 99.8th percentile for H_{m0} (See Table E.1). As for the possible values for δ, it was decided to study different declustering time lags spanning 14 days (in accordance with Antonini et al. (2019)) at an interval of 6

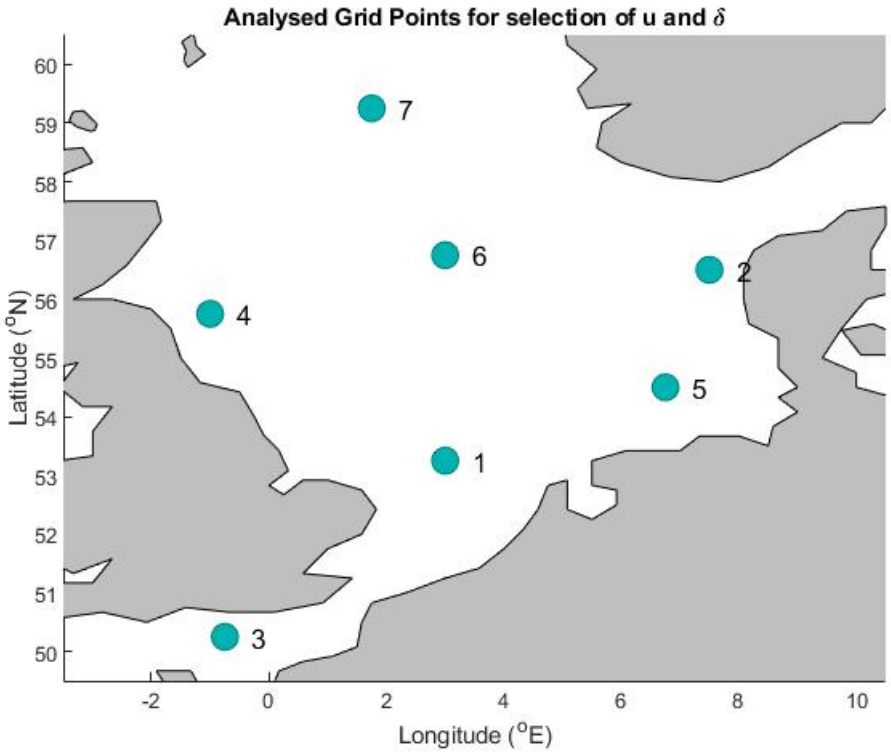


Figure E.1: Analyzed grid points for the selection of the threshold u and declustering time lag δ for the HAC-Analysis in Section 5.1.

hours, resulting in 84 potential values for δ.

The plots for Mean Residual Life, Dispersion Index, Extremal index and the stability of the GP distribution parameters have been constructed for 7 different locations throughout the research domain (See figure E.1). These 7 locations have been selected based on the expectation that the extreme waves in these locations potentially show different characteristics, and therefore different appropriate values for u and δ prevail in these areas. The goal is to find a value for u and δ that is appropriate for all 7 grid points. Then, the assumption is made that if the values u and δ are appropriate for these 7 locations, they are appropriate for all grid points in the research domain.

Looking carefully at figures E.2 - E.8, it can be seen that threshold 5 ($p_{99.5}\bar{H_{m0}}$) is the minimum threshold

Table E.1: Tested threshold values for the POT analysis. $\overline{H_{m0}}$ and $\sigma_{H_m 0}$ are the mean and standard deviation of the significant wave height (from all directions). $p_i H_{m0}$ is the ith percentile of the significant wave height (from all directions).

Threshold	Expression	Source
th_1	$\overline{H_{m0}} + 1.4 \cdot \sigma_{H_m 0}$	Viselli et al. (2015)
th_2	$\overline{H_{m0}} + 1.9 \cdot \sigma_{H_m 0}$	Viselli et al. (2015)
th_3	$p_{97.5} H_{m0}$	Viselli et al. (2015)
th_4	$\overline{H_{m0}} + 3.0 \cdot \sigma_{H_m 0}$	Arns et al. (2013)
th_5	$p_{99.0} H_{m0}$	Arns et al. (2013)
th_6	$p_{99.5} H_{m0}$	Sartini et al. (2015)
th_7	$p_{99.7} H_{m0}$	Sartini et al. (2015)
th_8	$p_{99.7} H_{m0}$	NA

that is appropriate to give a Poisson process at each of the 7 considered grid points. Selecting the lowest possible threshold is important to increase the number of extreme observations in the analysis. Further analysis of the Mean Residual Life, Dispersion Index and the GP distribution parameters shows that threshold 5 respects all of the requirements to be a suitable threshold. Therefore, threshold 5 is the selected threshold for the HAC analysis. Looking at the extremal index curve for threshold 5, it becomes apparent that a high value for δ is required to satisfy the requirement for independent extreme observations. In the end, it was found that the most suitable value for δ is 8 days, or 192 hours.

Grid Point 1

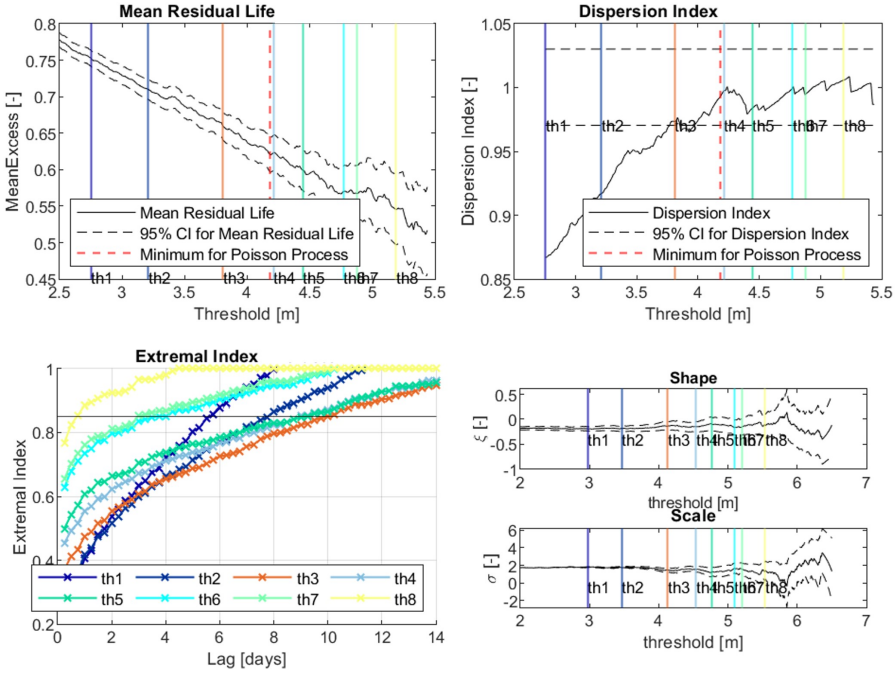


Figure E.2: Diagnostic plots for grid point 1 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 2

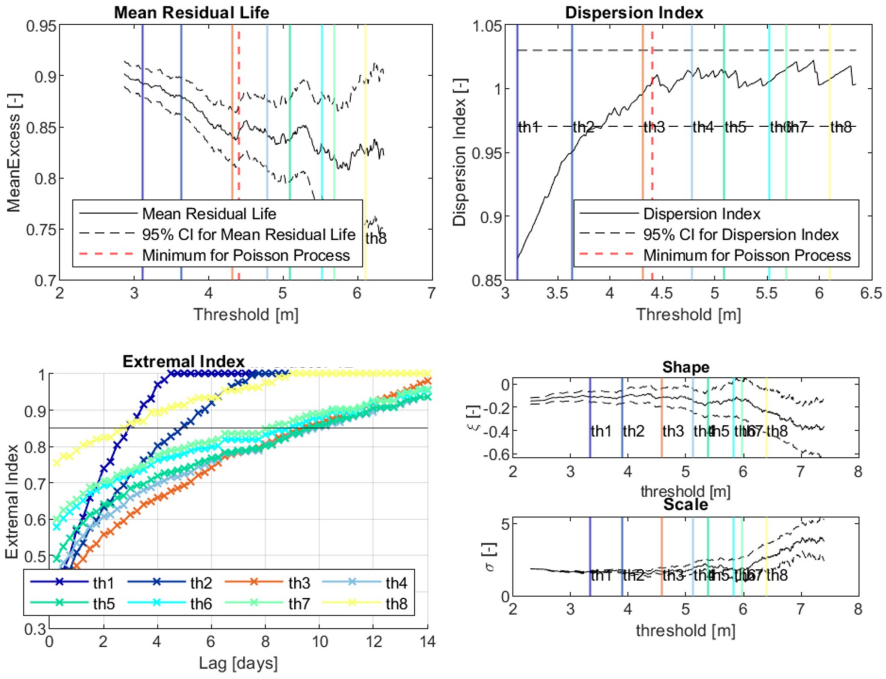


Figure E.3: Diagnostic plots for grid point 2 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 3

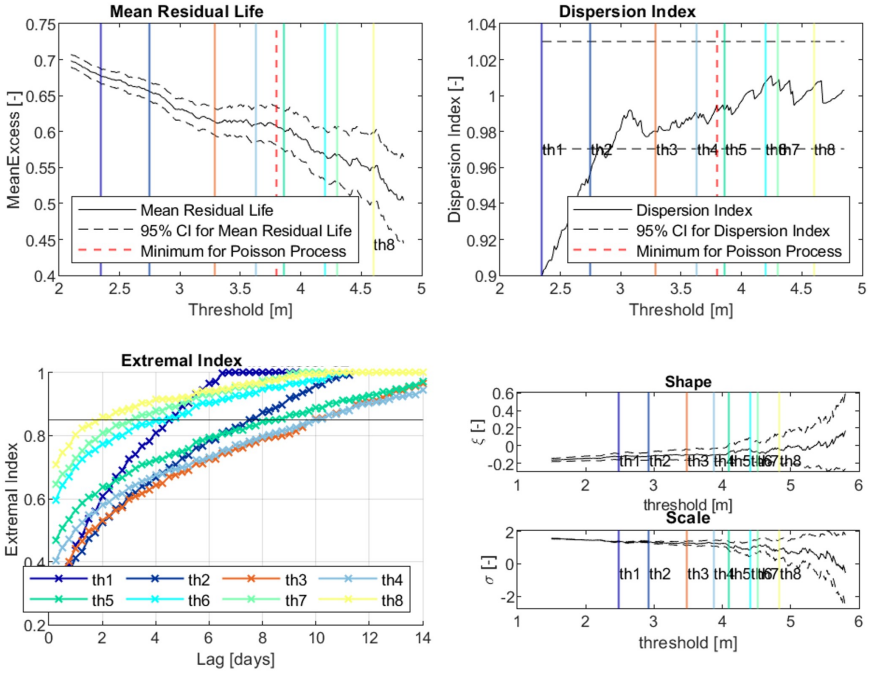


Figure E.4: Diagnostic plots for grid point 3 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 4

