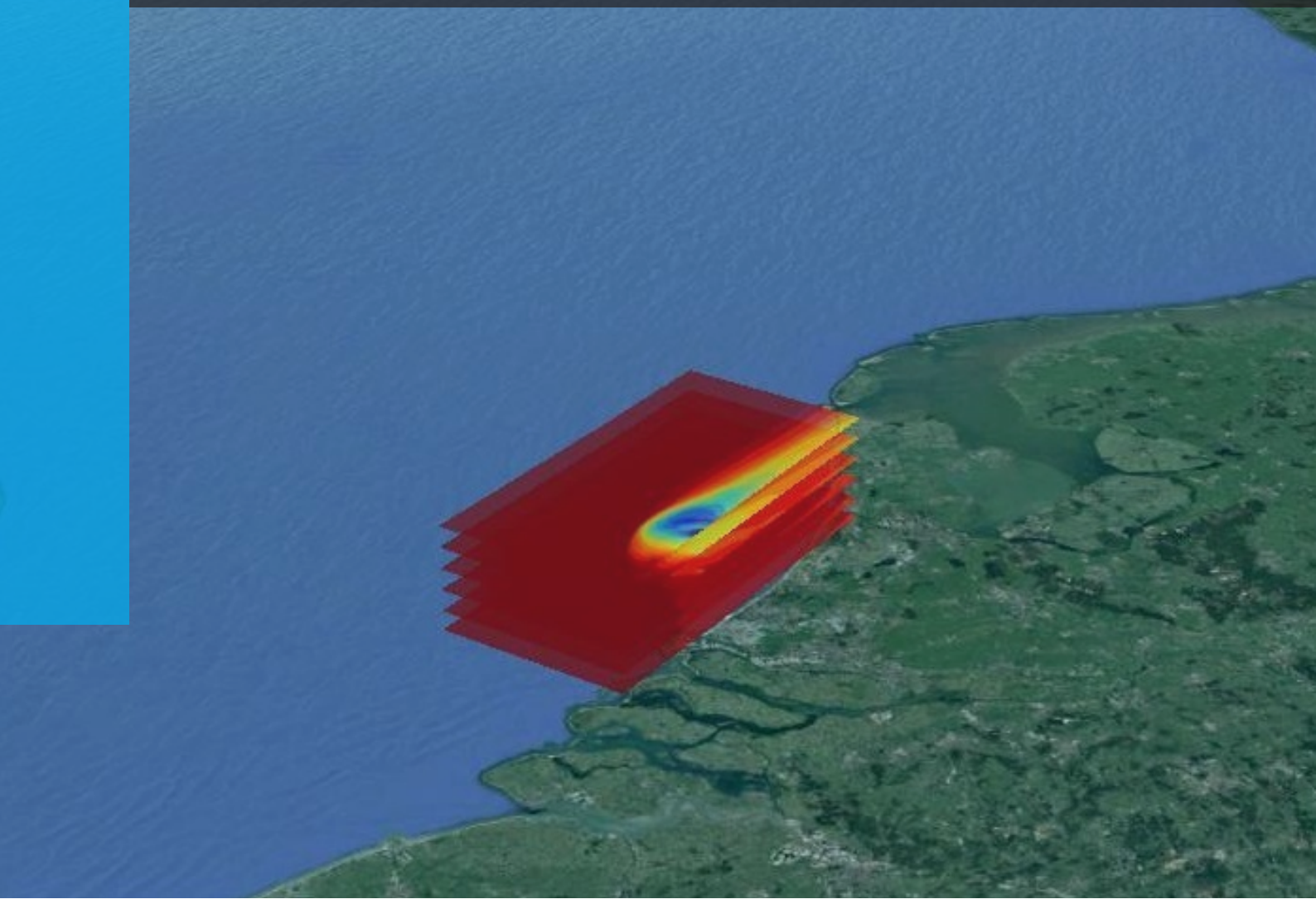
The influence of fresh water discharge on coastal water levels

By a combination of physical and probabilistic models

B. van Es

Technische Universiteit Delft



TU Delft

Deltares
Enabling Delta Life

The influence of fresh water discharge on coastal water levels

By a combination of physical and probabilistic models

by

B. van Es

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on THursday May 3, 2018 at 13:15 PM.

Student number:	4415574	
Project duration:	August 14, 2017 – May 3,2018	
Thesis committee:	Dr. ir. O. Morales Napoles,	TU Delft, Comittee chairman
	Prof. dr. ir. M. Verlaan,	Deltares and TU Delft, Supervisor
	Prof. dr. J. D. Pietrzak,	TU Delft

TU Delft

Deltares
Enabling Delta Life

Preface

This thesis is the final step in my student career at TU Delft. The aim of this thesis is to estimate the size and hydraulic effect of fresh water plumes induced by river outflow to coastal waters for worldwide applications. Due to computational limitations this effect can not be included in the current numerical models. Therefore, these insights are potentially essential in order to improve worldwide water level predictions.

Through a combination of data analysis, a numerical model and -experiment, parameterization and statistical analysis insights have been gained in the complex behavior of fresh water river plumes, compound floods and the effect of salinity on water levels. A lot of people helped me in obtaining the skills that I have needed to write this thesis over the past twenty seven years. Therefore, I would like to thank a few people and I would like to start with the members of my thesis committee.

First, I would like to thank the chairman of my thesis committee, Oswaldo Morales-Napoles. After finishing the course Probabilistics for Hydraulic Engineering in my second quarter of my master, Oswaldo was already excited about my graduation. Although, it was still a whole year away. Oswaldo, your support on the statistics and guidance throughout this thesis are well appreciated.

Second, I would like to thank my supervisor at Deltares and committee member, Martin Verlaan. Your initiation of the thesis topic gave me the opportunity to work on this beautiful project. Furthermore, Martin found a position as graduate student for me at Deltares. Martin, your supervision and guidance during this thesis were very helpful. On a regular basis we have had very interesting discussions, which were very helpful.

Third, I would like to thank my professor from civil engineering, Julie Pietrzak. Julie, your experience with numerical modeling has been essential in the development of the idealized numerical model. Furthermore, you have shown great enthusiasm for this project during the progress meetings and it was a honor to work with you.

Furthermore, my colleagues at Deltares have helped me developing the numerical model. Therefore, I would like to thank Firmijn Zijl and members of the Delft3D-FM software team: Sander van der Pijl, Herman Kernkamp, Arthur van Dam and especially Mohamed Nabi, who helped me a lot with the boundary conditions of the idealized model. Furthermore, I have had the pleasure of working with the other (graduate) interns at Deltares, thanks for the good times we have had together.

Last but most definitely not least, I would like to thank my parents, my sister and especially Jitske. Thank you for your love, your unconditional support and for providing me a place that I call home.

B. van Es
Delft, May 2018

Abstract

This thesis focusses on the size estimation and hydraulic effect of fresh water plumes induced by river outflow to coastal waters for worldwide applications. Due to computational limitations this effect can not be included in the current numerical models. Therefore, these insights are potentially essential in order to improve worldwide water level predictions. Through a combination of data analysis, a numerical model and -experiments, parameterizations and statistical analysis insights have been gained in the complex behavior of fresh water river plumes, compound floods and the effect of salinity on water levels.

The data analysis focuses on the location of the Rhine ROFI, the effect on water levels, the correlation structure of compound floods (the combination of coastal set-up and high river discharge) and return periods. Based on these analysis, the effect salinity on water levels is minor compared to the effect of tidal amplitude and coastal Delft3D-FM is dominant. However, the scale of the Rhine ROFI is significant and visible in for instance mean water levels. Compound flood events (coinciding high coastal water levels and - river discharges) are quite low correlated in the Netherlands, if no lag is introduced between the time series of water levels and discharge. Furthermore, the correlation becomes largest ($r = 0.24$) for a lag of 6 days if the discharge time series is lagged with respect to the water level time series. The semi-correlation r_{NE}, which is an important measure for the tail dependence, is maximum at a lag of 4 days. The effect of salinity on water levels is not well visible in the analysis of return periods. The effect of tidal amplitude (which increases towards the south of the Netherlands) and coastal shape dominate the analysis.

The numerical model has been developed to research the effect for the Rhine-Meuse river mouth, which can be modified (based on bulk parameters; discharge, width river mouth, depth, tidal amplitude and latitude), while performing up to standard. The model is able to simulate fresh water river outflow in combination with tides, in which tides are approximated by a Kelvin wave with linearized bottom friction. Great efforts have been spend on the development of this model, which resulted in additional boundary conditions programmed by the Delft3D-FM software development team. The model shows similar results compared to the Delft3D model of De Boer (2009), but this model is computationally more efficient.

The numerical model is used in a numerical experiment and especially getting boundary conditions correct for each numerical experiment was challenging. By varying the discharge, the uniform bathymetry, width of the river, salinity and tidal amplitude the influence of each parameter can be estimated. The results of the numerical experiment are used to analyze (1) the cross-shore plume width and (2) the hydraulic effect of salinity on water levels by dimensionless parameters used in previous research. The plume width is most responsive to discharge, depth and tidal amplitude. Furthermore, restarting the numerical model with a higher discharge has shown that the plume will grow over time to the expected plume width and under the presence of surges the hydraulic effect of salinity on water levels decreases (based on two numerical simulations).

In literature comparable experiments, without tidal components in water level, have been performed and scaling these results was challenging. For this reason and the fact that complex behavior in the numerical results is observed, the results of the numerical experiment have been parameterized by symbolic regression. The symbolic regression are intended for interpolation, physical analysis and prediction purposes. The parameterizations shows good results and the parameterized plume width shows a high sensitivity to tidal amplitude. Furthermore, the slope a in the hydrostatic balance equation has been parameterized. For each parameterization g' is needed and an effort was made to parameterize that parameter as well. Two methods have been assessed for obtaining the salinity difference between the river mouth and coastal waters. First an observation based method, which is capable of predicting daily averaged salinity based on random forests. Although the performance is good ($r^2 = 0.95$), this method can not predict depth-averaged salinity at the river mouth due to the input data. Therefore, a parameterization of the numerical results has been made. This is unfortunately a parameterization with dimension mismatch which is performing well ($r^2 = 0.99$. Using the last method and the parameterization for the plume width and a combination of parameterizations the performance of parameterizations for the plume width and hydraulic effect are assessed. Both parameteri-

zations are able to capture the complex behavior which is present in the results of the numerical experiments and able to produce predictions for world wide applications.

Multivariate statistical (vine) copula models have been used with two purposes; (1) to estimate uncertainty of the parameterizations and (2) for the application to compound floods. The uncertainty of the parameterization is quantified through a conditional copula sampling procedure and the uncertainty in the dataset of the river Rhine has been researched through a vine copula model. The c-vine structure has been selected by Cramér-von Misses, based on a vine copula with a limited number of copula families. The final c-vine copula model is able to reproduce the complete correlation structure and sample a matching dataset.

The ultimate goal of this thesis is to understand the influence of salinity on water levels, the relation to compound flood events and estimate the importance of this effect for global flood forecasts. The maximum observed water level different between a numerical model with salinity and without salinity (but with the same discharge, depth, salinity, tidal amplitude and width) under stationairy conditions (no weather inluence, astronomical tide and constant discharge) is 0.31 m, a result that is closely approximated by the combination of parameterizations. Furthermore, numerical for results two analysis showed that the hydraulic effect of salinity decreases under the presence of surges, which were induced by extreme wind conditions.

Contents

Nomenclature

List of symbols

α	Bottom slope
δh	Steric height anomaly
Δh	Observed water level difference
ϵ	Scaled tidal amplitude
η	Water level (Kelvin wave)
γL	Measure for the cross-shore plume width (by Garvine (1995))
κ	Linearized alongshore friction coefficient
ω	Angular frequency of the tide
$\Phi(X, \mu, \sigma)$	Cumulative distribution function normal distribution, with mean μ and standard deviation σ
π_i	Weight of ith component in gaussian mixture
ρ	Pearson correlation coefficient
$\rho_{1,2\|3,\ldots,n}$	Partial correlation between variable 1 and 2 given variables 3,. . . ,n
ρ_i	Density at location i
σ	Friction parameter
σ_X	Standard deviation of variable X
τ	Scaled inlet transport
θ	Parameter(s) for parametric distribution
$\hat{\theta}$	Maximum likely hood estimator
υ	Vertical eddy viscosity
μ_X	Mean of variable X
ω	Angular velocity of tide (M_2 or S_2)
a	Slope scaling parameters r_{rho} and r_h
AIC	Akaike criterion information
c_0	Wave speed ($= \sqrt{(gD)}$)
c_i	Internal wave speed ($= \sqrt{(g'D)}$)
$c(u_1, \ldots, u_n)$	copula probability density function (multivariate)
$C(u_1, \ldots, u_n)$	copula cumulative density function (multivariate)
CM_n	Goodness of fit - Cramér von Misses
D	Depth
f	Corriolis parameter
$f(x)$ or $f(x_1, \ldots, x_n)$	Probability distribution function
$F(x)$ or $F(x_1, \ldots, x_n)$ and $\hat{F}(x)$ or $\hat{F}(x_1, \ldots, x_n)$	Parametric- and a fitted umulative distribution function
g	Gravitational acceleration ($= 9.81\,[m^2/s]$)
g'	Reduced gravitational acceleration ($= \frac{\rho_{sw}}{\rho_i} g\,[m^2/s]$)
$h(u_1, u_2, \rho)$	Conditional copula
$h^{-1}(u_1, u_2, \rho)$	Inverse conditional copula
H_s	Significant wave height
i	Real, in-phase part of a complex number
I_r	Inlet ratio
$Im(x)$	Imaginary part of x

k	Alongshore water level variation parameter
k_0	Initial alongshore water level variation parameter
K	Kelvin number
K_i	Inlet scale in Rosby radii
m	Cross-shore water level variation parameter
N	Number of elements in variable
N_s	Number of sigma layers in numerical model
P	Pressure (air)
Q	Discharge
Q_X	Quantile function of variable X
r	Spearman rank correlation coefficient
r_{NE}, r_{SE}, r_{SW} and r_{NW}	Semi correlations (Spearman rank correlation coefficient)
r_ρ and r_h	Scaling factor density and water level respectively
r^2	Coefficient of determination
r_i	The Rosby inlet radius
$R(l)$	Mean squared error in tree prediction
$\Delta R(s,l)$	Decrease in mean squared error due to split s
Re	Reynolds number
$Re(x)$	Real part of x
R_i and S_i	Rank of value X_i and Y_i in variable X and variable Y respectively
s_{wl}	Surge
s	Normalized salinity
S	Salinity
S_i	Salinity inlet
S_{sw}	Salinity shelft waters
T_r	Return period
$\hat{Y}(x)$	Tree predictor
u	Independent random samples
u_x and u_y	Wind in x- and y-direction respectively
U_{10}	Windspeed 10 m above ground level
v	Dependent random sample
w or $w(t)$	Water level
w_t	Tidal component of water level
w_{t_0}	Tidal amplitude of water level
W	River mouth width
X and X_i	Variable X and value of X at index i
y_p	Cross-shore fresh water plume width
$\hat{Y}(x)$	Tree predictor
$\hat{Y}_{t^*}(x)$	Ensemble tree predictor/bootstrap-aggregated ensemble

List of Figures

List of Tables

1 Introduction

Floods are among Earth's most common and most destructive natural hazards. The risk of flooding is increasing worldwide due to two main factors: socioeconomic- and physical drivers. Socioeconomic drivers are for example growing population and economic developments, while on the other hand physical drivers are i.e. land subsidence and climate change (Muis et al., 2016). The last, of which climate change especially, is expected to be a large driver of increasing flood risk in the future (Hirabayashi et al., 2013; Milly et al., 2002).

The Intergovernmental Panel on Climate Change (IPCC) has raised awareness on compound events, such as compound floods, through a special report on "Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation" (Seneviratne et al., 2012). Furthermore, a recent study on compound floods (the presence of simultaneous fluvial- and coastal extremes) in the United States shows non-stationarity in the dependence between storm surge and precipitation, which could lead to more frequent compound floods in the future (Wahl et al., 2015).

With increasing risks of both physical- and socioeconomic drivers the need for accurate flood predictions is evident. After all, flood predictions can be used to give out flood warnings, evacuate people, save lives, reduce damages and initiate a flood relief. Most poor countries can't afford these high tech flood prediction systems. Therefore, the world bank, Deltares - an independent applied research institute - and others are providing and developing world wide operational flood forecasts (Verlaan et al., 2015).

1.1. Problem statement

World wide flood predictions are made real time by large numerical models, based on Delft3D-FM. These numerical models contain over 5 million grid cells and solve 2D-depth averaged linear shallow water equations. Although good flood predictions can be made by solving these equations, it also limits the physical phenomena that can be modeled. Preferably, flood predictions are made by solving 3D linear shallow water equations, since interactions in the vertical can be incorporated in the solution of the numerical model.

In order to function operationally, the numerical models which are currently used for world wide flood forecasts need to be partitioned and the use of data assimilation techniques is necessary. Therefore for this application, it is not expected that any further computational optimizations are possible which allow solving 3D linear shallow water equations in the near future.

One of the effects that plays a role in the vertical is stratification, through salinity differences. A relevant cause of salinity differences is the discharge of fresh water from rivers to coastal waters. The discharged fresh water forms a fresh water plume at coastal waters. Such a plume is known in literature as a Region Of Freshwater Influence (ROFI) (Simpson et al., 1993). Furthermore, the fresh water plume influences coastal water levels. Currently, no specific research has focused on the relevance of this effect for this application and these insights are needed for the improvement of world wide flood forecasts. Furthermore, combining this topic with research on compound floods is particularly interesting given the fact that there is a lot of overlap between these topics.

1.2. Motivation

The main motivation of this research is to increase the physical understanding of relation between salinity- and water level differences and the its relation to compound floods. These effects will be quantified with both physical (numerical) and probabilistic models. The aim of this research is to provide an understanding of the influence of density differences caused by salinity differences on the spatial variability in water levels and the exceedance probability of water levels. Hereby, it is assumed that additional research is necessary that quantification on the effect of compound floods is needed.

The research will be centralized on the mouth of the river Rhine-Meuse in the Netherlands. For this location plenty of data is available, similar numerical models (with structured grids) have been developed for this region. Furthermore, the river has the largest river discharge to the North Sea and is the second largest river in Europe. Therefore, it is expected that the hydraulic effect of salinity differences is visible at this location. All these aspects makes studying this river mouth convenient.

1.3. Research objectives

The objective of this research is to combine physical and statistical models in order to better capture the influence of this physical processes observed in compound events. A physical understanding needs to be obtained of the processes which govern the behavior of the water level, to investigate the non-linear behavior of water levels near a river mouth. The objective of the numerical model is to study the influence of salinity differences on the (spatial) variability in water levels and capture, with a limited amount of characteristics, the most important dynamics in a simple model. If all results from the numerical modeling have been obtained, the results will be parameterized. This, dimensionless, parameterization will be used in order to achieve estimates which are computationally more efficient. The main goal is to parameterize the spatial scale and the effect on water levels in such a way that these are applicable for global forecasting applications (such as the Global Tide and Surge model). Finally, the combined parameterization and statistical analysis will be used together with a traditional exceedance probability water level analysis. In the interest of all these objectives a main research question has been prepared.

Main research question

Can the effect of salinity differences on (spatial) variability in water levels near a river mouth be computed by a combination of numerical modelling and probabilistic models and reduce the amount of data needed while holding good estimates of exceedance probabilities for water levels?

In order to achieve the main objective, the following sub-questions are specified:

1. What is the exceedance probability in water levels near the mouth of the Meuse based on observations only?
2. What is the relationship between salinity differences and water levels?
3. What is the influence of salinity differences on the variability in water levels at the mouth of the Meuse compared to the variability in water levels at other locations at the North Sea?
4. How does the numerical model perform compared to real observations and is this performance sufficient for this application?
5. What is an efficient way of parameterizing the results from the numerical model for the statistical analysis?
6. Does a probabilistic model combined with a simple physical model give a better fit compared to a probabilistic model based on observations only?

1.4. Outline and reader

Chapter 1 contains an introduction to the topic and presents the aim of this research. The theoretical background is presented in chapter 2. Chapter 3 presents an extensive data analysis. The numerical model and the results of the numerical experiments are presented in chapter 4. Chapter 5 describes the prediction process and the uncertainty of these predictions. The conclusions and recommendations are presented in chapter 6.

2

Overview of literature on fresh water plumes, compound events, numerical modeling and data based methods

The literature study consists of four parts: physical processes, flood risk assessment, physical modelling and data based methods (which consist of random forests, symbolic regression and statistical analysis).

2.1. Physical processes

In order to understand the governing physical processes which are relevant for regions of fresh water influence, this section gives an overview of ROFI's, the Rhine ROFI, characterization of ROFI's. The region of freshwater influence is governed by spatial differences in salinity, therefore the last subsection presents results in world wide salinity trends.

2.1.1. Region of fresh water influence

Over the last decades a significant amount of research has been performed to understand the hydrodynamics and to model the region of freshwater influence (ROFI) near the Rhine. The term ROFI was introduced by Simpson et al. (1993). Due to these great efforts the effect of the ROFI on the coast and how it influences the current structures are quite well understood.

The presence of fresh water, which is discharged by a river, in coastal waters is the essence of a ROFI. ROFI's can mostly characterized by stratification, as shown in figure 2.1. Therefore, this term and it's implications should be well understood. In oceanography and climate research the term is often used to describe a phenomena, which is characterized by the presence of layers in water (or the atmosphere) due to differences in properties of these layers. There are multiple processes that can generate stratification. For instance freshwater from rivers, the tide and heat from the sun can generate stratification. Stratification can be broken down due to mixing by i.e. tide (with a large tidal amplitude during spring tide), waves and wind. Furthermore, high energetic conditions such as storms, where large waves and strong wind mix up the water, break down the stratification, resulting in a well mixed ROFI.

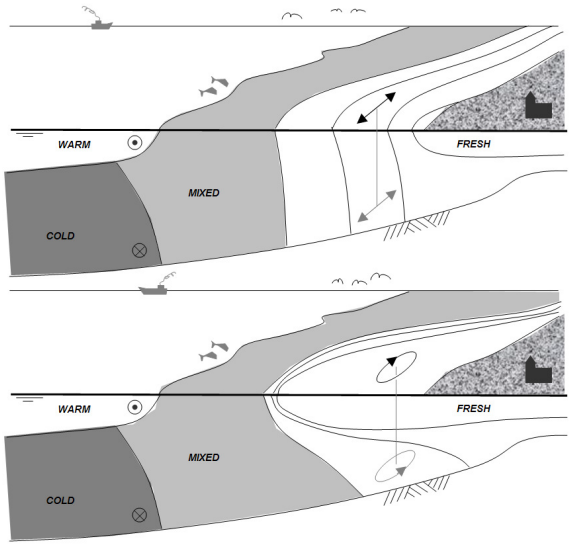


Figure 2.1: The Rhine ROFI during well-mixed conditions (top graph) and during stratified conditions (bottom graph), by De Boer (2009).

2.1.2. Rhine ROFI

This subsection gives first an overview of the different coastal water types in the North Sea, in order to understand the mixing processes on a large scale. The next paragraph gives an overview of the Rhine ROFI.

De Boer (2009) studied the spatial character of water properties at the North Sea. Several different water types can be distinguished and for this application the focus will be on the coastal waters. The coastal waters, which are characterized by the interaction of the inflow of freshwater from rivers and the saltwater from the Atlantic ocean, can be found close to the shore near the Netherlands, Germany, Denmark and the United Kingdom. Figure 2.2 shows the location of coastal waters in front of the Dutch, German and Danish coast (marked with A4, continental coastal water) and the English coast (marked with A2, English coastal water). The continental coastal waters are generated by the river outflow of freshwater at the river mouths of mainly the river Rhine/Meuse and the river Elbe (and some other much smaller rivers and outlets). Respectively, the English coastal waters are created by the river Thames and some smaller rivers.

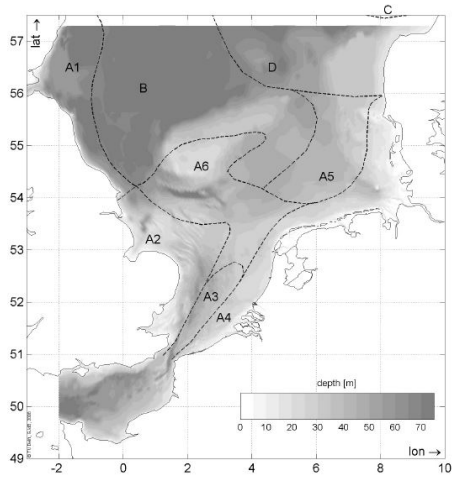


Figure 2.2: The coastal water types of the North Sea, by De Boer (2009)

The presence of stratification in the Rhine ROFI can be charaterized by differences in salinity, temperature and, a property which is strongly related to the previous, density. The outflow of freshwater from the river Rhine and Meuse causes a fresh water plume at the North Sea, which is periodically stratified due to the tidal structure present at this location (De Boer, 2009; Simpson et al., 1993). Research by Simpson et al. (1993) found that during spring tides the whole ROFI is well mixed and during neap tides (and low energetic conditions) the ROFI is stratified. Due to the rotation of earth outflowing freshwater from the river Rhine/Meuse is deflected to the North. Research by de Ruijter et al. (1997) found that edge the plume is at a distance of 20 to 40 km crossshore and 100 km alongshore from the river mouth, however the position of the ROFI is of course dependent on (1) the freshwater outflow of the river Rhine and (2) the wind and wave conditions at the North Sea and (3) the tidal amplitude. The scale of this plume and the region of stratification is shown in Figure 2.3. This figure shows the position of the Rhine ROFI under avarage and extreme conditions (Giessen et al., 1990).

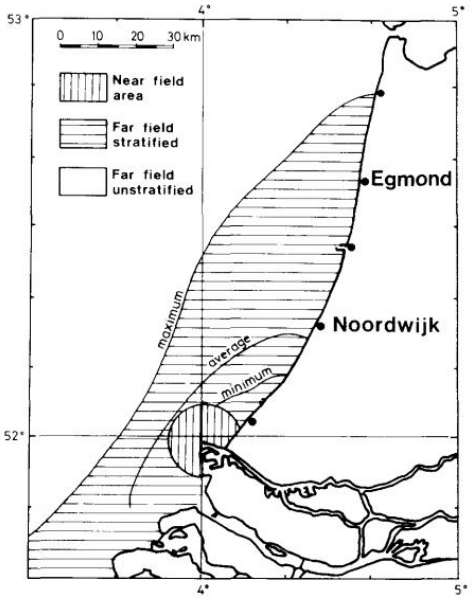


Figure 2.3: Extent of the stratified area under average and extreme conditions by Giessen et al. (1990)

2.1.3. ROFI characterization

This subsection gives an overview of ROFI characterizations that have been used in literature.

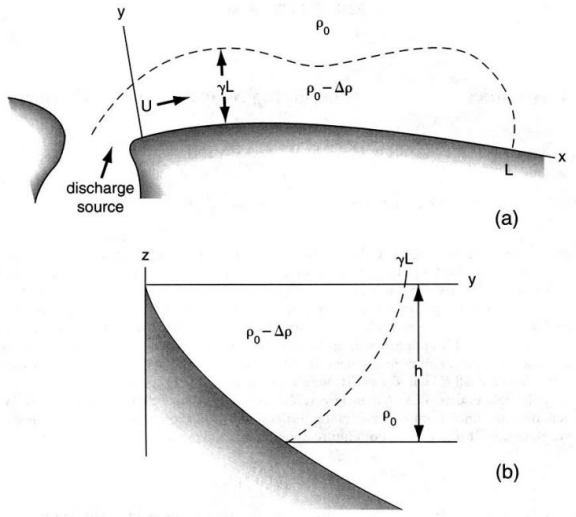


Figure 2.4: Parameters for the dynamical classification system of buoyant coastal discharges by Garvine (1995)

Over the years many efforts have been made to understand general characteristics of river plumes. Garvine made a dynamical system for classifying buoyant coastal discharges (Garvine, 1995), which states that the Kelvin number (K) (eq. 2.1 and figure 2.4) offers a consistent system for the classification of these fresh water river discharges. In these equations r_i is the internal Rosby radius, c_i is the internal wave speed, g' ($= \frac{\rho_{sw}-\rho_i}{\rho_{sw}} g$) is the reduced gravity, f is the Coriolis parameter and γL a measure for the cross-shore plume width, as denoted in figure 2.4.

$$K = \frac{\gamma L}{r_i} \qquad with \quad r_i = \frac{c_i}{f} \qquad c_i = \sqrt{g'D} \tag{2.1}$$

However, this system needs a priori knowledge of the systems observed characteristics and a system that requires only bulk properties of the river mouth (such as discharge, width, depth and salinity) is preferred. In a later study Garvine (1999) made a great effort to parametrize the observed characteristics of a buoyant coastal discharge based on the bulk properties of the river mouth (Garvine, 1999).