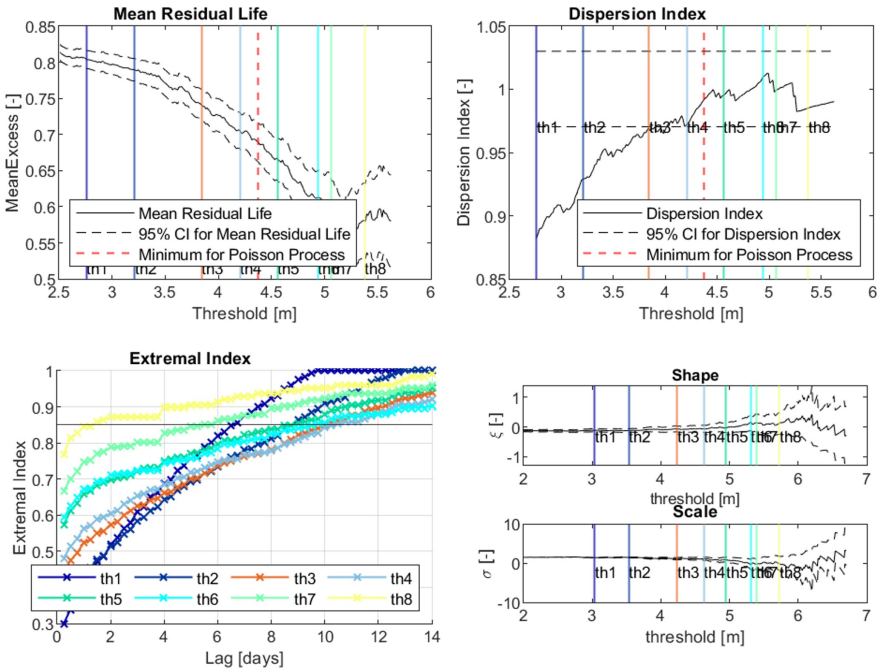


Figure E.5: Diagnostic plots for grid point 4 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 5

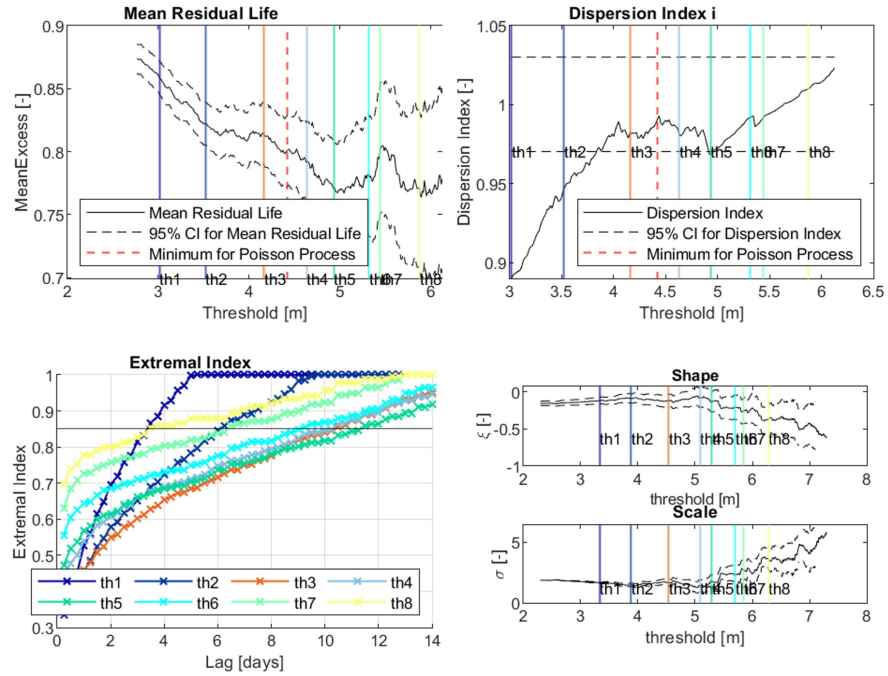


Figure E.6: Diagnostic plots for grid point 5 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 6

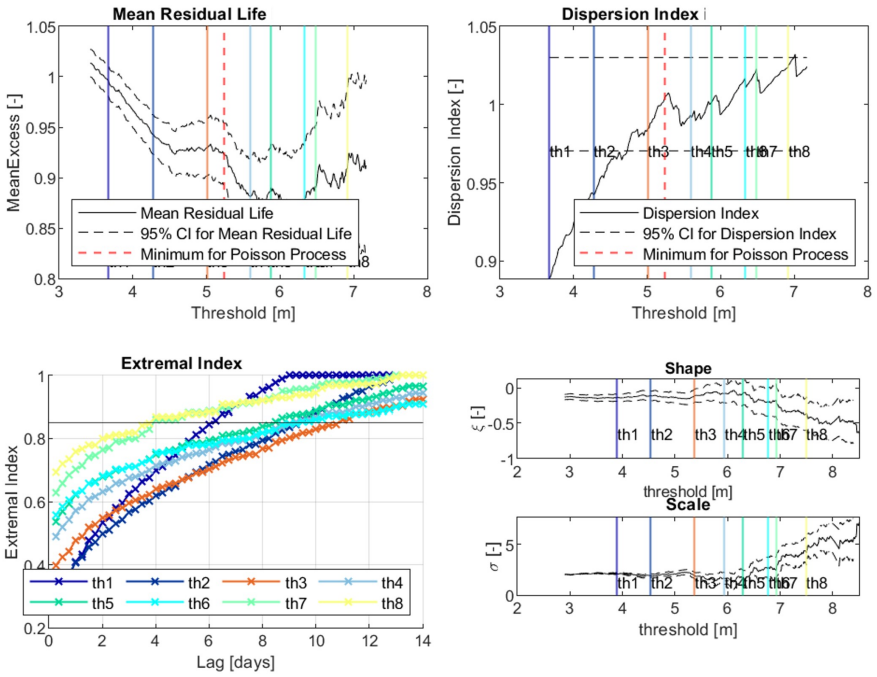


Figure E.7: Diagnostic plots for grid point 6 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Grid Point 7

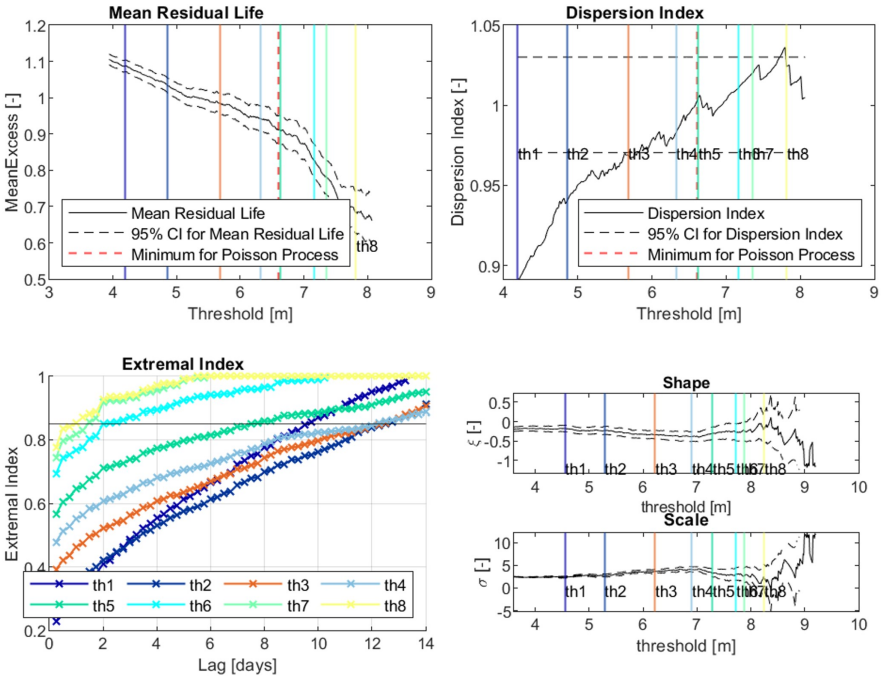


Figure E.8: Diagnostic plots for grid point 7 in Figure E.1. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

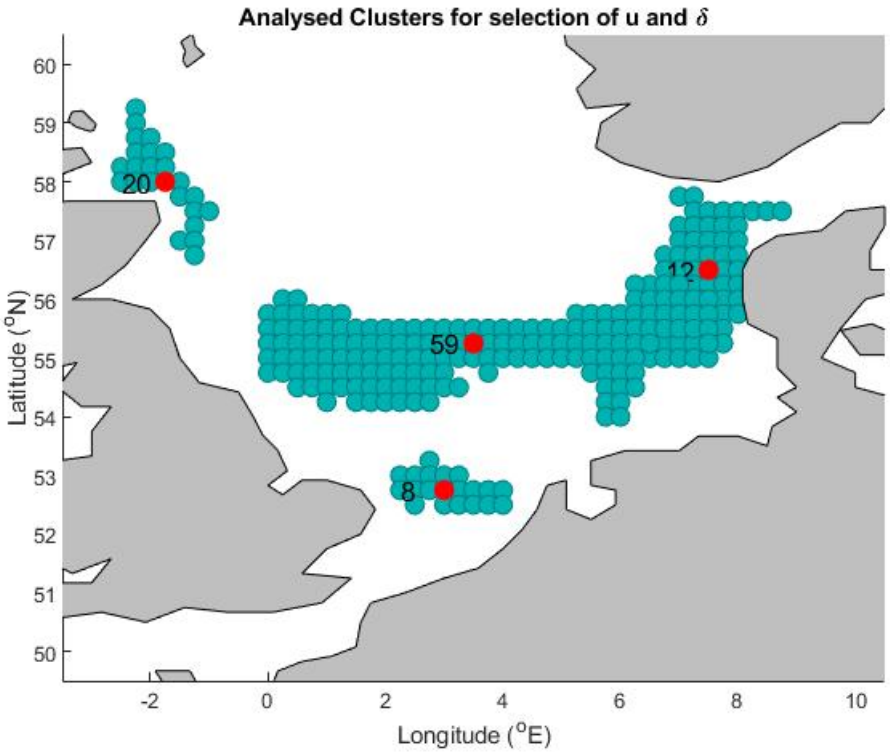


Figure E.9: Clusters subjected to intra-cluster spatio-temporal trend analysis. The grid points given in red are the centroids of the clusters of which the time-series was used to determine the appropriate values for U and δ

E.3. Threshold Selection for Intra-Cluster Trend Analysis

For the intra-cluster spatio-temporal trend analysis of the extreme wave and wind climate in the North Sea in Chapter 4, the appropriate threshold (u) and declustering time lag (δ) are selected for each of the analyzed clusters. To do so, the knowledge of the dominant wave directions derived in the clustering analysis (Chapter 5) is applied to find the dominant wave directions in each of the clusters. The time-series for the cluster centroids where filtered based on the dominant wave directions (see Figure E.9. Subsequently, the appropriate values for u and δ were determined based on Peak over Threshold procedures. The clusters subjected to analysis are visualized in Figure E.9. The determined boundaries for the dominant wave directions (θ_{LB} and θ_{UB}) and the values for u and δ can be found in Table E.2. Some remarks about the threshold selection for the different clusters include:

- For cluster 12 (Danish coast), th_6 is selected despite not meeting the requirement of a Poisson process according to the Dispersion Index. However, th_8 does not meet the requirements for the MRL, DI and the stability of the GP parameters. Thresholds th_6 and th_7 lie the closest to the Poisson ratio. In the end, threshold th_6 is selected since it meets all other requirements;

Table E.2: Selected Boundaries for the dominant wave directions, threshold and declustering time lag for the selected clusters subjected to intra-cluster spatio-temporal trend analysis in Chapter 4. The subscripts 1 and 2 for Cluster 8 indicate the primary and secondary dominant wave direction respectively.