In that particular research a parametrization has been made using non-dimensional numbers, such as the scaled inlet transport (τ), inlet scale in Rosby radii (K_i), scaled tidal amplitude (ϵ), the bottom slope (α), the inlet ratio and the inlet Reynolds number (Re). In these equations W is the width of the river mouth, D is the coastal depth and w_{t_0} the tidal amplitude.

$$\begin{aligned} \tau &= \frac{2Qf}{g'D^2} & K_i &= \frac{W}{r_i} & \alpha &= bottomslope \\ \epsilon &= \frac{w_{t_0}}{D} & I_r &= \frac{W}{D} & Re &= \frac{Dc_i}{v} \end{aligned} \tag{2.2}$$

Salinity is normalized in this paper. The shelf waters have a salinity (S_{sw}) and the river has different, lower, salinity (S_i). The used normalization of salinity is shown in eq. 2.3 and figure 2.5. Furthermore, this normalization is used to describe the cross- and alongshore length of the plume. The cross- and alongshore plume length are determined, in Garvine's and this research, as the distance between the river mouth and $s = 0.05$.

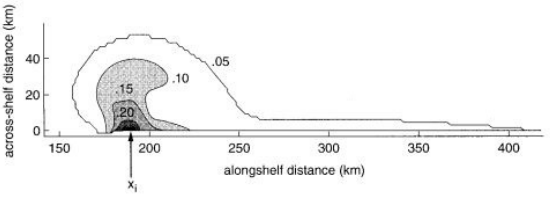


Figure 2.5: Normalization of sality by Garvine (1995)

$$s = \frac{S_{sw} - S}{S_{sw} - S_i} \tag{2.3}$$

The paper shows that the parametrization is made based on no tides and the influence of tides has been studied in a later phase. Therefore, the interaction of tides on the plume has not been incorporated in the parametrization. However, many (later) researches found an interaction between tides and stratification. For example De Boer (2009) found that, due to the influence of spring and neap tides, periodic stratification can be observed in the Rhine ROFI.

2.1.4. Salinity trends

Salinity levels are measured nowadays by measurement stations, boats and even satellites. Furthermore salinity levels have been monitored all around the globe and several researches have focused on analyzing these results. The focus in these researches is on analyzing the current data sets and making predictions on future worldwide salinity levels.

Based on these researches it is expected that salinity in the oceans will change due to climate change. The process of changing salinity values due to climate change is a slow process, however slight changes in salinity are already visible. Based on several researches it is expected that waters have currently a high salinity will get even higher salinity values. This mechanism is label as “rich get richer”. Furthermore, this is a clear sign of an intensified global water cycle.

Regions with high salinity levels can be found near the equator, i.e. warm regions. Gobally, in oceans, salinity levels tend to decrease with increasing latitude. For this application, the North Sea, which has not extremely high salinity values, is therefore expected to encounter a slight drop in salinity values due to climate change (Durack and Wijffels, 2010; Durack et al., 2012).

2.2. Flood risk

This section gives an overview of flood risk estimation, compound events, compound flood analysis in the Netherlands and methods for assessing compound events. Theoritical background on (multivariate) extreme value analysis can be found in section 2.4.

The focus in this chapter is on compound effects, since this is potentially very relevant for the hydraulic effect of salinity differences and fresh water plume size estimation. Furthermore, nearly all accepted analysis for climate- and weather extremes, which are relevant for flood risk analysis, focus on the extreme value of a single variable. However, as stated by the IPCC: "*there is no direct relation between the extremeness of a single a climate- or weather variable and the impact it causes*" (Seneviratne et al., 2012). Kumar et al. (2013) have shown that the impact of a extreme event is not directly related to the extremeness of a single climate- or weather variable in most cases.

In literature there are several cases in climate and weather related research (i.e. cyclones, weather and climate, influence of weather on the risk of wild fire and coinciding storm surges (as defined in appendix C) and high river discharge) in which a combination of multiple variables are not extreme by itself and still causes an extreme situation (Flannigan et al., 2009; Marin, 2010; Rappaport, 2000; Svensson and Jones, 2002; Van den Brink et al., 2005). Therefore, a method is necessary for assessing extreme situations that detects all relevant conditions that lead to significant impact (Fischer and Knutti, 2013).

2.2.1. Compound events

Compound events are defined by the IPCC as: *“(1) two or more extreme events occurring simultaneously or successively, (2) combinations of extreme events with underlying conditions that amplify the impact of the events, or (3) combinations of events that are not themselves extremes but lead to an extreme event or impact when combined. The contributing events can be of similar (clustered multiple events) or different type(s).”* (Seneviratne et al., 2012).

In a compound event framework review compound events are defined as followed: *“a compound event is an extreme impact that depends on multiple variables or events. The variables or events must have a causal relationship to the impact.”* (Leonard et al., 2014). Dependence between variables is obvious when the variables are characterized by deterministic equations. A different approach is needed if variables are characterized empirically (i.e. regression equations). Leonard et al. (2014) state that *“casual reasoning is required (whether by mutual external drivers or latent variables); otherwise non-sense correlations might be used to spuriously explain the impact.”*.

2.2.2. Compound flood analysis in the Netherlands

Historically the dependence between storm surge and high river discharge at the river Rhine was not taken into account in calculating flood risks (De Quay, 1967). The conclusion that it is unlikely that there could be dependence between those variables is drawn after a study on weather patterns in the Netherlands.

Quite recently, in 2012, a near flooding with a coinciding pluvial flooding and storm surge conditions occurred in the province of Groningen (Hurk et al., 2015). This event triggered the debate whether storm surge and high discharges are dependent and of statistical significance to flood risk estimation.

A slightly older statistical analysis shows a slight positive correlation between high sea water levels at the North Sea and river discharges at the Rhine (Geerse, 2013). Furthermore, this research suggests that taking into account the correlation between high sea water levels at the North Sea and river discharges at the Rhine will increase the water level indicated in the Dutch flood protection norms. Moreover, in a different study a meteorological data analysis shows dependence between storm surge and high sea water levels at the North Sea (Kew et al., 2013). The study shows the difference in the multivariate probability distribution function (PDF) for the correlated data and an independent data set. It should be noted that both studies state that it is uncertain whether the correlation observed in the tails of discharge and sea surges is the same correlation as observed in complete dataset, in other words it is not clear whether tail dependence is observed. A study by Klerk et al. (2015) found however that "*the dependence between high sea water levels and high river discharges is strongest for a discharge lagged by 6 days with respect to the water level series at Hoek van Holland* ".

2.2.3. Methods for assessing compound flood events

As previously explained it is essential to take compound flood event into account for flood risk analysis. In literature compound flood events have been analyzed with several methods: statistical methods, physical modelling or a combination of both. The statistical methods such as copulas, Bayesian Networks (BNs) and bivariate extreme value models have been applied in order to quantify the effect of compound flood events (Gutierrez et al., 2011; Lian, 2013; Zheng et al., 2015). Furthermore, physical modeling methods have been applied (Kew et al., 2013) in order to quantify the effect of compound flood events. Combinations of these methods have been applied as well by using a joint probability approach and a simple hydrodynamic model (Zhong, 2013). Within this thesis a complex physical model (3D numerical model) will be combined with a complex statistical model (vine-copulas).

2.3. Physical modeling

Physical modeling will be done with Delft3D-FM. This program solves, in 2D (depth averaged) or 3D, non-linear shallow water equations. These equations are derived by the Boussinesq approximations and the Navier-Stokes equations for incompressible free surface in shallow water (Kernkamp et al., 2011). Fluid motions are described by the Navier-Stokes equations and these equations can be seen as Newton's second law for motion in fluids.

The shallow water approximations are based on the assumption that the horizontal length scale in the equations is much larger than the vertical length scale. This leads to a reduction of the vertical momentum equation in to the hydrostatic pressure relation. The vertical accelerations are not taken into account based on the assumption that those accelerations are small compared to the gravitational acceleration. The last assumption is the Boussinesq approximation. This approximations ignores density differences, except the terms that are multiplied by the gravitational acceleration term. In Delft3D-FM the pressure term takes the effect of variable density into account, since the vertical turbulent exchange coefficients and the horizontal pressure gradients take density differences in the vertical into account.

Another advantage of Delft3D-FM is the ability to handle structured and unstructured grids within one model and due to this ability, the program is able to handle local refinements in the grid which is particular useful for the application in this thesis. Additionally, the orthogonality of the grid is of importance. For this application orthogonality is not of importance, due to the use of only rectangular cells and quads refinement with triangular border. In other words, all used grids are completely rectangular which ensures perfect orthogonality.

2.4. Data based models

In this thesis, some data based model are used for analyzing the complex behavior of fresh water river plumes. These data based models consist of Random forests, symbolic regression and statistical methods.

2.4.1. Random forests

Random forests are a type of ensemble learning, which is amongst others suitable for regression and classification. Random forests are also known as bagged tree ensemble or bootstrap aggregation regression ensembles and were invented by Breiman (1996).

Regression trees

Let x be the vector containing all information for the predictors (x_i) and all possible vectors are captured in the n-dimensional covariate space X. Furthermore, y is the to be predicted value given predictor x. Under the assumption that the training data set covers the complete solution space, the basis in constructing of regression trees is formed by repeated algorithmic partitioning of the covariate space. Through algorithmic division of the covariate space a decision tree is formed, which is usually suitable for classification. Figure 2.6 shows a regression tree and the principle of partitioning the (for this case n-dimensional) covariate space X.

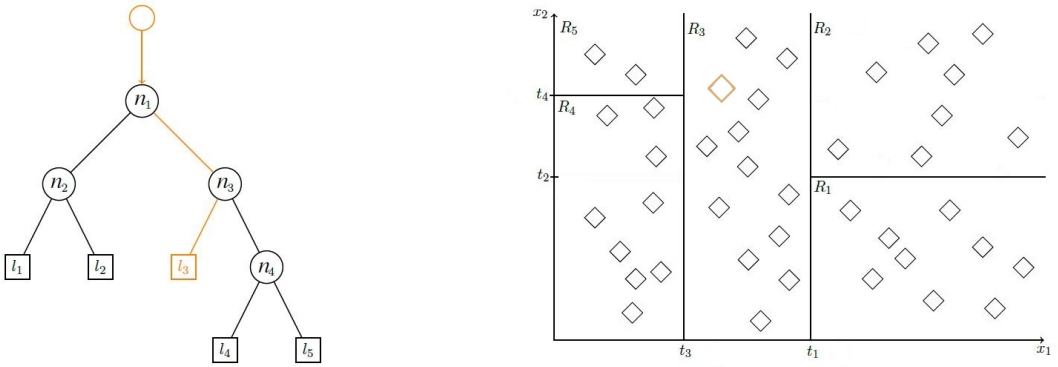


Figure 2.6: The left figure shows a regression tree with nodes $(n_1,\ldots,n_4)$ and leafs $(l_1,\ldots,l_5)$ and the right figure shows the partitioned covariate space, with regression surfaces $R_1,\ldots,R_5$

The regression tree evaluates node after node until a leaf is reached. Every leaf corresponds with a regression surface $(R_1,\ldots,R_5)$. Therefore, the regression tree adds a dimension by assigning a predicted value to each regression surface. At the nodes the following evaluations are made:

- n_1: $X_1 \leq t_1$
- n_2: $X_2 \leq t_2$
- n_3: $X_1 \leq t_3$
- n_4: $X_2 \leq t_4$

Algorithmic division of the covariate space

Partioning of the of the covariate space into two descendant subsets forms the basis of the algorithmic division of the covariate space. During construction of the tree the tree-nodes (n_i) are considered temporary leafs at the moment of splitting. The choice for each node is based on three methods, which are shown below. The method for determining a prediction at each (temporary) leaf (1) is used in the final tree predictor $Y_t(x)$ and the method to select a split at every intermediate node (2). The procedure is stopped using the method for stop splitting (3).

1. Method for determining a prediction at each (temporay) leaf
2. Method to select a split at a (temporay) leaf
3. Method for stop splitting

1. Method for determining a prediction at each (temporay) leaf
The method for determining a prediction at each (temporary) leaf is based on a simple regression. The tree predictor $\hat{Y}(x)$ predicts at leaf l based on the mean over the data points y_i which are included by x in leaf l, as shown in eq 2.4. In this equation $N(x)$ is the number of data points x_i which are in l.

$$\hat{Y}(x) = \frac{1}{N(x)} \sum_{x_i \in l} y_i \tag{2.4}$$

2. Method to select a split at a (temporay) leaf
In order to explain the method to select a split at a (temporay) leaf first the method for computing the error should be understood. The errors are computed by the mean squared error as shown in eq. 2.5, in which $\widetilde{L}$ is the collection of all temporary nodes l at the moment of splitting. The mean squared error is then given:

$$\begin{aligned} R(l) &= \frac{1}{N} \sum_{l \in \widetilde{L}} \sum_{x_i \in l} \left(y_i - \hat{Y}(x) \right)^2 \\ &:= \frac{1}{N} \sum_{l \in \widetilde{L}} R(l) \end{aligned} \tag{2.5}$$

If S are all possible splits at l and s splits the temporary leaf l in l_L and l_R, then the mean squared error of the total tree is decreased by:

$$\Delta R(s, l) = R(l) - R(l_L) - R(l_R) \tag{2.6}$$

Now the optimal split, s^*, is selected by maximizing the decrease in mean squared error of the total tree.

$$\Delta R(s^*, l) = \max_{s \in S} \Big(\Delta R(s, t) \Big) = \min_{s \in S} \Big(R(l_L) + R(l_R) \Big) \tag{2.7}$$

3. Method for stop splitting
The splitting procedure is stopped if $R(l)$ reaches a predefined threshold or new nodes contain less observations than a certain predefined leaf size (in other words, $N(l)$ in eq 2.4 is smaller than a predefined leaf size).

Bootstrap-aggregated ensemble
By sampling n observations with replacement from the dataset (this dataset is denoted t^*) and creating regression trees based on the algorithmic partitioning method mentioned above predictions can be made through each of these regression trees. By averaging each prediction from the regression trees (the ensemble) a final prediction is made. Equation 2.8 shows this process, with N bagged trees and each individual tree ($Y_{t_i^*}(x)$) is trained by the dataset t_i^*, which is resampled with replacement from dataset t.

$$\hat{Y}^*(x) = \frac{1}{N} \sum_{i=1}^{N} \hat{Y}_{t_i^*}(x) \tag{2.8}$$

2.4.2. Symbolic regression

Symbolic regression is not only able to calculate parameters in an equation, but also able to construct the equation itself. In literature it is commonly used to find explicit and differential equations, however even implicit equations can be found with this technique. Furthermore, it is able to detect the underlying invariant equations and physical conservation laws (Schmidt and Lipson, 2009). For the symbolic regression the software package Eureqa is used, which is gaining popularity among researchers with more than 850 publications to date. According to Schmidt and Lipson (2009) unlike (non-)linear regression the method is based on *"evolutionairy computation for searching the space of mathematical expressions while minimizing various error metrics"*.

In this type of regression equations are formed through binary trees, by randomly selected mathematical building blocks. These mathematical building blocks are predefined and can consist for example of constants, state variables, mathematical operators $\{+, -, \div, *\}$ and/or functions (for example $e^x, log(x)$). Figure 2.7 gives an overview of a binary tree. Of course, by randomly selecting building blocks and growing binary trees the number of possible equations is infinite and grows exponentially. However, through evolutionary computations (an term that is elaborated in the next paragraph) it is possible to find the most effective equations and limit the number of possibilities.

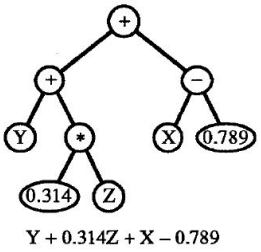


Figure 2.7: Binary tree example by Koza (1994)

The process of growing binary trees consists of extension and mutation. At the start of the analysis the solutions, fitness predictors and the trainers are randomized. The program evaluates the performance of individual solutions in parallel. After these evaluations the best performing solutions are selected and new solutions are generated by extending and mutating their binary trees. This process is infinite, repeated until convergence is reached and this process is defined as the evolutionary process.

2.4.3. Statistical analysis

As discussed earlier, modelling dependency in climate related variables for extreme value estimation is of importance. Determined by the variable structure and the question at hand a copula or a Bayesian Network (BN) can be used to model dependency between variables. The use of copulas is relatively new in climate change analysis (Schoelzel and Friederichs, 2008) and the use of copulas and BN's is growing in the field of Hydraulic engineering. Ultimately vine copulas will be used in the analysis and these statistical models are used to model multivariate dependency structures based on marginal distributions. This subsection will first explain the obtainment of the pseudo observations by the fitting process and fit evaluation, then a brief introduction to copulas is given and finally the method of vine copulas is described.

Distributions

In order to perform a statistical analysis datasets are fitted to parametric distributions. These parametric distributions can be used to approximate for example return periods or the probability of occurrence (or exceedance) of a certain event. Furthermore, this step is essential for further steps in the statistical analysis, analyzing the dependency structure. For this application all datasets are fitted to several parametric probability distributions[1] by maximum likelihood estimators (mle). Mle's $\hat{\theta}$ find parameters θ for a given parametric distribution, by maximizing the likelihood for these parameters by:

$$\hat{\theta} \in (\arg\max_{\theta \in \Phi} L(\theta, x)) \tag{2.9}$$

Using the parametric distributions and their mle, the equations of the pdf $f(x)$ and cdf $F(x)$ are known. Some basic conventions are given in eq. 2.10. For further reference the notation $u_i = \hat{F}(X_i)$ will be used.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(x)\,dx \\ F(x) &= P(X \leq x) \end{aligned} \tag{2.10}$$

Mixed gaussian models

Apart from parametric distributions, mixed gaussian models are used in the fitting process. Mixed gaussian models are a summation of several weighted normal distributions (with $F(X, \mu, \sigma) = \Phi(X, \mu, \sigma)$, the weighted normal distributions have g components), as shown in eq. 2.11. Gaussian mixtures are fitted using an Expectation Maximization algorithm (McLachlan and Peel, 2004), a standard method of fitting gaussian mixtures in Matlab. The pdf of a gaussian mixture is presented on the next page:

[1] Beta, Binomial, Birnbaumsaunders, Burr, Discrete Uniform, Exponential, Extreme Value, Gamma, Generalized Extreme Value, Generalized Pareto, Geometric, Half-Normal, Inversegaussian, Logistic, Loglogistic, Lognormal, Nakagami, Negative Binominal, Normal, Poisson, Rayleigh, Rician, Stable, t-location scale, Uniform and Weibull distributions are used in the fitting procedures for this application

$$
\begin{aligned}
&F(X) = \sum_{i=1}^{g} \pi_i \Phi_i(X, \mu_i, \sigma_i) \\
&0 \leq \pi_i \leq 1 \quad (i = 1, \ldots, g) \\
&and \quad \sum_{i=1}^{g} \pi_i = 1
\end{aligned}
\tag{2.11}
$$

Choosing the correct number of components is essential for the performance of the mixture model. According to McLachlan and Peel (2004): "Testing for the number of components g in a mixture is an important but very difficult problem which has not been completely resolved". For this application, density estimation, based on the AIC the number of components will be estimated. This method does not underestimate the true number of components, asymptotically (Leroux, 1992). Based on a combination of the first local minima for number of components to AIC and a manual check the number of components is selected.

Distribution evaluation
The Akaike Information Criterion (AIC) can be used to compare fits to data from different statistical models, which is a convenient way to compare the fits of distributions to data (Mutua, 1994). The AIC value in itself is only meaningfull if a comparison is made between AIC values of fits of other distributions. Comparing a distribution fit between different datasets is not possible with this method. The lower the AIC value, the better the fit of that particular distribution for the dataset. The formula for the AIC is given by equation 2.12, in which $length(\theta)$ is the number parameters of the fitted distribution and L is the likelyhood value of the distribution fit. Therefore, the AIC is a measure for relative goodness of fit, it measures both goodness of fit and model simplicity. By ordering the distributions by AIC (from low to high) the distribution that is most representative to that dataset is chosen.

$$
AIC = 2length(\theta) - 2log(L(\theta, X)) \tag{2.12}
$$

Correlation
In order to take the dependency between multiple variables into account in a statistical analysis the correlation between these variables needs to be calculated. Correlation is most commonly calculated by the Pearson correlation coefficient (denoted by: ρ_p) (Kurowicka and Cooke, 2006). The correlation coefficient is a basal coefficient, indicating the likely hood of the combined occurrence of high, and low, values in two datasets. The correlation coefficient can, in that sense, be compared to a mean or a standard deviation for 1D datasets.

Another measure for quantifying correlation is the Spearman rank correlation coefficient (denoted by: r) (Hollander et al., 2013), which is shown in equation 2.14. In this equation the correlation between variable X and Y is calculated. The Spearman rank correlation is calculated in the same manner as the Pearson correlation coefficient, but only with the ranks (denoted in the equations by R_i and S_i for X_i and Y_i respectively and n observations) of the dataset and not the actual samples of the dataset. The (p% and $(1-p)$%) confidence bounds, obtained by the Fisher transformation, of the Spearman rank correlation coefficient (r_{conf}) are given by eq. 2.15. The advantage of a rank correlation is that the correlation coefficient stays the same if any or all of the variables are transfered by a monotone increasing function (and this happens in a transformation from variable x to u, which is equal to $F(x)$). For further reference, if correlation is mentioned, rank correlation r is meant.

$$
\rho_p = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} = \frac{E\big([X - \mu_X][Y - \mu_Y]\big)}{\sigma_X \sigma_Y} \tag{2.13}
$$

$$
r = \frac{cov(Rank(X), Rank(Y)}{\sigma_{Rank(X)} \sigma_{Rank(Y)}} = 1 - \frac{6 \sum_{i=1}^{n} \big(R_i - S_i\big)}{n(n^2 - 1)} \tag{2.14}
$$

$$
r_{conf} = tanh\Big(atanh\big(r\big) \pm \frac{\Phi(1 - \frac{p}{2}, 0, 1)}{\sqrt{n-3)}}\Big) \tag{2.15}
$$

Copulas

In statistics copulas are used to model the dependence between variables. Through Sklar's theorem, the joint CDF of variables $x_1,\ldots,x_n$ can be written as $F(x_1,\ldots,x_n) = C(u_1,\ldots,u_n)$. In this notation the copula C is n-dimensional and u_i as the pseudo observation ($u_i = F_i(x_i)$). Under the assumption that copula C is unique and the marginal distributions are continuous the multivariate pdf can be decomposed according to eq 2.16, where c is the copula density.

$$f(x_1,\ldots,x_n) = f_1(x_1)\cdot,\ldots,\cdot f_n(x_n)\cdot c(u_1,\ldots,u_n) \tag{2.16}$$

Copulas are continuous models for the marginal distributions of each variable. Several copula families can be fitted to multivariate datasets. The fitting process of a multivariate distribution (or copula family) is similar to fitting a univariate distribution. As expected, evaluation of the quality of the fit is similiar; through several types of goodness-of-fit. However, the complexity of these applications increases obviously with an increasing number of variables.

Vine copula

The use of multivariate copula is restricted by the fact that one copula family has to be used for all variables. Therefore, the use of a multivariate copula doesn't ensure, for all variables, a proper fit and the dependency structure is not well represented by a single copula family. In order to cope with this limitation vine copulas (or pair copula constructions) can be used. Vine copulass, are able to apply a different copula family for different pairs of variables. This way there is more flexibility to find a copula family that fits good for each pair of variables.

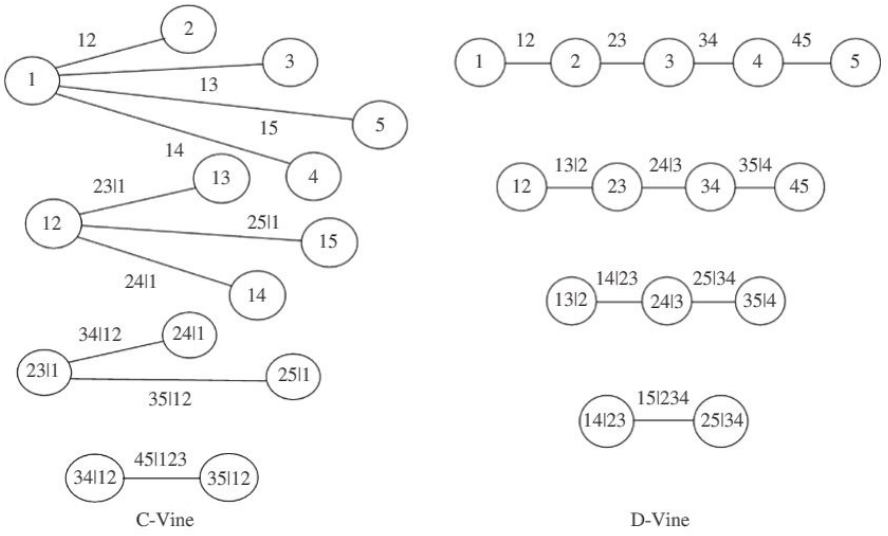


Figure 2.8: Graphical representation of a C- (left) and a D-vine (right), by Zhang et al. (2015)

Vine structures can be structured according to the work of Bedford and Cooke (2001) and Aas et al. (2009). Canocial (or c-) vines and D-vines are most regularly used in dependent variable sampling, these structures are shown in figure 2.8. The corresponing density functions are shown in eq. 2.17 (c-vine) and 2.18 (d-vine)

$$f(x_1,\cdots,x_n) = \prod_{k=1}^{d} f_k(x_k)\cdot\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{j,(i+j)|1,\cdots,(j-1)}\left\{F(x_j|x_1,\cdots,x_{j-1}),F(x_{i+j}|x_1,\cdots,x_{j-1})\right\} \tag{2.17}$$

$$f(x_1,\cdots,x_n) = \prod_{k=1}^{d} f_k(x_k)\cdot\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,(i+j)|(i+1),\cdots,(i+j-1)}\left\{F(x_i|x_{i+1},\cdots,x_{i+j-1}),F(x_{i+j}|x_{i+1},\cdots,x_{i+j-1})\right\} \tag{2.18}$$

Goodness of fit

In order to assess the performance of a several vine copula models goodness of fit will be applied. For this application Cramér-von Mises will be applied. For this method the squared difference between the emperical and the fitted copula family is calculated, as shown in eq. 2.19. Comparing these value to seperate models gives an overview of the performance of the copula family.

$$CM_n = n\sum_{i=1}^{n}\left\{C_n\left(\frac{R_i}{n+1},\frac{S_i}{n+1}\right) - C_{\theta_n}\left(\frac{R_i}{n+1},\frac{S_i}{n+1}\right)\right\}^2 \tag{2.19}$$

Sampling

The sampling process for vine copula can become complicated, since the correlations between multiple variables has to be accounted during sampling. The sampling procedure is followes the tree structure of the vine. The following example will be a d-vine as shown in figure 2.8 (right graph) with Gaussian copula families (others are possible, but not all formulas for each distribution is given in this thesis).

The sampling procedure for five dependent variables (v) is shown in eq. 2.20 and starts by sampling five independent random variables $u_1,\ldots,u_5$ (uniform[0,1]). The first variable (v_1) is equal to u_1 and the second variable (v_2) is computed through the inverse conditional copula ($h(u_1,u_2,\rho)$, see eq. 2.22 for Gaussian conditional copula), exactly the same as for a bivariate copula. The third variable (v_3) is more complex, the correlation between the first variable and the second variable needs to be correct. This is achieved by the partial correlation (the correlation between v_1 and v_3 given v_2) and the conditional copula between these variables, then the correlation between variable v_2 and v_3 is calculated. This process, with partial correlations and conditional copulas, is continued for variable v_4 and v_5, as shown in eq. 2.20. As shown in these equation, the process gets significantly more complex with an increasing number of variables.

$$\begin{aligned}
v_1 &= u_1\\
v_2 &= F^{-1}_{r12;x_1}(u_2)\\
v_3 &= F^{-1}_{r23;x_2}\left(F^{-1}_{r13|2;F_{r12|x_2}(x_1)}(u_3)\right)\\
v_4 &= F^{-1}_{r34;x_3}(F^{-1}_{r24|3;F_{r23|x_3}(x_2)}(F^{-1}_{r14|23;F_{r13|2;F_{r23;x_2}(x_3)}F_{12;x_2}(x_1))}(u_4)))\\
v_5 &= F^{-1}_{r45;x_4}(F^{-1}_{r35|4;F_{r34|x_4}(x_3)}(F^{-1}_{r25|34;F_{r24|3;F_{r34;x_3}(x_4)}(F_{r23;x_3}(x_2))}(a)))\\
a &= F^{-1}_{r15|234;F_{r14|23;F_{24|3;F_{r23;x_3}(x_2)}(F_{r34;x_3}(x_4))}(F_{r13|2;F_{r23;x_2}(x_3)}(F_{r12;x_2}(x_1))}(u_5)
\end{aligned} \tag{2.20}$$

Partial correlation

The partial correlation is calculated through equation 2.21, which becomes iteratively for correlations given multiple variables.

$$\rho_{1,2|3,\ldots,n} = \frac{\rho_{1,2|3,\ldots,n-1} - \rho_{1,n|3,\ldots,n-1}\rho_{2,n|3,\ldots,n-1}}{\sqrt{1-\rho^2_{1,n|3,\ldots,n-1}}\sqrt{1-\rho^2_{2,n|3,\ldots,n-1}}} \tag{2.21}$$

Conditional copula

The conditional is formally defined as the partial derivate to u_2 as shown in eq. 2.22, furthermore the Gaussian conditional copula is given by eq. 2.23. The conditional copula for different copula families are of course defined, but not shown in this report.

$$h(u_1,u_2,\rho) = \frac{\partial C(u_1,u_2,\rho)}{\partial u_2} \tag{2.22}$$

$$h(u_1,u_2,\rho) = \Phi\left(\frac{\Phi^{-1}(u_1,0,1) - \rho\Phi^{-1}(u_2,0,1)}{\sqrt{1-\rho^2}},0,1\right) \tag{2.23}$$

The inverse conditional copula for the Gaussian copula is defined in eq. 2.24. For some copula families it is not possible to estimate the inverse conditional copula. If that's the case a numerical approximation has to be used.

$$h^{-1}(u_1,u_2,\rho) = \Phi\left(\Phi^{-1}(u_1,0,1)\sqrt{1-\rho^2} + \rho\Phi^{-1}(u_2,0,1),0,1\right) \tag{2.24}$$

3

Spatial data analysis, correlation structure and return periods

First, the spatial patterns of the Rhine ROFI are researched in the first section. The second section shows the correlation between salinity, water levels and discharge close to the Meuse-Rhine river mouth. The return periods on water levels and surges along the dutch coast are researched in the third section. In this section the relationship between the Rhine ROFI and the return periods are researched.