Location	Cluster	θ_{LB} (deg N)	θ_{LB} (deg N)	u	δ (days)
Dutch Coast	8-1	270	360	th_5	2
	8-2	180	270	th_6	2
Danish Coast	12	210	330	th_6	5
Scottisch Coast	20	60	180	th_6	2
Central North Sea	59	260	360	th_6	5

- For cluster 59 (Central North Sea), it can be seen that the selected values for u (th_6 and δ (5 days) do not meet the requirements of the Shape parameter of the GP distribution. However, since this is the only combination that meets the requirements for MRL, DI and EI, it is decided that this was a necessary compromise in order to be able to perform the spatio-temporal analysis in this cluster.

Cluster 8

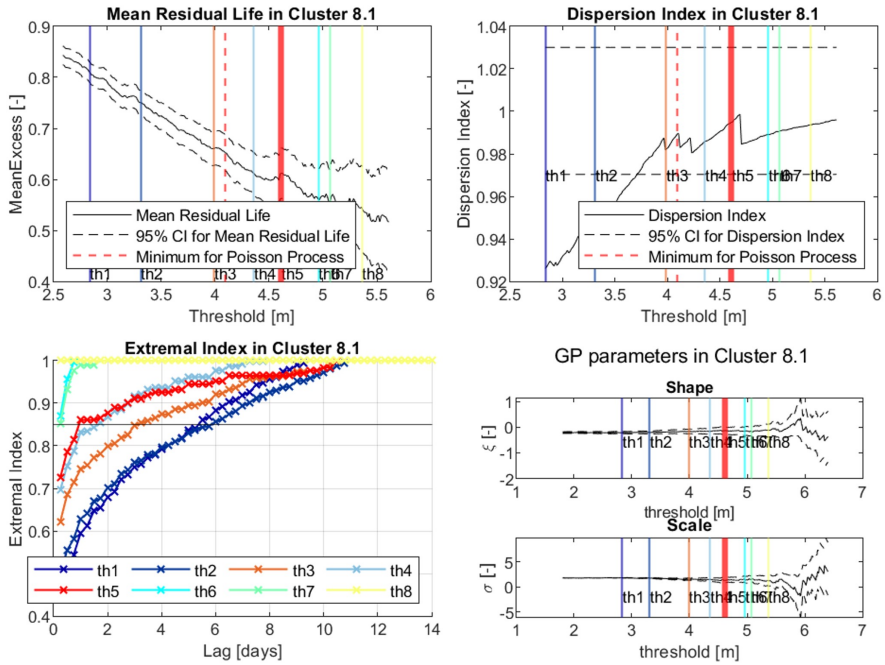


Figure E.10: Diagnostic plots for cluster 8 (primary dominant wave direction) in Figure E.9. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

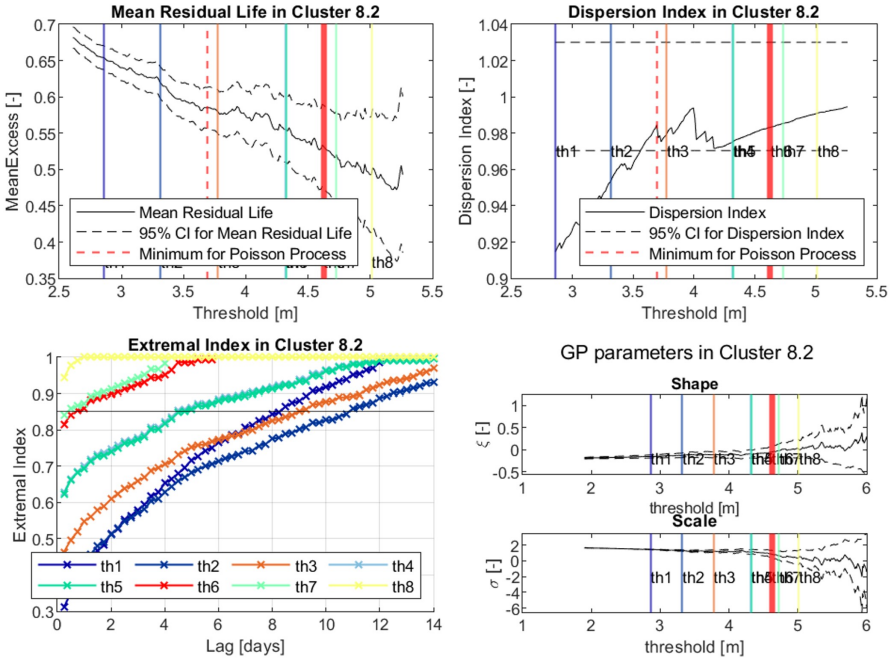


Figure E.11: Diagnostic plots for cluster 8 (secondary dominant wave direction) in Figure E.9. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Cluster 12

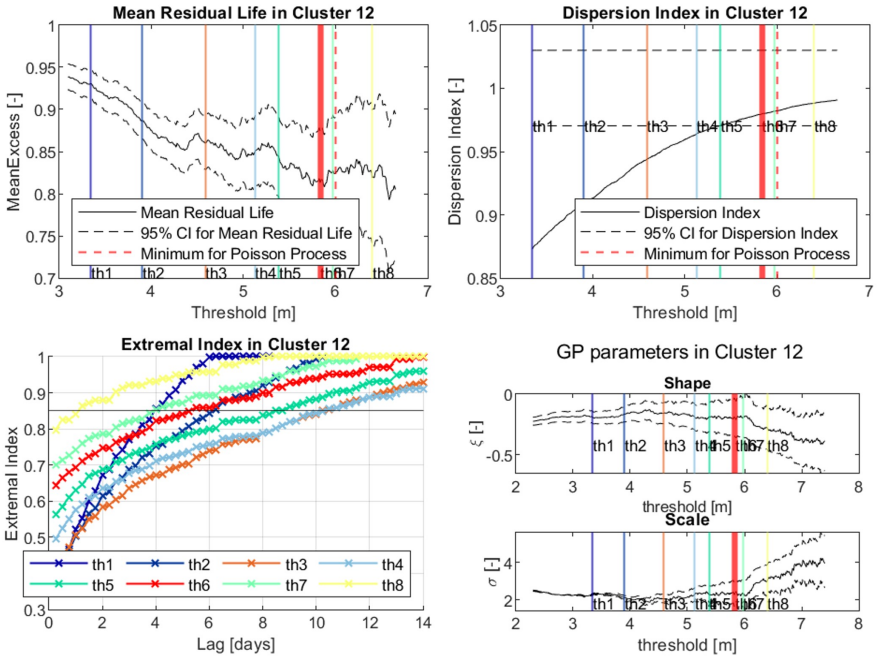


Figure E.12: Diagnostic plots for cluster 12 (primary dominant wave direction) in Figure E.9. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Cluster 20

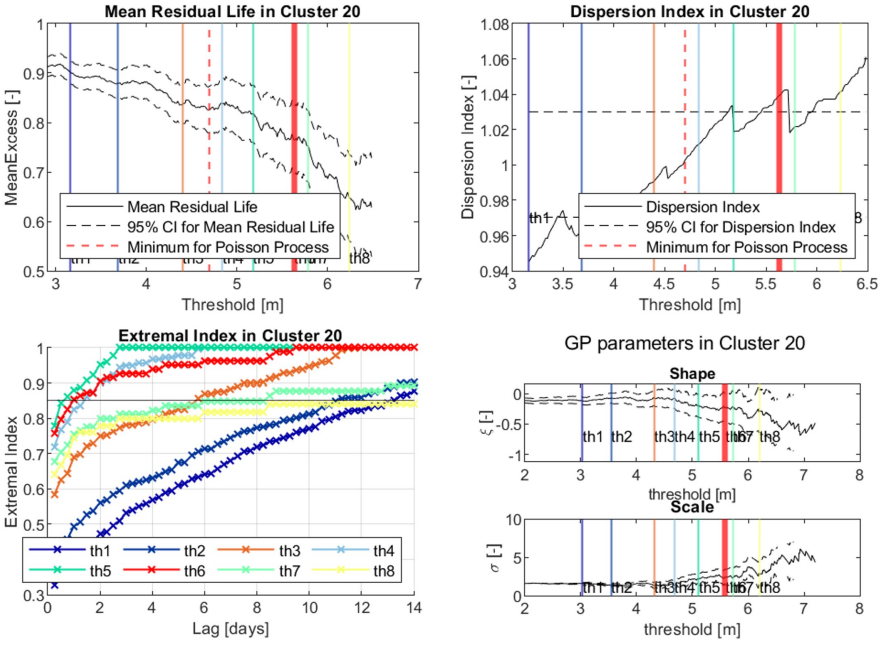


Figure E.13: Diagnostic plots for cluster 20 (primary dominant wave direction) in Figure E.9. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