Through out this chapter data sets are used for a variety of data analysis, all data used in these data analyses is publicly available via Rijkswaterstaat and can be found at live.waterbase.nl. Rijkswaterstaat is a Dutch government agency of the Ministry of Infrastructure and Environment, which monitors biological and hydraulic parameters as one of its activities. Rijkswaterstaat obtains data by performing tests and measurements. Furthermore, a dataset on discharges has been obtained through PhD candidate D. Paprotny (Centre, 2017).

3.1. Spatial analysis

A first insight regarding water levels and salinity trends can be obtained by studying the spatial patterns near the mouth of the river Rhine-Meuse. First the salinity is researched and second the water levels.

3.1.1. Spatial salinity analysis

The salinity data is measured in the period 1975 until 1985 ('75-'85) is converted by Rijkswaterstaat to practical salinity unit (Tables and Standards, 1981). The locations of the salinity observations are shown (with magenta) in figure 3.1. The observations are done by taking samples at surface level with a boat and these samples are taken on average over all locations 1.8 times a month.

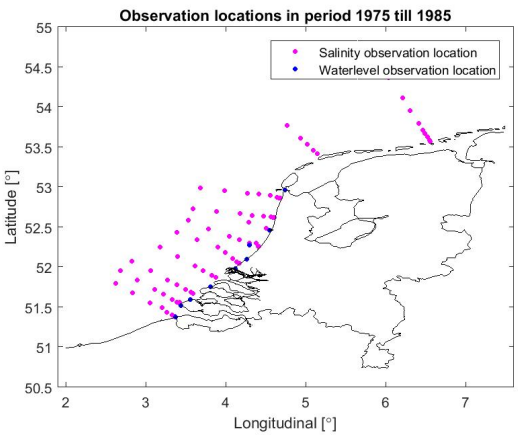


Figure 3.1: Locations of salinity (shown by magenta dots) and waterlevel (shown by blue dots) observations

The salinity analysis consists of the calculation of the mean and standard deviation at every measurement location in the period 1975 until 1985. An overview of the observation locations and the periods of observation per observation location are shown in appendix C. The mean is calculated by equation 3.1 and the standard deviation is calculated by equation 3.2, in these formulas X_i is the value of variable X at index i. These formulas are quite straight forward and won't need any further explanation. By projecting the results on a map, spatial patterns can be analysed.

$$\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{3.1}$$

$$\sigma_X = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\mu)^2} \tag{3.2}$$

The results of the salinity analysis, which are shown in figure 3.2, are created by a natural neighbour interpolation (Sibson, 1981) to visualize the spatial character of the results. The figure shows that the Rhine ROFI is characterised with a significant lower mean and a higher standard deviation in salinity at surface level. This lower mean and higher standard deviation is of course caused by the fact that these observations are taken at water level surface and carried out close to the outflow of the river. The river generates an outflow of fresh water at the North Sea and this outflow of fresh water is visible in these statistics.

Suijlen and Duin (2002) have made a spatial analysis for mean salinity over the same time period using almost identical observation locations using a interpolation method based on triangulation. The results from figure 3.2 are very similar to the figure made by Suijlen and Duin (2002) (Chart 2) and the fact that these results are so alike says a lot about the certainty of the used method applied in this analysis.

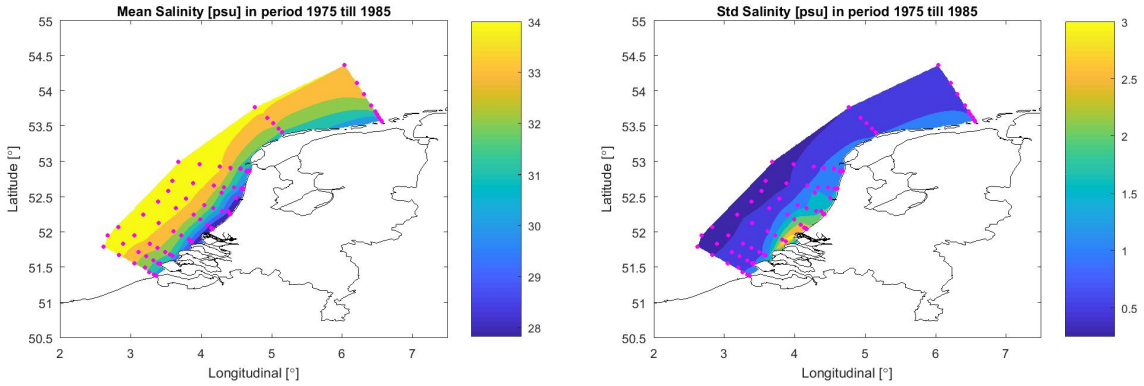


Figure 3.2: The left figure shows the mean salinity in period 1975 untill 1985 and the right figure shows the standard deviation. Locations of salinity observations (shown by magenta dots) and the contourplots are created by a natural neighbour interpolation method (Sibson, 1981)

3.1.2. Spatial water level analysis

The same kind of analysis, as the spatial salinity analysis, has been carried out for the water level analysis. Figure 3.1 shows (marked with blue) the locations of the water level observations at the North Sea, for which data is available in the period '75-'85. The unit of water level data is cm relative to a standard Dutch reference level, which is called N.A.P. The method for this analysis is the same as the method salinity analysis. The mean and standard deviation have been computed for each water level observation location and the results are visualized using natural neighbour interpolation (Sibson, 1981). The results of the analysis are shown in figure 3.3.

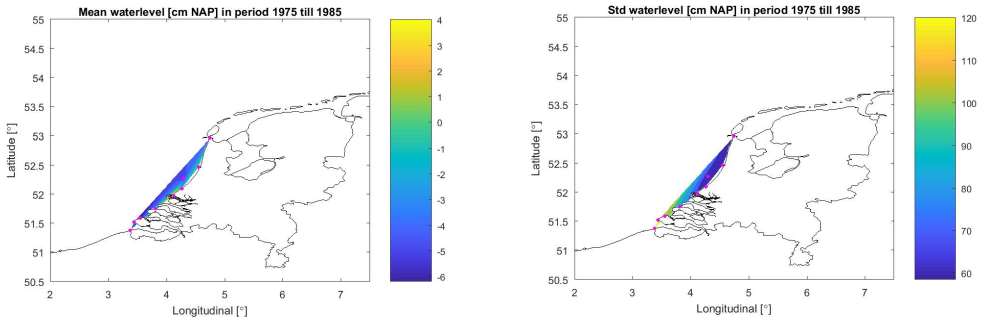


Figure 3.3: Locations of salinity (shown by magenta dots) and waterlevel (shown by blue dots) observations. Contour plots are created by a natural neighbour interpolation method (Sibson, 1981).

Due to the fact that the observation locations for water levels are more or less on a straight line along the coastline spatial patterns are not easily visualized and detectable. Table 3.1 shows the results of the data analysis and these results show that at Hoek van Holland (at the mouth of the Rhine-Meuse) the mean is by far the highest. Figure 4 does not show these results graphically due to the applied interpolation method and the scarcity of observation locations. However, the results from table 3.1 are in line with the finding of Vinogradov and Ponte (2010), close to the river mouth of rivers with a high discharge a higher mean water level is observable.

Table 3.1: Mean and standard deviation per location [cm (NAP)], locations are ordered from North to South.

Location	Mean [cm]	Std [cm]
Den Helder	0.87	57.71
IJmuiden buitenhaven	4.49	61.07
Noordwijk Meetpost	1.31	60.59
Scheveningen	2.36	65.53
Hoek van Holland	9.48	66.72
Euro platform	2.58	61.14
Lichteiland Goeree	2.82	70.93
Brouwershavensche Gat 08	1.91	89.31
Westkapelle	-0.79	119.33
Cadzand	-1.60	129.15

New time period

In order to visualize the higher mean near the river mouth of the Rhine-Meuse a second time period has been analysed, 2001 till 2016. Data availability in this period is better for water levels, but the data spatial resolution for salinity data is less good. The results for the water level analysis are shown in figure 3.4. The results for the salinity analysis are not shown in this thesis, but these show similar results as the spatial salinity analysis of course with a lower spatial resolution due to a lower number of salinity observation locations. This graphical representation clearly shows the higher mean water level near the river mouth. The figure also shows in the standard deviation plot very clearly the increase in tidal amplitude near the south of the Netherlands by an increasing standard deviation towards the south of the Netherlands.

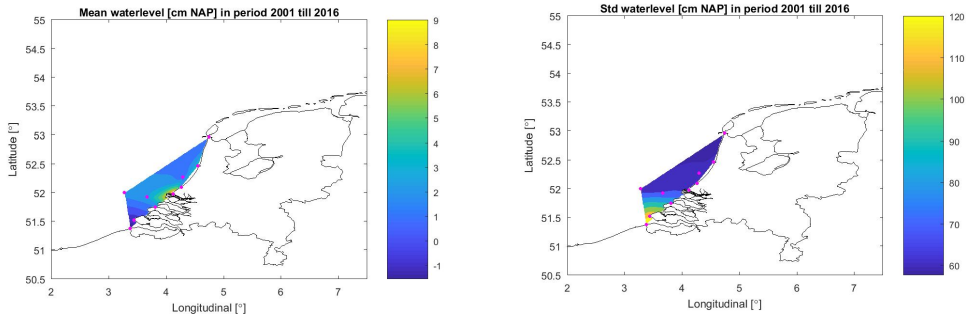


Figure 3.4: Locations of salinity (shown by magenta dots) and water level (shown by blue dots) observations. Contour plots are created by a natural neighbour interpolation method (Sibson, 1981).

3.2. Correlation between salinity, water levels and discharge

In order to understand the relationship between salinity, (mean) water levels (and surges) and discharge the dependence structure has been researched for these variables. For this analysis data from the spatial analysis has been used. By studying the dependence structure a better understanding of the governing processes can be obtained.

3.2.1. Correlation between salinity and mean water levels

An effort has been made to find a relationship between mean water levels and salinity through data analysis. Furthermore, an attempt has been made to find a relationship storm surge and salinity. For this analysis salinity and water level data in the period 1975 (when salinity measurements started) until 2016 has been analyzed.

Unfortunately the observation location for water levels and salinity are not exactly the same. In order to find relevant correlations the correlation between water level and salinity at nearby observation locations has to be calculated. Therefore the distance between the location for water levels observations and a location for salinity observations should be minimized, while both locations should be close to the river mouth. Therefore the water level measurement location "Hoek van Holland" and the salinity observation location "Ter Heide 1 km uit de kust" have been selected. The distance between these locations is less than 10 km, while both are situated close to the river mouth.

The influence of salinity on water levels could be visible if the mean water level is plotted against the salinity observation. Water levels have a sampling frequency of one hour, therefore the waterlevels are averaged over 2 tidal (M2) periods (which is equivalent to roughly 25 hours).

Figure 3.5 shows the samples, the corresponding correlation and a linear trend line. It should be noted that the sampling intensity of salinity observations is much lower (and more irregular) compared to the sampling intensity of water level observations (as shown in appendix A, table A.1, A.2 and A.3). Therefore, the salinity observation intensity determines how many observations can be used for calculating the correlation. The salinity observations with their corresponding water level (which is always measured within 30 minutes of the moment of sampling for the salinity observation) have been used in order to calculate the correlation. Figure B.1 shows the data has transformed to standard normal, which shows the semi-correlations as well.

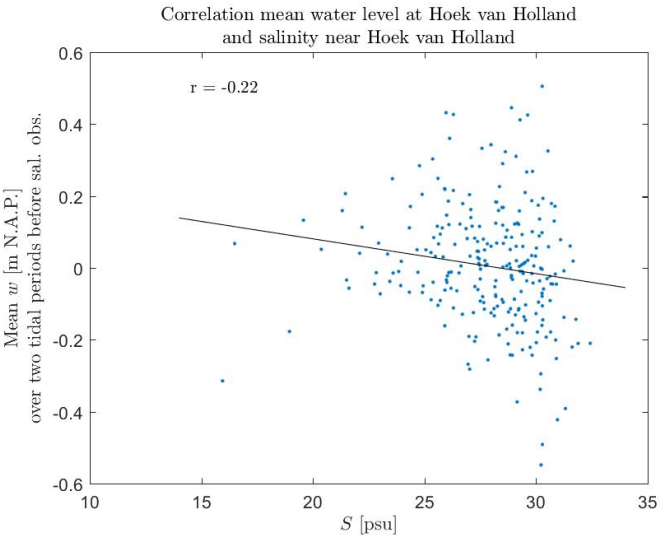


Figure 3.5: Mean water level (averaged over two tidal periods) and salinity observations are shown in blue dots, black lines indicates a linear trend line obtained by linear regression.

Conclusions

The results of this analysis show no large (negative) correlation between water levels and salinity. Several factors could be governing this correlation, i.e. there is no correlation present between these variables or the sampling method is not suitable (due to errors or the sampling method) in order to find a correlation.

It is probable that the last a important factor influencing this analysis. The sampling method could introduce a non representative salinity time series. This is supported by the fact that the water level time serie shows a surge larger than 1 meter. However, surges larger than 1 meter have been regularly present in the datasets for surges.

Salinity observations are done by a boat of Rijkswaterstaat and taking a sample requires a person going close to the water to physically take a sample. This is most certainly not possible in stormy conditions (i.e. $U_{10} > 20m/s$ wind and $H_s > 2m$). Therefore it can be concluded that the sampling method introduces some restrictions for this analysis, furthermore making it in fact not suitable for the conducted analysis.

In order to make this analysis work properly time continuous measurement of salinity observations are preferred, and even necessary. In the current dataset the salinity observation sampling frequency is in most cases less than twice a month and establishing a accurate mathematical relationship in such a dynamic environment is nearly impossible.