Cluster 59

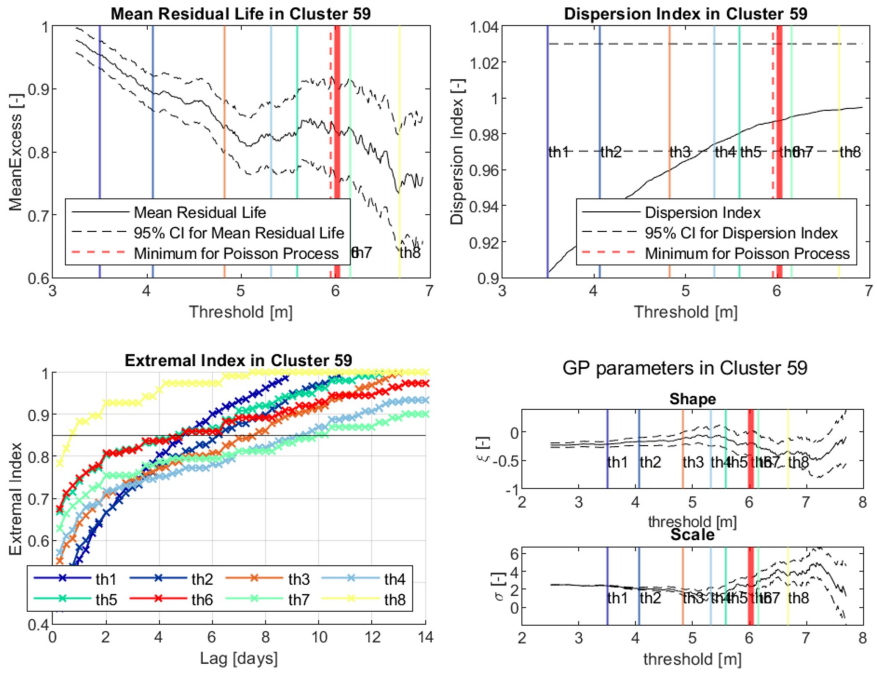


Figure E.14: Diagnostic plots for cluster 59 (primary dominant wave direction) in Figure E.9. Top Left: Mean Residual Life plot, top right: Dispersion Index, bottom left: Extremal Index and finally bottom right: the GP distribution parameters.

F

Supplementary Material Clustering

This appendix provides supplementary information corresponding to the clustering analysis of the North Sea (as performed in chapter 5. Comments about the clustering results for the different configurations are given, and for each configuration, the best clustering solution is determined. For the analysis of the different clustering results, the following criteria are evaluated:

- When comparing the results for different linkage mechanisms for the same parameter combination and weight configuration, how do the scores for the quality of the dendrogram (cophenetic correlation C) and the scores for the optimal clustering solution - Silhouette Coefficient (SC), Calinski Harabasz criterion (CH) and the Davies-Bouldin value (DB) - compare.
- For the cluster evaluation criteria (SC, CH and DB), how much agreement is there regarding the ideal number of clusters ($\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$) for each respective evaluation metric. When there is a lot of agreement, there is a clear optimal clustering solution. Also the stability of SC, CH and DB for varying values for $\mathcal{K}$ is assessed. Higher stability gives more room to find the optimal clustering solution
- How much emphasis is placed on one parameter compared to the other parameters included in the analysis. Weight can be assigned to the parameters to place more emphasis on different parameters in the clustering which are deemed more important. However, too much weight can cause the clustering algorithm to miss some essential details, which may become problematic in further stages of this report.

F.1. Combination 1: Significant Wave Height

The first combination of wave parameters includes the primary dominant wave direction (θ_1), secondary dominant wave directions (θ_2), the number of dominant wave directions at the grid point (N_θ) and the 99.5 percentile level for significant wave height ($H_{m0,P99.5}$).

Configuration 1.1

For configuration 1.1, most weight was assigned to $H_{m0,P99.5}$. Analyzing the clustering results (see Figure F.1), the following comments can be made:

- When comparing the scores for C, SC, CH and DB for average and ward linkage, it can be observed that Average linkage produces the better dendrogram followed by Ward linkage. When comparing the SC, CH and DB, it can be seen that Ward and Average linkage have similar scores for SC and DB. However, the score for CH for Ward linkage is much higher than for Average linkage.
- For both linkage mechanisms (Ward and Average), there is some disagreement regarding the ideal number of clusters as determined by the internal evaluation metrics. Therefore, finding an optimal clustering solution difficult. However, the values for SC and DB are quite stable for varying (high) values of $\mathcal{K}$, meaning that a compromise can be found without losing too much "cluster quality."

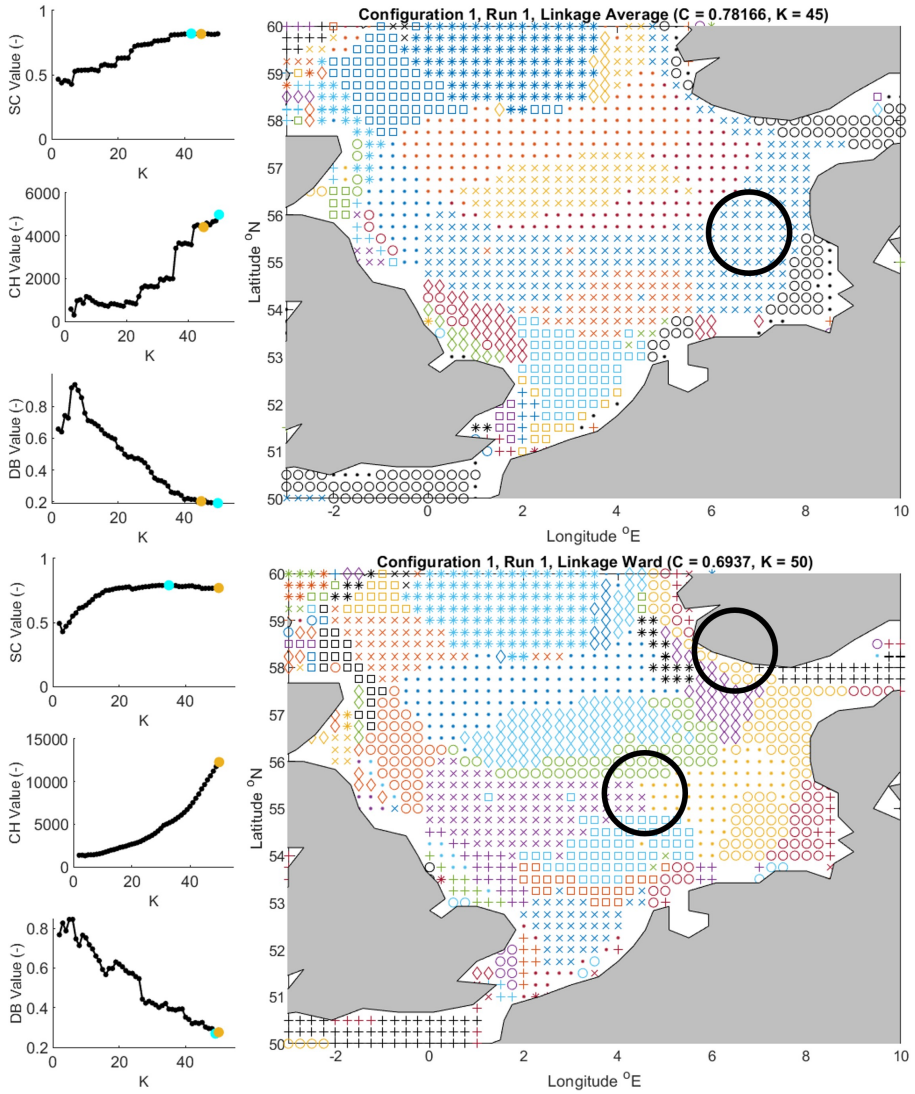


Figure F.1: Clustering results for configuration 1.1 using average (top) and Ward (bottom) linkage mechanisms.

- Average linkage places a lot of emphasis on cluster divisions based on $H_{m0,P99.5}$. This can be especially noticed when studying the area between the central North Sea and the German Bight of the German Blight. One would expect a division based on the primary dominant wave direction (see Figure 4.8).
- For Ward linkage, although there is more division on wave direction (see cluster division based on primary dominant wave direction in the German blight, there are still some areas with no cluster divisions where one would expect different wave directions to prevail. Also, there is a explainable cluster division in the central North Sea.

Due to the superior values of CH for Ward linkage over Average linkage, and the better cluster divisions based on θ_1, θ_2 and N_θ, it is decided that the best clustering solution for this configuration is by using Ward Linkage. Due to the sharp rise in the CH plot (see Figure F.1), it is decided the ideal vor K is 50.

Configuration 1.2

Compared to configuration 1.1, more weight was moved from $H_{m0,P99.5}$ to the parameters for dominant wave direction. Analyzing the clustering results (see Figure F.2), the following comments can be made:

- When comparing the scores for C, SC, CH and DB for average and ward linkage, it can be observed that Average linkage produces the better dendrogram. The quality of the dendrogram

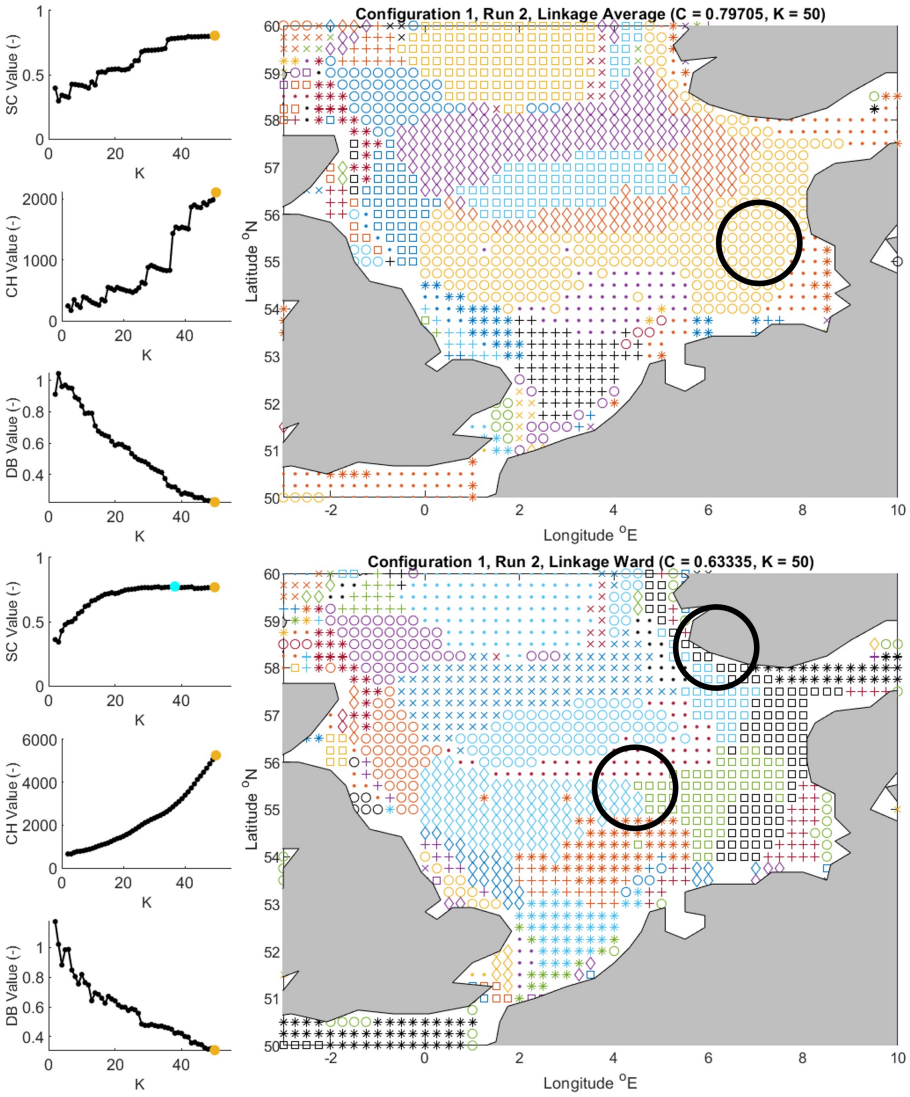


Figure F.2: Clustering results for configuration 1.2 using average (top) and Ward (bottom) linkage mechanisms.

for Ward linkage is even reduced when comparing it with configuration 1. When comparing the SC, CH and DB, it can be seen that Ward and Average linkage have similar scores for SC and DB. However, the score for CH for Ward linkage is much higher than for Average linkage (more than twice as high).

- For Average linkage, all scores for SC, CH and DB agree that the optimal number of clusters $\mathcal{K}$ is 50. For Ward linkage, SC does not agree with CH and DB for the ideal number of clusters, therefore making an optimal clustering solution difficult.
- Average linkage places a lot of emphasis on the wave height when clustering. This can be especially noticed when studying the area of the German Bight. One would expect a division based on the primary dominant wave direction (see Figure 4.8).
- For Ward linkage, although there is more division on wave direction (see cluster division based on primary dominant wave direction in the German blight, there are still some areas with no cluster divisions where one would expect different wave directions to prevail. Also, there is an inexplainable cluster division in the central North Sea which cannot be attributed to either $H_{m0,P99.5}$ or the wave direction parameters.

Although there is a clear agreement in the ideal number of clusters for Average linkage, we still decided that Ward linkage is the better option for this configuration. This is decided based on the better subdivision for the wave direction parameters and the higher CH score for Ward linkage than for Average

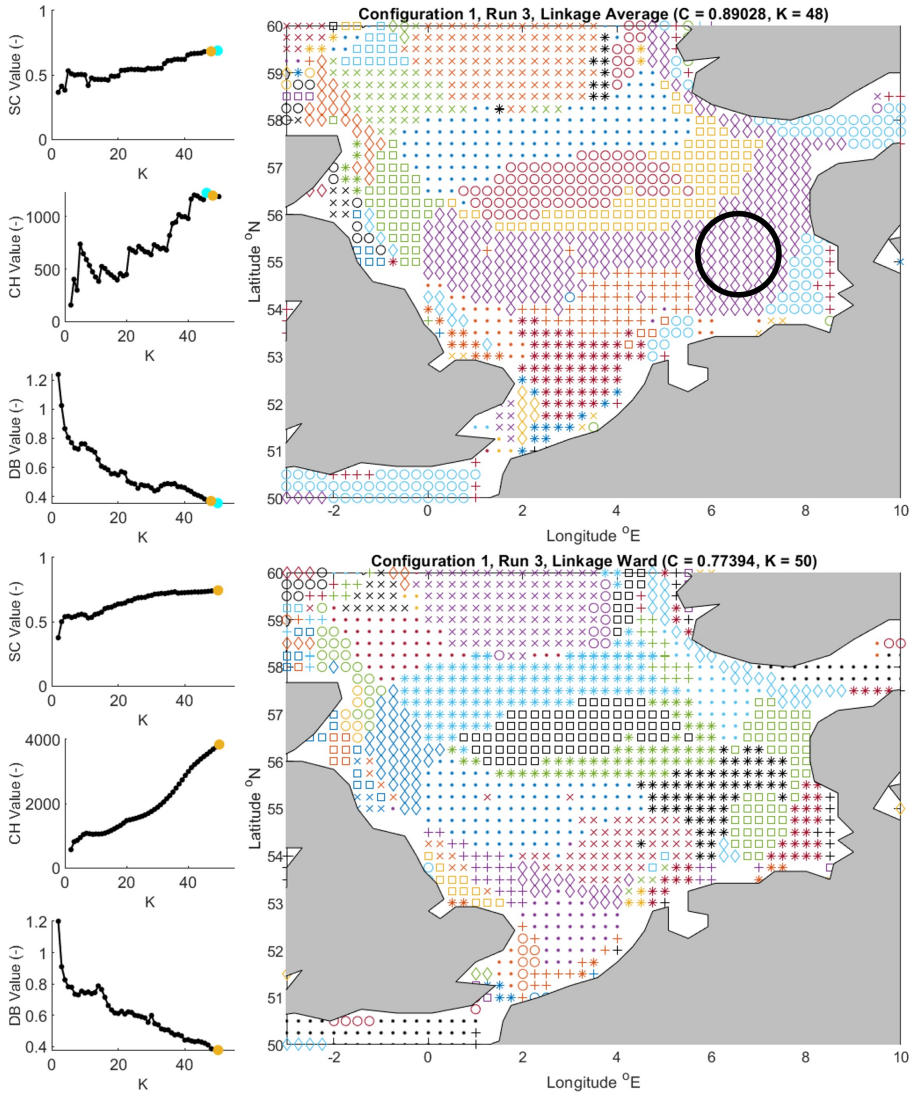


Figure F.3: Clustering results for wave parameter combination 1, weight configuration 3 and average (top) and Ward (bottom) linkage mechanisms.

linkage. However, it must be noted that neither clustering solution performs extremely well for this configuration

Configuration 1.3

Compared to configuration 1.3, equal weight was assigned to all parameters. Analyzing the clustering results (see Figure F.2), the following comments can be made:

- It can be observed that Average linkage produces the better dendrogram based on the $C - score$, again followed by Ward linkage. When comparing the SC, CH and DB, it can be seen that Ward and Average linkage have similar scores for DB. However, the score for SC and CH for Ward linkage are (much) higher than for Average linkage (CH is more than 3 times as high).
- For Ward linkage, all scores for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ agree that the optimal value for $\mathcal{K}$ is 50. For Average linkage, SC does not agree with CH and DB for the ideal number of clusters. However, values for SC are stable for varying values of $\mathcal{K}$, making it possible to find a good clustering solution for this configuration.
- Despite the equal weights assigned to all parameters, average linkage still results in a lot of cluster division based on $H_{m0,P99.5}$. Once again, the area of the German Bight seems problematic. One would expect a division based on θ_1 there.

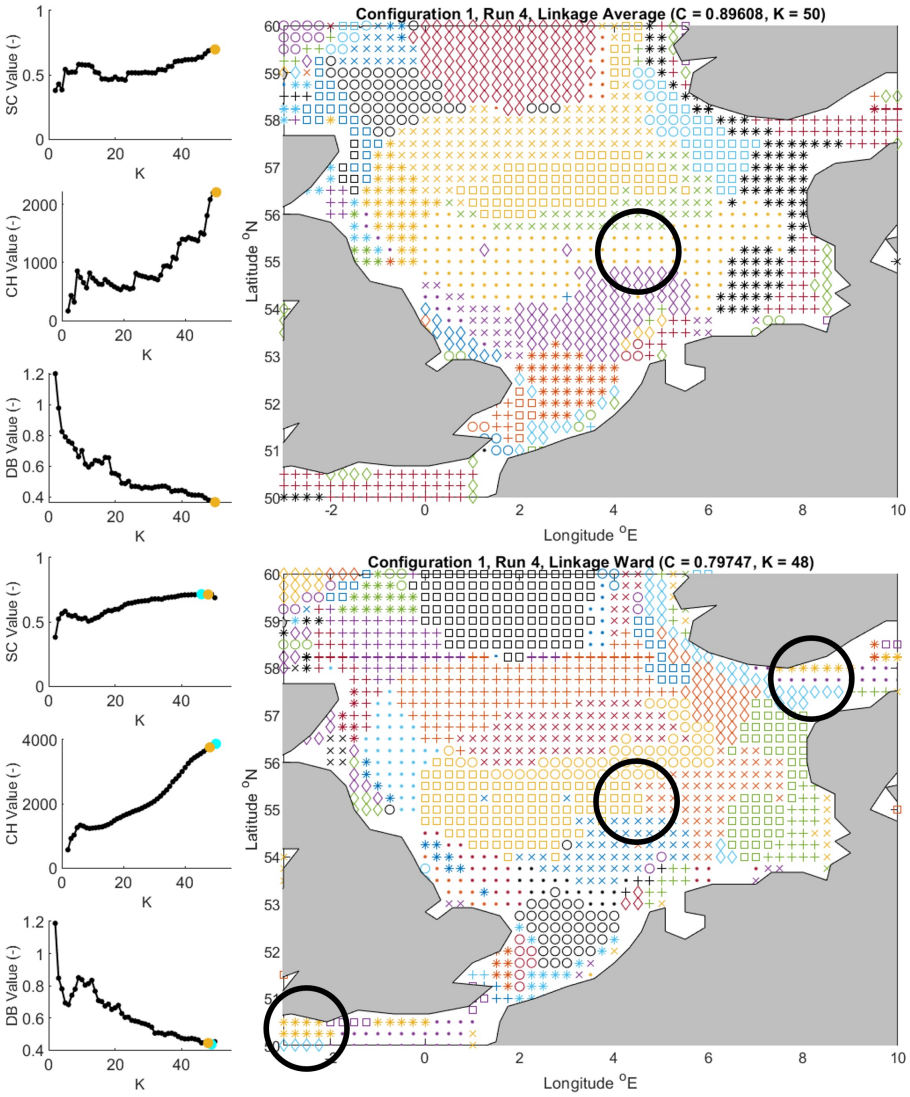


Figure F.4: Clustering results for wave parameter combination 1, weight configuration 4 and average (top) and Ward (bottom) linkage mechanisms.

- For Ward linkage, there are clear cluster divisions seen in the German Bight as well as near the coast of Norway. Also, there still is an in-explainable cluster division in the central North Sea which cannot be attributed to either $H_{m0,P99.5}$ or the wave direction parameters.