3.2.2. Correlation water levels and discharges

By analyzing the observations for water levels and discharges the correlation between these variables can be obtained or ultimately, by filtering the tide from the water levels, the correlation between surges and discharges can be obtained. The surge is defined in appendix B and obtained by filtering out the surge from the water level signal using the Matlab toolbox T_tide (Pawlowicz et al., 2002).

Figure 3.6 shows the correlation structure for the correlation between mean daily water levels at Hoek van Holland - mean daily discharge at Lobith (left graph) and mean daily surge at Hoek van Holland - mean daily discharge at Lobith (right graph). It should be noted that the correlation coefficient for the water levels is slightly higher than the correlation for the surge and that the semi correlation r_{NE} (which is of great importance for compound flood events) remains the same for water levels and surge.

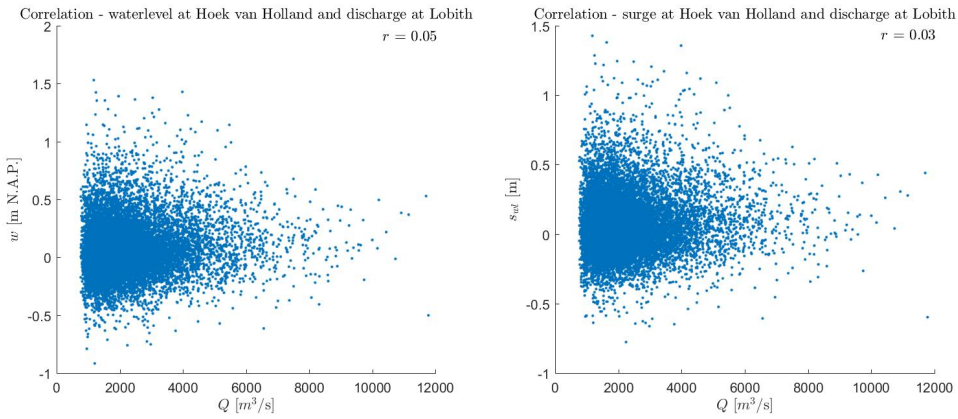


Figure 3.6: Overview of the correlation structure for the discharge at Lobith and the water level (left) and surge (right) at Hoek van Holland. The correlation coefficient for the complete dataset is given in the upper right corner of each scatter plot.

In order to get a better understanding of the behavior of compound events between high river discharges and high water levels or surges in the Netherlands the lagged correlation should be studied as well. After all a compound flood event can occur even if a high river discharge and a high water level or surge occurs very closely after each other. A lagged correlation is obtained by calculating the dependence between high water levels or surges and high river discharges if the discharge time series is lagged by a given number of days with respect to the water level time series. The 5% and 95% confidence bounds are computed by the Fisher transformation (eq. 2.15). The figures for 6 days lag and - the transformations to standard normal (obtained by fitting a distribution to each dataset) are shown in appendix D; figure D.1, D.2 and D.3.

Figure 3.7 shows the results of the analysis and the results show that the correlation, for both water levels and surge, is largest for a lag of 6 day, which Klerk et al. (2015) found as well for the lagged correlation between the water level series at Hoek van Holland and the discharge time series in Lobith for a slightly shorter time series. However, the semi correlation, r_{NE}, based on surge is a bit lower than the semi correlation for water levels for a time lag of 6 days.

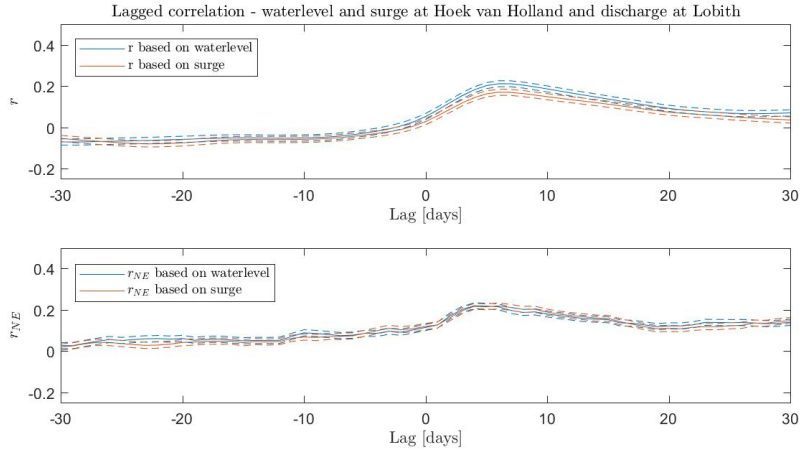


Figure 3.7: Cross correlation (upper figure) and the semi correlation r_{NE} for the discharge at Lobith and the water level and surge at Hoek van Holland with up to 30 days lag. The dashed lines indicate the 5% and 95% confidence intervals, computed by the Fisher transformation (eq. 2.15).

By studying the correlation between water levels or surges and salinity observations in the quantiles of these variables a better understanding of the behavior of the (higher) quantiles can be studied. This analysis is done by taking the residual dataset after taking the given quantile function, which is given in equation 3.3, for both variables (water level and discharge or surge and discharge) separately and calculating the correlation based on this residual dataset. Again, the confidence bounds are computed by the Fisher transformation (eq. 2.15).

$$Q_X(p) = x : \{P(X \leq x) = p\} \tag{3.3}$$

$$\begin{aligned} &r\Big(x \in A, y \in A\Big) \\ &A : (x > Q_x(p) \bigcap y > Q_y(p)) \end{aligned} \tag{3.4}$$

The top graph of figure 3.8 shows the correlation based on the quantile of both variables (water level and discharge or surge and discharge) or as mathematically described in equation 3.4. The figure shows a coverging correlation until a quantile of roughly 0.3, a stationary correlation between a quantile of 0.3 and 0.6. Furthermore, after this quantile the confidence bounds start to diverge from the calculated correlation (r).

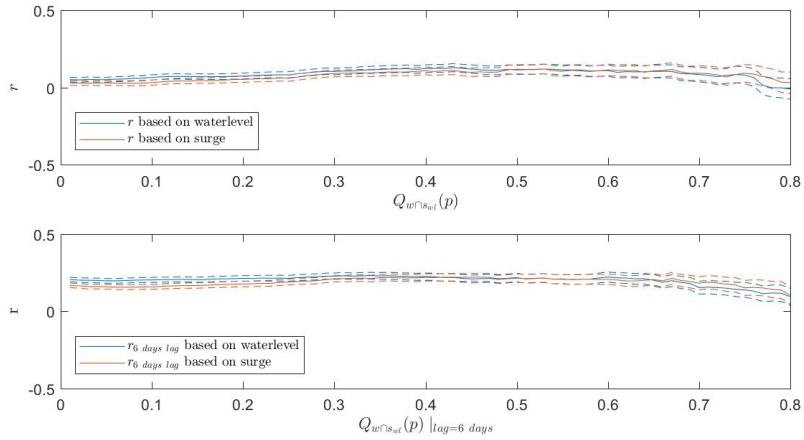


Figure 3.8: Correlation for the discharge at Lobith and the water level (top graph) and surge (lower graph) at Hoek van Holland depending on the (upper) quantile for discharge and water level or surge respectively. The dashed lines indicate the 5% and 95% confidence intervals, computed by the Fisher transformation (eq. 2.15).

3.2.3. Correlation salinity and discharge

The influence of discharge on salinity can be quantified by studying the correlation between these variables. Differences in salinity are primarily governed by the river discharge. Furthermore, due to mixing (by for example tides and wind stress) high salinity values could still be present simultaneous with high river discharge. Figure 3.9 shows the salinity - and discharge observations and the corresponding correlation. Low salinity values tend to correspond with high discharge values, therefore a quite strong negative correlation is present in the data. The transformed standard normal plot is shown in appendix E; figure E.1.

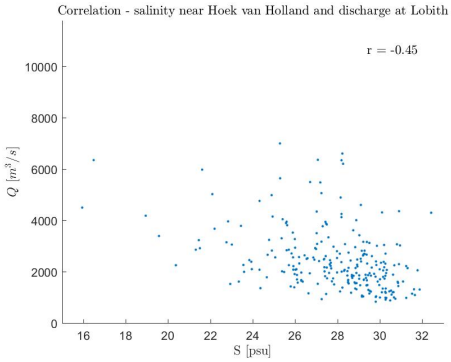


Figure 3.9: Correlation structure between discharge at Lobith and Salinity at Hoek van Holland.

By studying the influence of time (lag) on the correlation between river discharge near Lobith and salinity near Hoek van Holland the influence of this variable can be estimated. A lagged correlation is obtained by calculating the dependence between discharge and salinity if the salinity time series is lagged by a given number of days with respect to the discharge time series at Lobith. Figure 3.10 shows the results of the analysis and the correlation is minimal for a lag of two days. These two days corresponds quite well to the average travel time of a flood wave (distance/average waterspeed) of water from Lobith to Hoek van Holland.

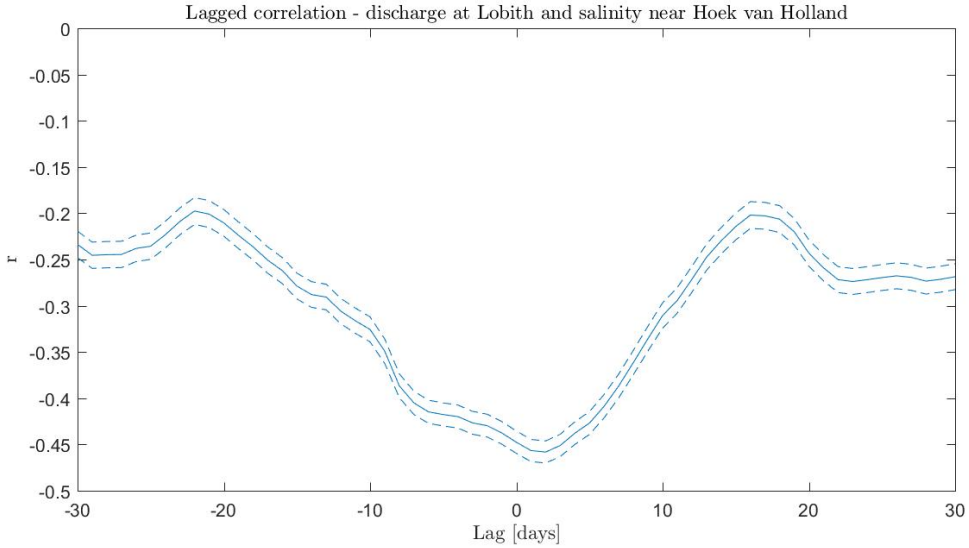


Figure 3.10: Lagged correlation discharge near Lobith and salinity near Hoek van Holland. The dashed lines indicate the 5% and 95% confidence intervals, computed by the Fisher transformation (eq. 2.15).

3.3. Return periods

Through univariate extreme value theory I would like to prove that return periods based for Hoek van Holland are different compared to location close by which are not directly influenced by the discharge of the river Rhine. This could prove the necessity of a multivariate approach, accentuate the possible influence of salinity differences on water levels and emphasize the relevance of this study to flood risks.

In the Netherlands flood defences are periodically checked whether the flood defence still meets the required safety standards. Therefore the dutch flood safety institute Rijkswaterstaat develops, together with some other partners, hydraulic boundary conditions (HBC) for flood defences (van Verkeer en Waterstaat, 2006). The HBC specifies at each location an exceedance probability of the flood defence and prescribes the normative combination of water levels, wave heights and -periods. These conditions are the conditions that flood defense at a specified location should be able to withstand. Figure 3.11 shows the normative water levels for the coast of Holland which are specified by the HBC for these respective locations. The exceedance probabilities in the HBC are calculated based on a risk based approach.

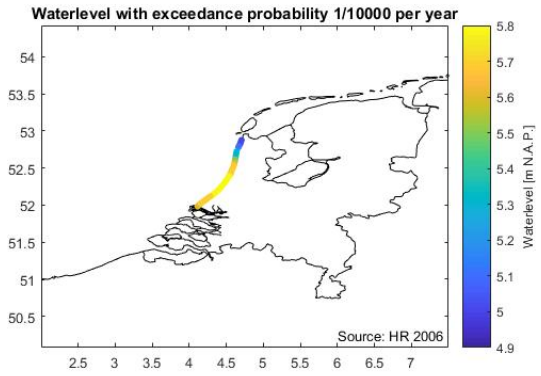


Figure 3.11: Matched salinity and water level observation locations

Since water level data is available this analysis can be made as well. The advantage of this is that the HBC only gives a normative combination of water levels, wave height and -periods for each location at one (high) return period, which makes analyzing the influence of the ROFI to this information not possible.

For this analysis a relatively simple approach has been carried out. An annual maxima approach has been applied, by fitting the annual maxima Z to a Gumbel distribution. According to Xu (2007) a Gumbel distribution is able to "reasonably predict the short term return period such as less than 30 year return period". In this analysis the return period is calculated by equation 3.5 and 3.6 and the pdf of the Gumbel cumulative distribution function is given by equation 3.7.

$$T_r = \frac{1}{1 - P(Z <\leq)} \tag{3.5}$$

$$F(z) = P(Z \leq z) \tag{3.6}$$

$$F(z|\mu,\sigma) = e^{e^{-\frac{z-\mu_G}{\sigma_G}}} \tag{3.7}$$

The annual maxima are fitted to a Gumbel distribution by method of moments. An analysis with a maximum likely hood estimator has been used as well, however these estimates underestimated the higher return periods. Equation 3.8 shows the used equations to fit the annual maxima (Aydin and Şenoğlu, 2015; Xu, 2007). The results of this analysis are shown in figure 3.12.

$$\begin{aligned} \sigma_G &= \frac{\sqrt{6}}{\pi}\sigma(z) \\ \mu_G &= \mu(z) - 0.5772\sigma_G \end{aligned} \tag{3.8}$$

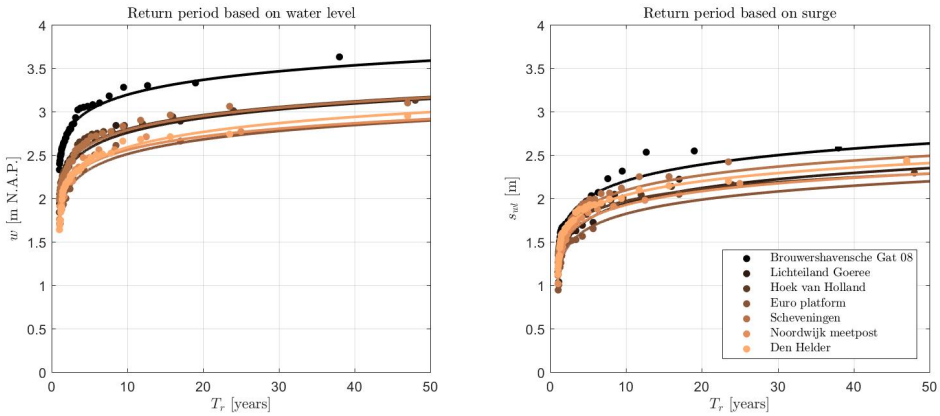


Figure 3.12: Return period estimated by a Gumbel distribution with method of moments fitting method. Dots represent the water levels and surges. The lines represents the respective (by color) fitted distribution.

Conclusion

The fact that the tidal amplitude is larger in the south of the Netherlands is visible in the plot for return period based on water levels. The influence of the larger tidal amplitude can be seen at the location "Brouwerhavensegat 08", wich is located most southernly. This location has by far the highest water levels per return period. On the other side, water level locations which are located at sea or most northern ("Den Helder") show a significant lower water level per return period. All these locations have a lower tidal amplitude, which could be one of the reasons for observing lower water levels per return period.

Still, after filtering the tidal amplitudes from the water levels at each locations, higher surge levels per return period can be observed at southern locations. However, the difference between all locations in water/surge levels per return period is smaller. High water levels, and surges, are generally forced by north western winds due to the funnel shape of the North Sea with this wind direction.

For the surge levels per return period at the location "Hoek van Holland", near the river mouth, a slight less steep gradient is observable. This could indicate a lower risk of flooding compared to other location for even higher return periods. However it should be noted that a univariate analysis has been conducted, in which compound flood events are not incorporated. Therefore this figure could give a false interpretation of the conditions that could initiate a flood at this location. On the other hand, the HBC (figure 3.11) shows this behavior as well, therefore the coastal shape of Holland could be a major influence on the spatial distribution of flood risks.

4

Numerical modeling of fresh water plumes

Numerical simulations are made in order to quantify the effects of density differences caused by salinity differences. This chapter covers all aspects from the numerical model. First a introduction on the numerical model is given, then the numerical model is explained in detail. After that, the numerical experiments are shown and the results of these experiments are presented. Furthermore, this chapter is finished by a discussion on the results.

4.1. Introduction to the numerical model

The goal of this numerical model is to capture the most important dynamic interactions and to be applicable all over the world. In literature, no usable and precise parameterization of the plume width has been described. Furthermore the effect of salinity difference which are induced by river outflow have not been studied before. Therefore, the model will be used to study the plume width and the hydraulic effect of salinity difference. Furthermore, the effect of salinity on local water levels needs potentially to be corrected in very large (worldwide) models, such as the GTSM which is part of GLOSSIS (Verlaan et al., 2015). The computational time of these 2D models is large, therefore such a localized- and 3D-effect can not be incorporated in the model, due to computational time restrictions.

The numerical Delft3D-FM model for this application is an idealized and simplified numerical model of the North Sea. The Delft3D-FM model has been inspired by the work of prof. dr. J.D. Pietrzak and dr. ir. G.J. de Boer. In multiple publications a Delft3D model of the Rhine ROFI is developed which successfully studied the interaction between tides and stratification (De Boer, 2009; de Boer et al., 2009) . After the development of unstructured grids in the Delft3D software prof. dr. ir. M. Verlaan made a replica model in Delft3D-FM, intended to study the spatial scale of salt intrusion on the river Rhine. Therefore, the ROFI is not modelled with great detail and some additional changes (which are presented in the next section) have been made to this Delft3D-FM model in order to model the Rhine ROFI correctly.

The simplified and idealized Delft3D-FM model schematizes the dutch coastline as a perfect northward pointing coastline and the depth of the North Sea uniformly. Figure 4.1 gives an impression of the numerical model. A tidal wave, with along- and cross shore descending tidal amplitude, is generated at the southern boundary. The western- and northern boundary are located far away from the river mouth, furthermore those boundaries are weakly to non-reflective. The steps in the river are modelled, since these are estimated to be essential for the penetration of salinity on the river. The next section provides more in-depth information on all of these aspects.

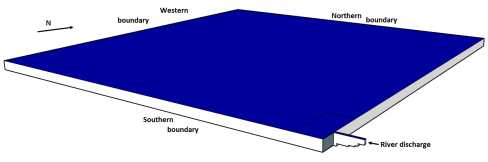


Figure 4.1: Simplified overview of the numerical model.

4.2. Numerical model

As stated above the numerical model from prof. dr. ir. M. Verlaan, which was developed for a different purpose (researching salt intrusion on the river), has formed the basis for the developed numerical model. Changes to this model were inevitable and have been made to improve the performance of the model and to make the model suitable for modeling the Rhine ROFI. First a brief introduction on the intrusion model is given, then a more in depth explanation of the developed ROFI model is presented.

4.2.1. Intrusion model

The goal of the intrusion model, developed by prof. dr. ir. M. Verlaan, is to study salt intrusion at the river Rhine and study the effect of the steps in the river bed on salinity intrusion.

The coastal grid in this model is much smaller compared to De Boer's (2009) numerical model (100x50 km compared to 750x750 km), which is completely justifiable for the application of the intrusion model. Furthermore, the Kelvin wave is model as an frictionless Kelvin wave, with an (shore) tidal amplitude of 1 m at all boundaries (south, west and north). The last important difference is the bathymetry, the coastal water is 20 m (compared to 25 m) and steps at the river Rhine (compared to a sloping river bed). Figure 4.2 gives an impression of the intrusion model.

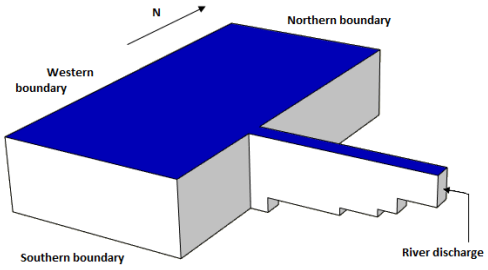


Figure 4.2: Simplified overview of the intrusion model.

4.2.2. ROFI model

The ROFI model is developed on the basis of the model made by prof. dr. ir. M. Verlaan and the findings of De Boer and prof. dr. J.D. Pietrzak. For the development of the model changes have been made to the intrusion model for computational settings, -grid and boundary conditions. These changes to the model are presented below. Furthermore, a detailed description on the bathymetry is shown in the computational grid explanation.

Computational settings

The intrusion model has been developed in April 2014, a period at which Delft3D-FM was still at early development. Due to these developments new functionalities, solving schemes and settings have been developped. Therefore, in consultation with ir. H.W.J. Kernkamp, dr. ir. M. Nabi (both developpers of Delft3D-FM) and ir. F. Zijl (senior researcher with a large experience in setting up and calibrating various 3D models of weakly dynamic, stratified seas) computation changes were made to the intrusion model. The applied computational settings are shown below.

- Timesteptype : 3 (automatic timestep for 3D)
- 2D Conveyance: -1 (3d setting)
- Advection solver: 33
- Tlfsmo: 1 day (spin-up time for tide)
- Layer type: sigma
- Number of (sigma) layers (see below)

Sigma layers

In order to reduce the numerical error in Delft3D-FM the number of sigma layers has been made dependent on the uniform coastal depth (D). Figure 4.3 and equation 4.1 shows the total number of sigma layers (N_s) on the whole grid depending on the depth. The layer thickness of a sigma layer is on average between 0.25 m and 2.0 m, ensuring sufficient accuracy and computational efficiency. For lower depths, the program is able to handle stratification in the model, while under higher depths the model is still sufficiently fast and not unnecessary inaccurate.

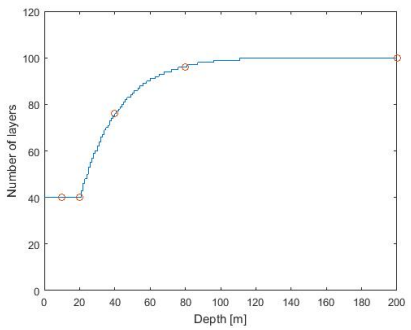


Figure 4.3: Number of sigma layers depending on the depth of the model

$$N_s(D) = (D \leq 20) * 40 + ceil((D > 20) * 100(1 - 0.6e^{0.045(20-D)})) \tag{4.1}$$

In the model the sigma growth factor is equal to 1, meaning equally spaced sigma layers over depth. This implies that for instance the depth averaged density is equal to the average density.

Computational grid

Furthermore changes to grid have been made. The computational grid has been severely enlarged, to decrease the influence of the boundaries on the computational results. Figure 4.4 shows the used computational grid, with the respective cell size at each location. Refinements in the grid have been made using the "Refine quads factor 2 (triangular border type)", ensuring a perfect orthogonal grid.

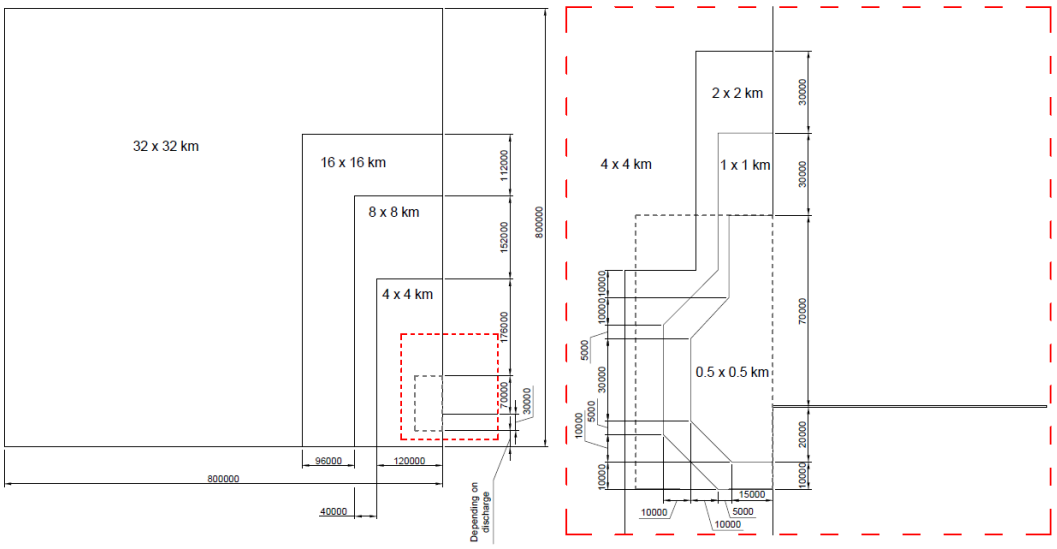


Figure 4.4: Computational grid used for parametrization. The left graph shows the complete grid and the right graph shows a close up with more detail of the red square of the left figure. Dimensions of the computational grid are shown in both figures and computational grid cell are shown in text at each location within a range of the computational grid.

The use of a new, much larger, computational grid does increase the computational time of the model. Due to partitioning the grid in four domains and calculating the model on the Deltares H6 server on one node the computational time is roughly 4 minutes and 28 seconds of computational modeling per modeled day (for a depth of 20 m), which is suitable for this application.

The region of interest for all analysis is enclosed by the red dashed rectangular shape which is 30 km south of the Rhine river mouth, 70 km north of the Rhine river mouth and 50 km offshore of the river mouth.

Bathymetry

The bathymetry of the numerical model has been simplified and only the rough characteristics of the North Sea bathymetry have been implemented. The bottom of the North Sea is modeled at a uniform depth of 20 meters. The bathymetry of the river Rhine is modeled with the steps which are present in the river. These steps in the river bed have a large influence on the intrusion of salinity on the river. Furthermore the steps are essential in order to model the in- and outflow characteristics of salinity at the North Sea.

Table 4.1: The characteristics of the modelled river steps

Distance to mouth	Depth
5 km	-16 m N.A.P.
25 km	-14 m N.A.P.
35 km	-12 m N.A.P.
40 km	-8 m N.A.P.

Boundary conditions
The tides at the North Sea between can be described by a Kelvin wave. The (frictionless) Kelvin wave, which is based on the shallow water equations, was first described by Taylor (1922). De Boer (2009) used, based on an approach of Jacobs (2004), a method where linearized, along- and cross shore, friction is incorporated in the analytical solution of the Kelvin wave. This Kelvin wave (which is mathematically described by $w_t(t,x,y)$) has an decreasing along- and cross-shore amplitude and a varying phase over the computed grid. Equation 4.2 shows the complex form of tidal amplitude applied at a boundary in the model. These equations need to be solved iteratively, since κ is depended on the depth-averaged velocity, which is dependent on $w_t(t,x,y)$.

In these equations is w_{t_0} the tidal wave amplitude, m is a measure for the cross-shore variation and k is a measure for the alongshore variation. The initial, frictionless, wave length is given by k_0, the Coriolis parameter is given by f and the initial velocity of the tidal wave is given by c_0. Furthermore, σ gives the relation between friction and inertia, by dividing the linearized alongshore friction parameter κ by the angular frequency of the tide ω.

$$w_t(t,x,y) = w_{t_0} e^{-Re(m)x - Re(k)y} cos(\omega t - Im(m)x - Im(k)y) \tag{4.2}$$

With:

$$k = i k_0 \sqrt{1 - i\sigma}$$
$$m = \frac{f}{c_0} \frac{1}{\sqrt{1 - i\sigma}}$$
$$\sigma = \frac{\kappa}{\omega}$$

Several combinations of boundary types for the southern-, western- and northern boundary have been tested. During these experiments combinations with water level, velocity, or Neumann (water level gradient) boundary types have been tested, unfortunately all with no satisfying results. The approach of dr. ir. G.J. de Boer will be applied for the parametrization, which consists of a water level boundary on the southern boundary and a Riemann (weakly reflective boundary type) at the western- and northern boundary. This required additional programming of boundary conditions by the Delft3D-FM software team, which is greatly appreciated. For easy tracking of time, S_2 tides have been used in all analysis.

Southern boundary
Based on testing the model a significant influence of the southern boundary was found. In order for the model to work properly preferably the southern boundary is as close to the river mouth as possible. However the southern boundary may not influence the solution, especially the spatial salinity distribution. The southern boundary uses a Harleman-Tatcher boundary for the salinity, which decreases the influence of this boundary. Depending on the discharge in the model the location of the southern boundary has been chosen:

- Discharge < 5000 m^3/s: Distance river mouth to southern boundary = 30 km (same as for Gerben's model)
- 5000 $m^3/s \leq$ Discharge < 10000 m^3/s: Distance river mouth to southern boundary = 40 km
- Discharge $\geq$ 10000 m^3/s: Distance river mouth to southern boundary = 50 km

4.3. Numerical experiments
Through numerical experiments the plume width and hydraulic effect of salinity on water levels are researched. A lacking understanding of these aspects is found in literature and these aspects are particulary relevant in this research. The plume width gives an indication of the spatial scale of the effect and this thesis is centralized on the hydraulic effect of salinity on water levels.

By varying individual- and combinations of parameters an understandig of the governing processes can be obtained. First a description of the varied parameters is given, then the method of these numerical experiments is explained.