In the end, due to the superior scores for SC CH and DB, the clear agreement on the ideal value for $\mathcal{K}$ and the clear cluster divisions based on wave direction parameters and $H_{m0,P99.5}$, the ideal clustering solution for this configuration is Ward linkage with $\mathcal{K} = 50$.

Configuration 1.4

In configuration 1.4, more weight is assigned to the wave direction parameters compared to $H_{m0,P99.5}$. Analyzing the clustering results (see Figure F.4), the following comments can be made:

- When comparing linkage mechanisms, it can be observed that Average linkage produces the better dendrogram based on the $C-score$. Also when comparing the results for SC CH and DB, it can be seen that Average and Ward clustering have similar results. The CH score for Ward clustering is still better than for Average clustering.
- For Average linkage, all scores for SC, CH and DB agree that the optimal number of clusters $\mathcal{K}$ is 50. For Ward linkage, there is some disagreement on the ideal number of clusters, making clustering potentially more difficult.

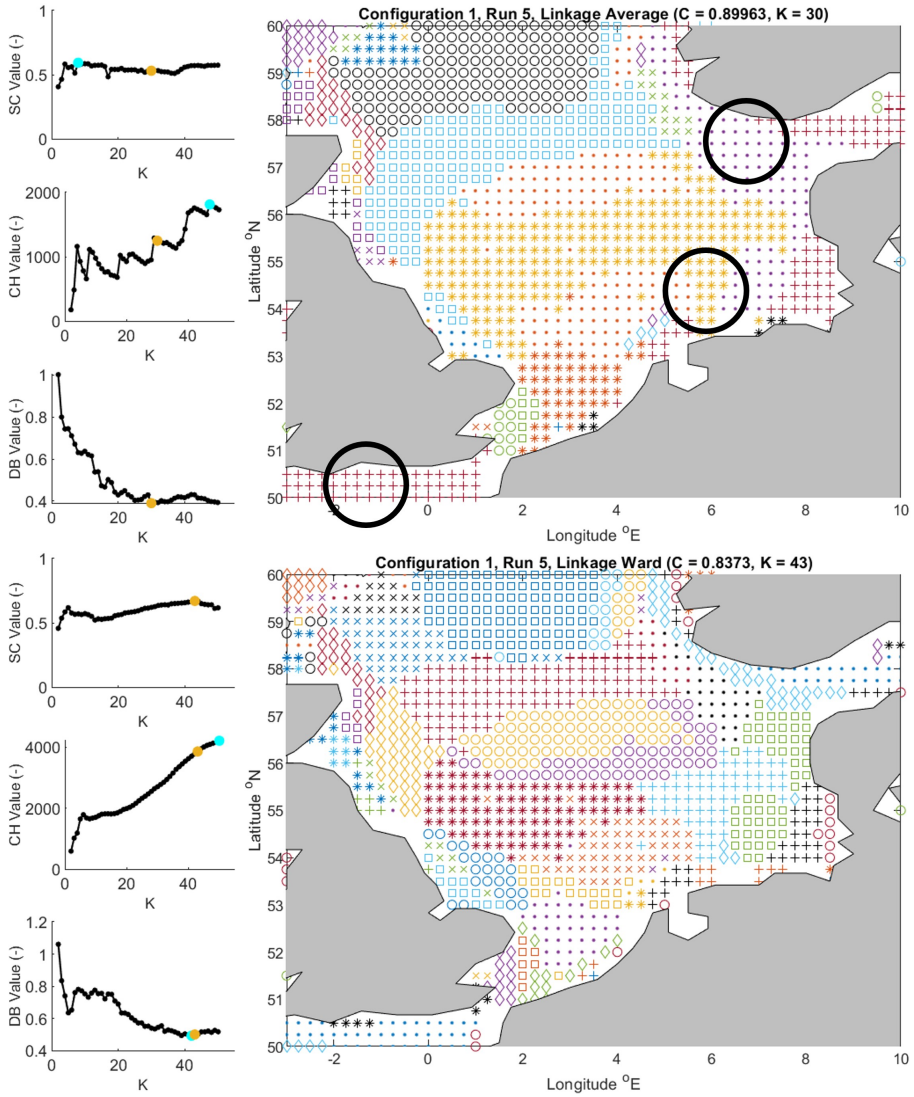


Figure F.5: Clustering results for wave parameter combination 1, weight configuration 5 and average (top) and Ward (bottom) linkage mechanisms.

- Despite the higher weights for wave direction parameters, Average linkage does not seem to place too much emphasis on the wave direction parameters. Cluster divisions are well defined by both differences in wave height and dominant wave direction.
- Ward linkage seems to place more emphasis on the wave direction. This can be noticed in the Skagerrak and to the South of the English coast. Here subdivisions are made on the wave directions where not necessarily a division would be expected.

Although, Ward linkage shows better performance for SC, CH and DB, Average linkage gives better clustering results for this configuration. This is mainly based on the established cluster divisions (see Figure F.4), the C score for Average linkage and the fact that the evaluation criteria have a clear agreement on the ideal number of clusters. The ideal value for $\mathcal{K}$ is 50.

Configuration 1.5

In configuration 1.5, most weight is assigned to the wave direction parameters compared to $H_{m0,P99.5}$. Analyzing the clustering results (see Figure F.5), the following observations can be made:

- When comparing the linkage mechanisms, it can be observed that Average linkage produces the better dendrogram based on C followed by Ward linkage. Looking at SC, CH and DB, Ward

linkage performs much better for both SC and CH, whereas Average linkage gives better results for DB.

- For both Average and Ward linkage, there is much disagreement between the ideal number of clusters, making it difficult to find an ideal number of clusters. Especially for Average linkage, it can be observed that $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ show a very wide range in their respective ideal number of clusters.
- For Average linkage, it can be observed that there is a lot of emphasis placed on the division based on the dominant wave directions rather than the wave height. Especially towards the south of the English coast, and near the coast of Norway, one might expect a subdivision based on $H_{m0,P99.5}$, which cannot be observed in Figure F.5.
- The clustering map for Ward linkage is very comparable to the clustering map for ward linkage in configuration 3. Some good distinctions based on both the dominant wave directions and significant wave height can be observed. Still, there is the in-explainable cluster division in the central North Sea.

Based on the findings for this clustering configuration, it can be stated that Ward linkage is the better approach for this configuration, due to its superior clustering output, better performance for the evaluation metrics and the better agreement between the ideal number of clusters.

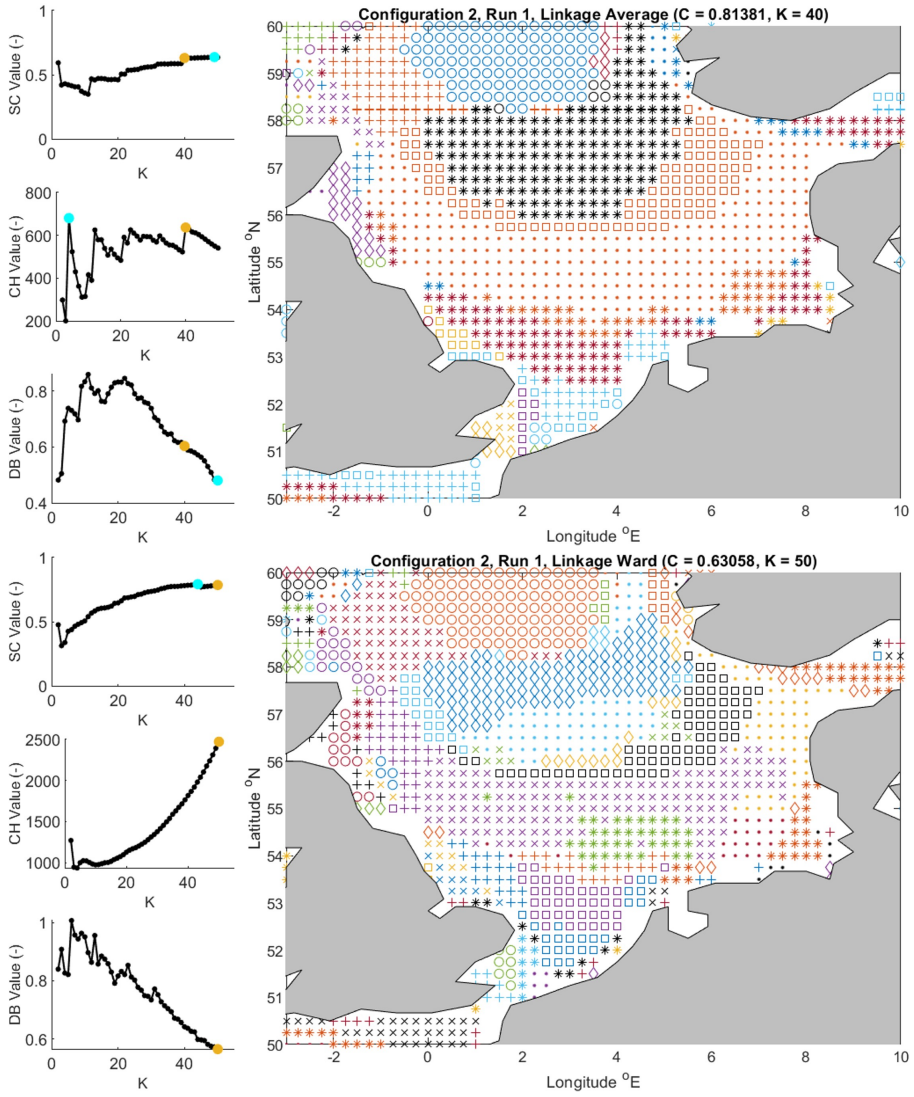


Figure F.6: Clustering results for wave parameter combination 2, weight configuration 1 and average (top) and Ward (bottom) linkage mechanisms.

F.2. Combination 2: Significant Wave Height and Peak Period

The third parameter combination includes the dominant direction parameters (θ_1, θ_2, N_θ) as well as the significant wave height ($H_{m0,P99.5}$) and peak period (T_P) for waves exceeding the 99.5 percentile level for significant wave height.

Configuration 2.1

In configuration 2.1, most weight was assigned to $H_{m0,P99.5}$ and T_P. See Figure F.6 for the clustering results. Whilst analyzing configuration 2.1, the following observations were made:

- Comparing the values for C between the different linkage mechanisms shows that Average gives the best performing dendrograms followed by Complete linkage. However, Ward outperforms Complete linkage regarding SC, CH and DB, thus it was decided to analyze Ward instead. Regarding the internal evaluation criteria, Ward performs much better for SC and CH, with CH being twice as high for Ward linkage. DB shows better results for Average linkage.
- For both Average and Ward linkage, there is much disagreement regarding the ideal number of clusters. This makes it difficult to find an ideal number of clusters for this configuration. Especially for Average linkage, it can be observed that $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ show a very wide range. Especially CH shows a lot of instability for varying values of $\mathcal{K}$.

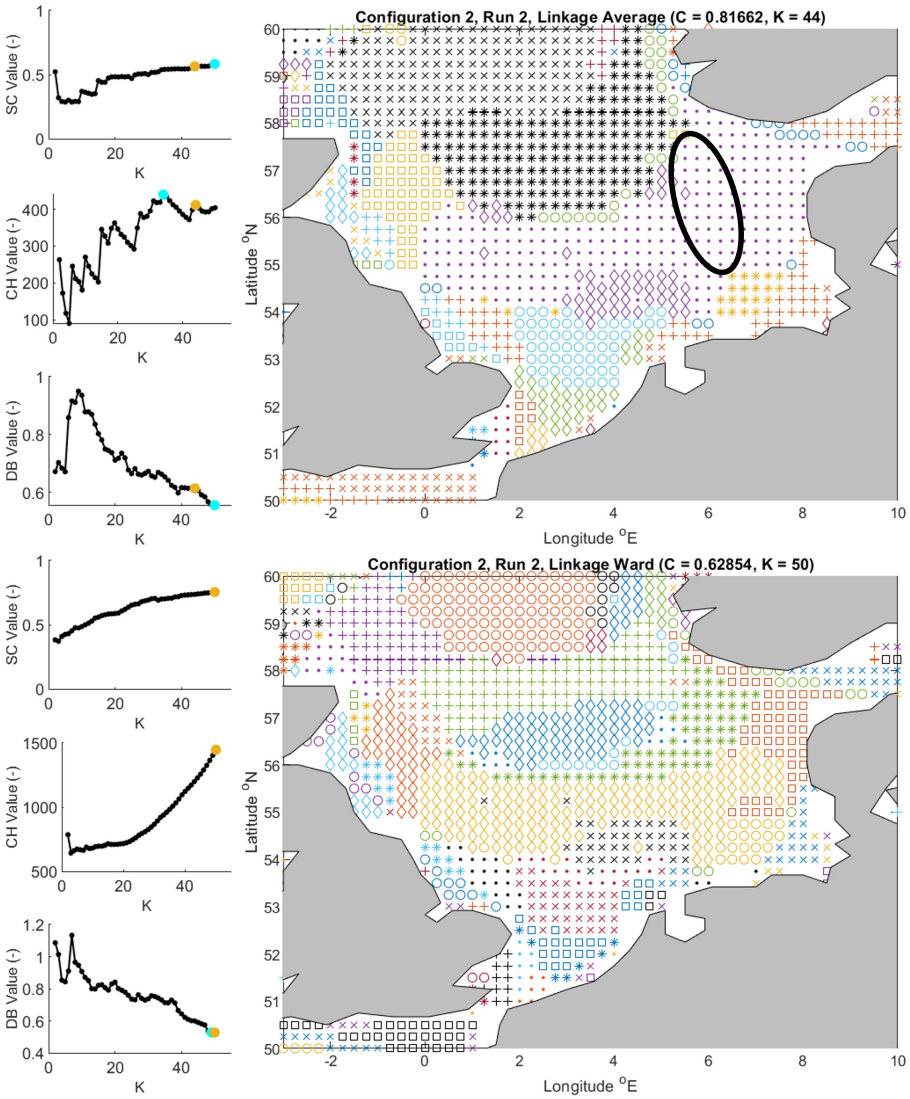


Figure F.7: Clustering results for wave parameter combination 2, weight configuration 2 and average (top) and Ward (bottom) linkage mechanisms.

- In Figure F.6, it can be seen that Average linkage places emphasis on T_P and $H_{m0,P99.5}$. This is not surprising given the assigned weight distribution. However, for an optimal clustering solution, one would prefer more cluster divisions based on the wave direction parameters.
- For Ward linkage, it can be seen that already more cluster divisions based on wave directions can be identified. This is comparable with results found for configurations 1.1.

Based on the findings for this clustering configuration, it is decided that Ward linkage with $\mathcal{K} = 50$ is the best clustering solution. This is mainly due to the superior scores for CH and SC and the better cluster divisions found in the clustering maps. Although it must be said that neither linkage mechanism scores particularly well for this weight configuration.

Configuration 2.2

In configuration 2.2, weight was shifted from $H_{M0,P99.5}$ and T_P to the wave direction parameters. See Figure F.7 for the clustering results. Whilst analyzing configuration 2.2, the following observations are made:

- Comparing the values for C shows that Average and Complete linkage give the best performing dendrograms. Once more, Ward outperforms Complete linkage regarding SC, CH and DB,

thus it was decided to further investigate Ward instead. Comparing the internal evaluation criteria shows that Ward performs much better for SC and CH. average linkage shows better values for DB.

- The values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ show that there still is not that much agreement regarding the ideal number of clusters for this configuration. The values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ for Ward linkage show preference for 50 clusters. For Average linkage there is still a lot of disagreement regarding the ideal number of clusters.
- Even though more weight is added to the wave direction parameters, it can be seen that Average linkage places emphasis on cluster divisions based on T_P and $H_{m0,P99.5}$ (see Figure F.6). This is not surprising given the weight distribution. However, for an optimal clustering solution, one would prefer more cluster divisions based on the wave direction parameters as well.
- For Ward linkage, it can be seen that already more cluster divisions based on wave directions can be identified.

Based on the findings for this clustering configuration, it is decided that Ward linkage with $\mathcal{K} = 50$ is clearly the best clustering solution. This is mainly due to the superior scores for the internal evaluation metrics SC, CH, the better agreement for the ideal number of clusters and the better cluster divisions found in the clustering maps.

Configuration 2.3

In configuration 2.3, all parameters have been assigned equal weight. See Figure F.7 for the clustering results. Whilst analyzing the results for configuration 2.3, the following observations were made:

- Based on the values for C, it is clear that Average and Ward linkage give the best performing dendrograms. Further analysis of the internal evaluation metrics, pointed out that average and ward linkages also performed best for SC, CH and DB. Comparing Average and Ward linkage shows that Ward linkage has better values for SC and CH. Average linkage and Ward linkage show similar values for DB.
- When comparing the values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$, it appears that there is better agreement over the ideal number of clusters for Average linkage. Also, CH is not stable for higher numbers of clusters. Ward linkage on the other hand shows much better results in that $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ are much closer together. Also, SC and DB are more stable for higher number of clusters making it easier to find an optimal clustering solution.
- Even though equal weight is distributed among all parameters, it can be seen that Average linkage put emphasis on divisions based on T_P and $H_{m0,P99.5}$ (see Figure F.6). However, For a useful clustering division, one would prefer more cluster divisions based on the wave direction parameters as well. Once again the cluster divisions West of the Danish coast seems to be problematic for average clustering.
- For Ward linkage, it can be seen that already more cluster divisions based on wave directions can be identified.

Based on the findings for this clustering configuration, it is decided that Ward linkage with $\mathcal{K} = 50$ is the best clustering solution. This is mainly due to the superior scores for CH and SC and the better cluster divisions found in the clustering maps. combined with the fact, because the values $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ lie closer together for Ward linkage than for Average linkage, we can be certain that this is the best clustering solution for this configuration.

Configuration 2.4

In configuration 2.4, more weight has been assigned to the wave direction parameters than on $H_{M0,P99.5}$ and T_P. The clustering results can be found in Figure F.9. About the clustering results for this configuration, the following comments can be made:

- The values for C show tat Average and Ward linkage produce the best dendrograms. Comparing Average and Ward linkage shows that Ward linkage shows better values for SC and CH. Average linkage shows better values for DB.

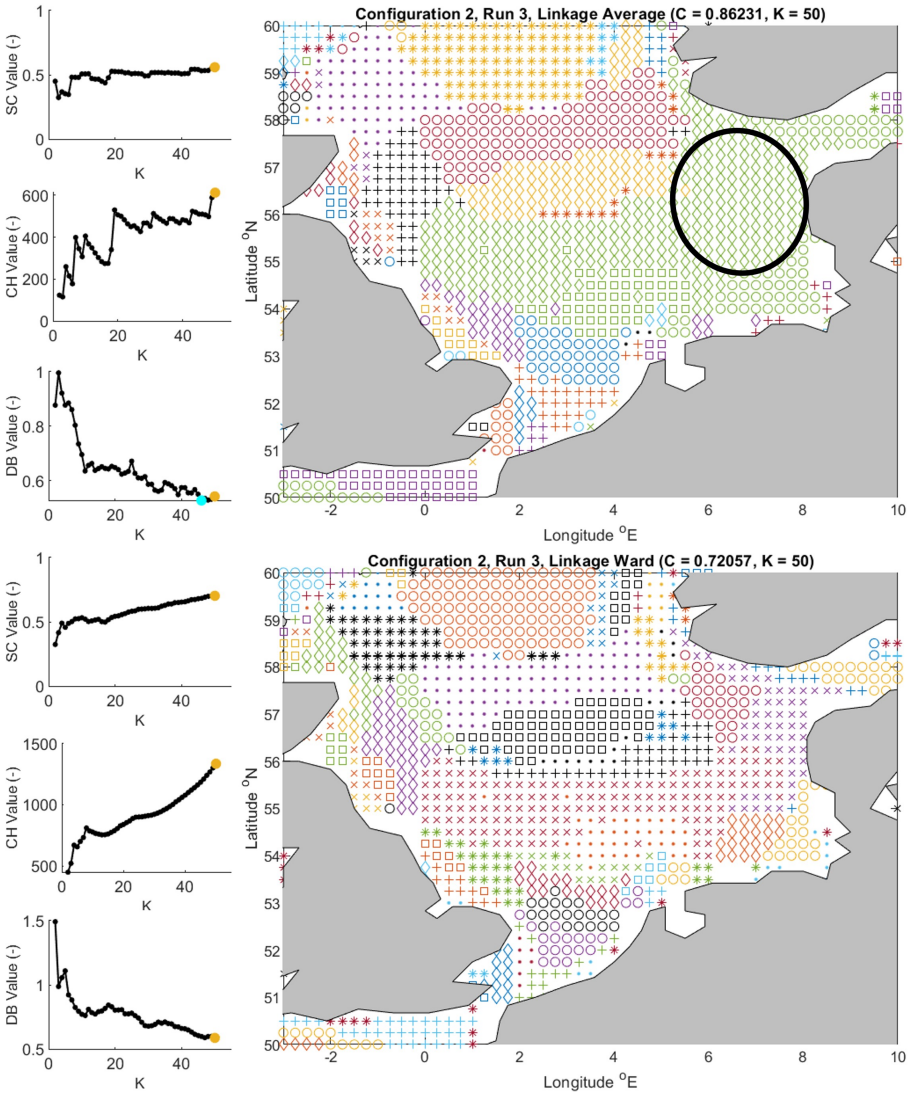


Figure F.8: Clustering results for wave parameter combination 2, weight configuration 3 and average (top) and Ward (bottom) linkage mechanisms.

- When comparing the values for $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$, It became apparent that there is a lot of disagreement on the ideal number of clusters, especially for Average linkage. Also both CH and DB are not stable for higher numbers of clusters. Ward linkage shows more agreement regarding the number of clusters. Especially $\mathcal{K}_{SC}$ is much lower than $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$. However, both DB and SC are more stable for higher number of clusters. For $\mathcal{K} = 50$, the value for SC is almost the same as for $\mathcal{K}_{SC}$.
- Average linkage shows to place high emphasis on cluster divisions based on the dominant wave directions. This is comparable with what was found for configuration 1.4. This can be noticed for points of the coast of Norway and to the South of Norway (see Figure F.9). Here one would expect cluster divisions based on $H_{m0,P99.5}$. For a useful clustering division, one would prefer more cluster divisions based on $H_{M0,P99.5}$ and T_P as well.
- Ward linkage shows better cluster divisions based on dominant wave directions and $H_{M0,P99.5}$ and T_P. However, still some emphasis seems to be placed on the dominant wave directions. This is especially apparent south of Norway, at the entrance of the Skagerrak. Here cluster divisions are made based on dominant wave directions, while based on Figure 4.8, this does not seem necessary.

Based on the findings for this clustering configuration, it is decided that Ward linkage with $\mathcal{K} = 50$ is

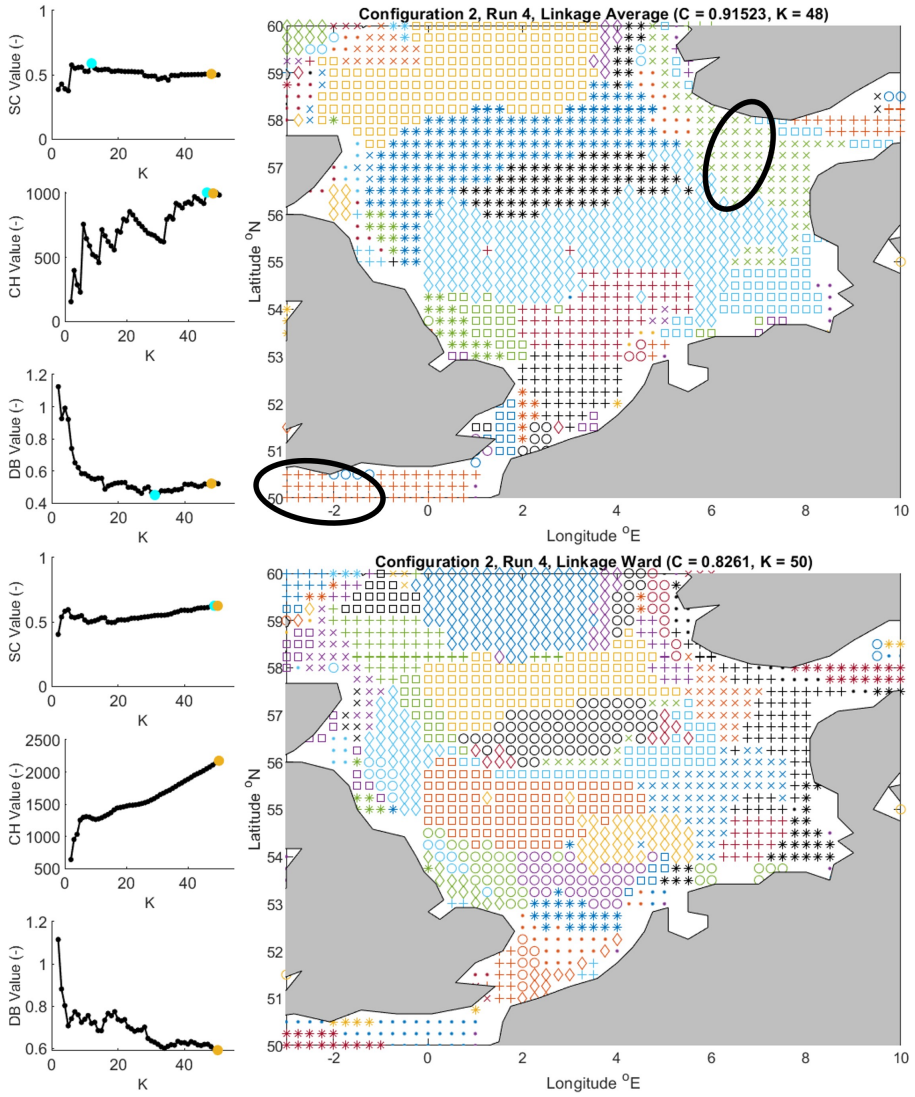


Figure F.9: Clustering results for wave parameter combination 2, weight configuration 4 and average (top) and Ward (bottom) linkage mechanisms.

the best clustering solution. This is mainly due to the superior scores for CH and SC and the better cluster divisions found in the clustering maps. combined with the fact, because the values $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ lie closer together for Ward linkage than for Average linkage, we can be certain that this is the best clustering solution for this configuration.

Configuration 2.5

In configuration 2.5, most weight was assigned to the dominant wave directions. See Figure F.10 for the clustering results. Whilst analyzing the clustering results, the following observations were made:

- Comparing the values for C between the different linkage mechanisms shows that Average and Ward linkage give the best dendrograms. Average and ward linkages also gave the best scores for SC, CH and DB. Comparing Average and Ward linkage shows comparable results for SC. Ward scores better for CH and Average linkage shows better values for DB.
- For both Average and Ward linkage, there is much disagreement regarding the ideal number of clusters. This makes finding an ideal number of clusters difficult. Especially for Average linkage, $\mathcal{K}_{SC}$, $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$ show a very wide range in their respective ideal number of clusters. Also, CH and DB are unstable for higher values of $\mathcal{K}$. Ward linkage also shows disagreement for the ideal number of clusters. Especially $\mathcal{K}_{SC}$ is much lower than $\mathcal{K}_{CH}$ and $\mathcal{K}_{DB}$. However, both DB

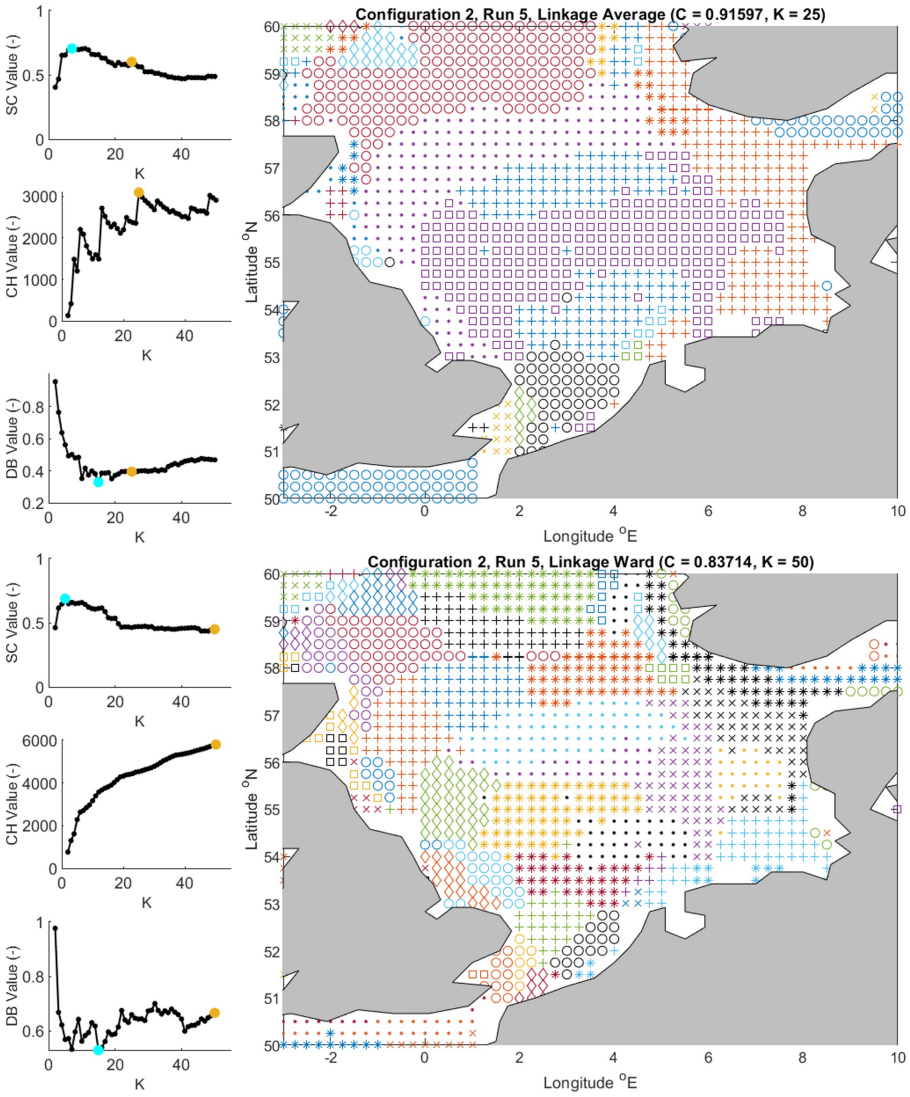


Figure F.10: Clustering results for wave parameter combination 2, weight configuration 5 and average (top) and Ward (bottom) linkage mechanisms.

and SC are more stable for higher values of $\mathcal{K}$.

- Average linkage places a lot of emphasis on cluster divisions based on dominant wave directions. This becomes especially apparent in the Eastern parts of the North Sea, where only divisions based on the dominant wave directions can be observed. While one would clearly expect a division based on $H_{m0,P99.5}$ and T_P as well.
- Ward linkage shows better cluster divisions based on dominant wave directions and $H_{M0,P99.5}$ and T_P. However, still some emphasis seems to be placed on the dominant wave directions. This is especially apparent south of Norway, at the entrance of the Skagerrak. This is comparable with the clustering results found for configuration 2.4.

Based on the findings for this clustering configuration, it is decided that Ward linkage with $\mathcal{K} = 50$ is the best clustering solution. This is mainly due to the superior scores for CH and SC and the better cluster divisions found in the clustering maps. Although it must be said that neither linkage mechanism scores particularly well for this weight configuration.