4.3.1. Variation of parameter space

An overview of all numerical experiments is given in table 4.2. The first five and the eighth experiments only one parameter was varied systematically, ensuring a complete span over the six dimensional space. Furthermore, the sixth and seventh experiment are done by varying both the width or the depth together with the discharge to study the combined effect of these parameters. The ninth and tenth experiment are performed at last to ensure high τ's.

Table 4.2: Overview of all numerical experiments

Experiment	Varied parameters	Fixed parameters
1	$Q: 500,\ldots,10000\ m^3/s$	S, D, W, w_{t_0}, f
2	$S: 30...40\ psu$	Q, D, W, w_{t_0}, f
3	$W: 500...10000\ m$	Q, S, D, w_{t_0}, f
4	$D: 10...200\ m^1$	Q, S, W, w_{t_0}, f
5	$w_{t_0}: 0.5,\ldots,2.0\ m$	Q, S, D, W, f
6	$Q: 2000,\ldots,32000\ m^3/s,\ Q/W = 4$	S, D, w_{t_0}, f
7	$Q: 1000,\ldots,16000\ m^3/s,\ Q/D = 100$	S, W, w_{t_0}, f
8	$f: 10°,\ldots,80°$	Q, S, D, W, w_{t_0}
9	$Q: 20000,\ldots,50000\ m^3/s$	S, D, W, w_{t_0}, f
10	$Q: 20000,\ldots,50000\ m^3/s,$ $for\ case\ D: 5, 10, 20\ m\ and\ W: 16000\ m$	S, W, w_{t_0}, f

4.3.2. Method

In order to get results on the plume width and the influence of salinity water levels numerical models are used. During each model run the model is performed twice, onces with salinity turned on and once with salinity turned off, while having the same coastal water density. The model with salinity is used to estimate the plume width and the hydraulic effect is studied by finding the difference in water levels between both models.

Results can only be obtained if the model has reached equilibrium. Therefore, under stationary conditions (i.e. one tidal amplitude, -period and constant discharge) the time dependent spatial change of salinity in the model at the moment before collecting results is essential. Results are unusable if the fresh water plume in the model is not fully developed at the moment of collecting the results. Therefore, it is essential to model a sufficient period of time for the model to reach equilibrium. For this reason, the model is performed for a computational modeling period of 270 days to ensure an equilibrium solution. Numerical results presented in this thesis are time averaged over the last five days of numerical simulation for all applications, of course except for time dependent process analysis.

By varying several (combinations of) parameters the dimensionless parameters which characterize the fresh water plume will vary as well. Some of these are dependent on difference in salinity at the river mouth and off shore salinity ($\tau\ and\ K_i$, see section 2.2). Therefore, those dimensionless parameters can only be resolved after performing the numerical model and reading the output files of the model.

The results for the hydraulic effect are presented as a observed water level difference (∂h). The observed water level difference is the difference in water levels between a model without- and a model with salinity, as shown in equation 4.3. Using this convention negative observed water level difference (∂h) are overestimations of the model without salinity. In other words, GLOSSIS would give in this case an over prediction in water levels, which is defined in this thesis as negative.

$$\partial h = w_{no\ sal}(t) - w_{sal}(t) \tag{4.3}$$

4.3.3. Results

The results consist the dimensionless numbers, the influence of each parameter to reduced gravity at the river mouth, plume width and the effect of salinity on water levels. An complete overview of the results of all experiments is shown in appendix F; table F.2.

Dimensionless parameters

Figure 4.5 shows the governing parameters on the right side, and the dimensionless parameters on the left (see eq 2.3) if mixing of the river is taken into account. A total of 55 experiments have been performed in order to cover a wide range of all dimensionless parameters.

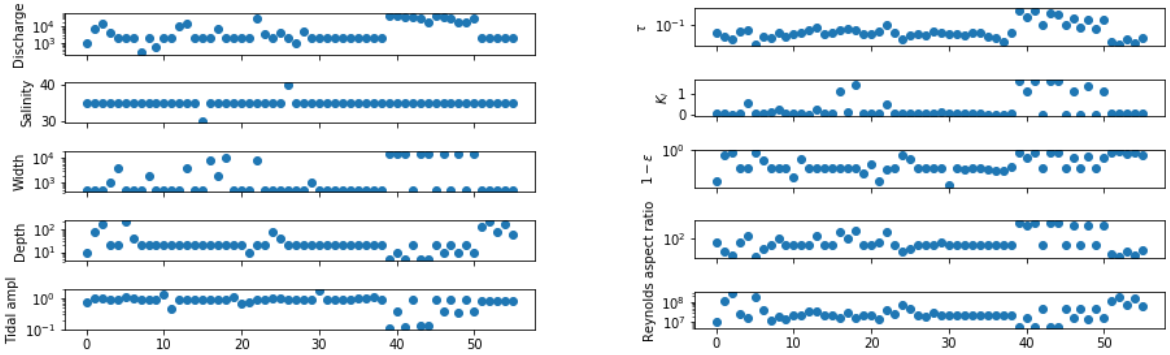


Figure 4.5: Left graph shows the dimensional bulk parameters of the numerical experiments and the right graph shows the corresponding dimensionless parameters. These are obtained after the numerical experiments.

Reduced gravity

The reduced gravity is a relative measure for difference between the depth averaged salinity at the river mouth and the coastal salinity. The reduced gravity will find a maximum, since it is governed by the difference in salinity at the river mouth and the coastal salinity. Of course, understanding the influence of several parameters on the reduced gravity is an essential first step in the complete process.

The results are shown in figure 4.6. As can be seen in the figure, discharge (which is bound by a maximum) and depth have the largest, positive and negative respectively, influence on the reduced density. Slight positive influence is found for salinity and tidal amplitude and slight negative influence is found for width and the coriolis parameter.

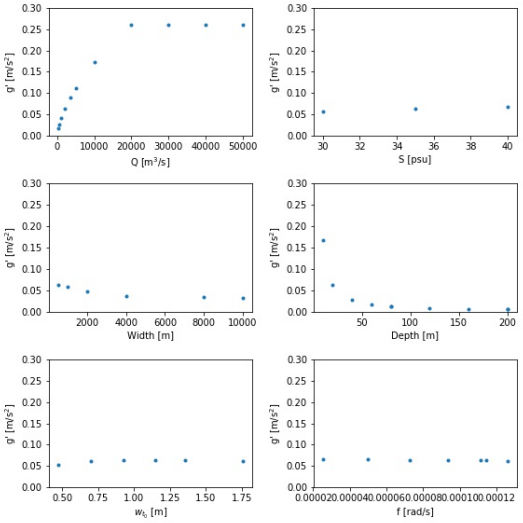


Figure 4.6: The influence of discharge, salinity, width, depth, tidal amplitude and coriolis parameter on the reduced gravity

Plume width

The cross-shore plume width is calculated based on the numerical results, using the approach of Garvine (described in section 2.2). First a brief description of cross-shore plume development will be given, then the plume width will be analyzed to the influence of several parameters.

Plume growth over time

The time dependent plume width is presented in figure 4.7. The fresh water plume forms in some models within 5 days and stay stationary in size after wards, while other numerical model (with a different discharge, salinity, width, depth or tidal amplitude) take much longer to become stationary. Furthermore, some models show nearly none scatter, while others show quite some significant scatter. The scatter is a sign of a chaotic

process, which is probably induced by tide - salinity interaction, since tides are modeled as perfect linear waves. The time dependent scatter in the plume width implies the need for time averaging plume width (as mentioned in 4.3.2.), since these time dependent processes are not part of this research. Nearly all plumes are completely formed and not growing anymore after 270 days of numerical simulation. Therefore, the conclusion can be drawn that 270 days of numerical modeling is sufficient for this application (plume width estimation).

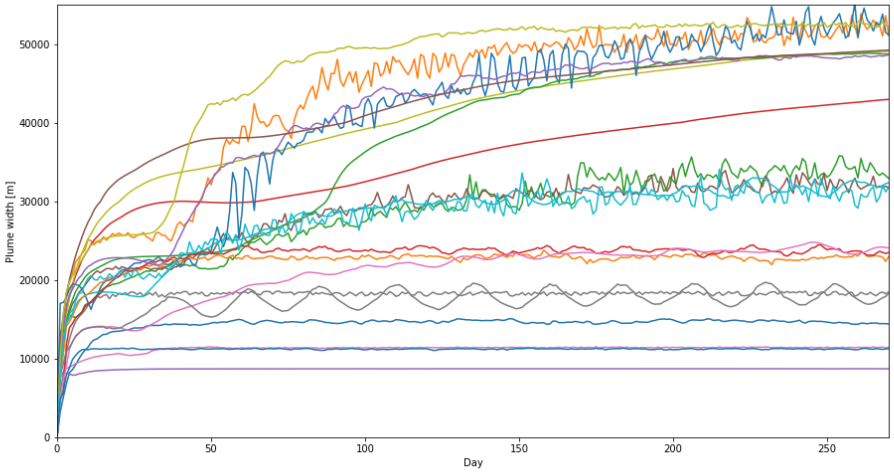


Figure 4.7: Time dependent plume response of selected numerical models, each color indicates a different model. The selection is based on plumes with a final plume width smaller than 50 km.

Plume width response to parameters

The cross shore width of the fresh water plume is of great importance for the spatial scale of the plume. The results from the numerical models are shown in figure 4.8. As can be seen in the figure, discharge and depth have the largest influence on the plume width. Slight positive influence is found for salinity, width and coriolis parameter and slight negative influence is found for tidal amplitude. Furthermore, for the discharge the plume width seems to be bound by a maximum or showing asymptotic behavior. Near linear trends are found for salinity and the coriolis parameter. Width shows a remarkable response, between 4000 m and 8000 m the plume with increases suddenly while being (near) stationairy for lower and larger widths.

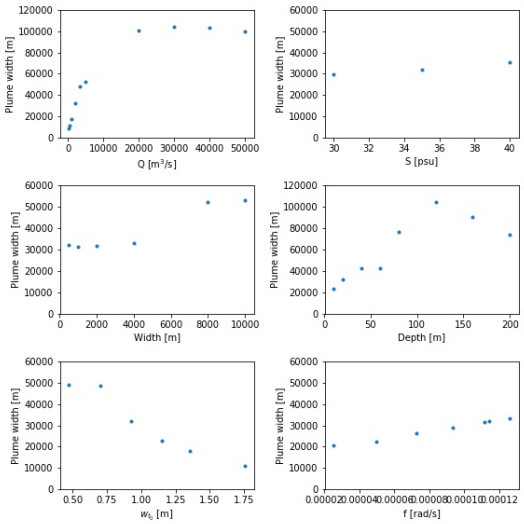


Figure 4.8: The influence of discharge, salinity, width, depth, tidal amplitude and coriolis parameter on the plume width

The fresh water cross-shore plume width response to depth is remarkable. A maximum fresh water plume width is visible in the graph, possibly indicating that different mixing processes become dominant with larger depths. Furthermore, possibly the characteristic mixing process in the Rhine ROFI, induced by tidal ellipses observed by de Boer et al. (2009), disappears with larger depths. Unfortunately, the output results of the numerical models are not usable to analyse these current structures and mixing processes. However, this observation alone is of great scientific value and can most definitely be researched in the near future.

Hydraulic effect of salinity on water levels

Fresh water river discharge has a local effect on water levels. This local effect on water levels is most likely governed by the presence of the fresher, lighter, water. In this section the static effect of salinity on water levels will be quantified.

When comparing two models with the same characteristics (discharge, river width, tide, depth and lattitude) where one is calculated with salinity and the other is not, a difference in water level between these models can be observed. This difference in water level is time dependent and the results for 25 models from figure 4.5 are shown in figure 4.9. For future reference, only time averaged water levels differences will be analyzed.

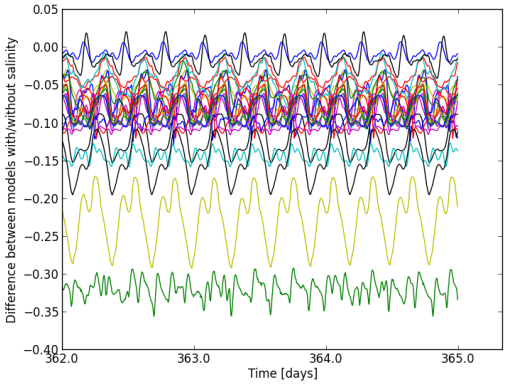


Figure 4.9: Time dependent difference between models with - and without salinity.

The hydraulic effect of salinity on water levels has been researched by two methods: steric height anomaly and a hydrostatic balance equation. Both methods are explained in this order.

Steric height anomaly

The relation between steric height anomaly (Δh) and the observed water level differences (∂h) will be expressed in a graph. The steric height anomaly is a measure for density induced water level differences, which are in this case induced by salinity. The formula for steric height anomaly (Δh) is given in eq. 4.4 and it is based on equal pressures at the sea bed.

$$\begin{aligned} &\Delta h = \int_{-D}^{h} \frac{\rho_0 - \rho(z)}{\rho_0} dz = h + D - \frac{1}{\rho_0} \int_{-D}^{h} \rho(z) dz \\ &and \int_{-D}^{h} \rho(z) dz = \overline{\rho}(h+D) \\ &with\ \overline{\rho} = \frac{\sum_{i=1}^{N} \Delta h_i \rho_i}{\sum_{i=1}^{N} \Delta h_i} \quad (as\ the\ depth\ averaged\ density) \end{aligned} \tag{4.4}$$

In figure 4.10 the results for the analysis for the observed water level differences which are induced by salinity are shown, with each color indicating a different model. The steric height anomaly seems to be a good predictor for the observed water level difference, although scatter and a square root relation is observed.

However, as shown in the left graph of figure 4.11 a purely steric fit ($\partial h = \sqrt{\Delta h}$, with $r^2 = 0.58$). This seems to be the lower bound of the hydraulic effect, which is observed in some cases. Through a linear regression on $y = ax$ the slope of $\partial h = a\sqrt{\Delta h}$ can be found. After such an analysis an $a = 0.684$ is found, with an $r^2 = 0.92$.

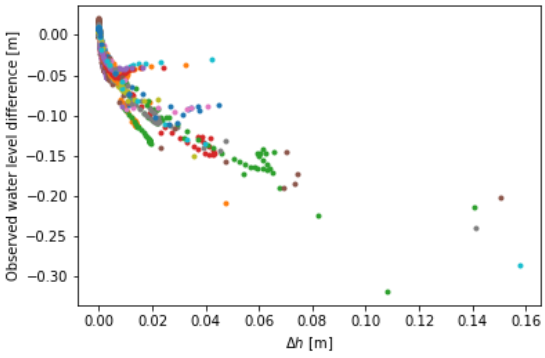


Figure 4.10: Steric height anomaly plotted against observed water level difference. Each color indicates results for a different model.

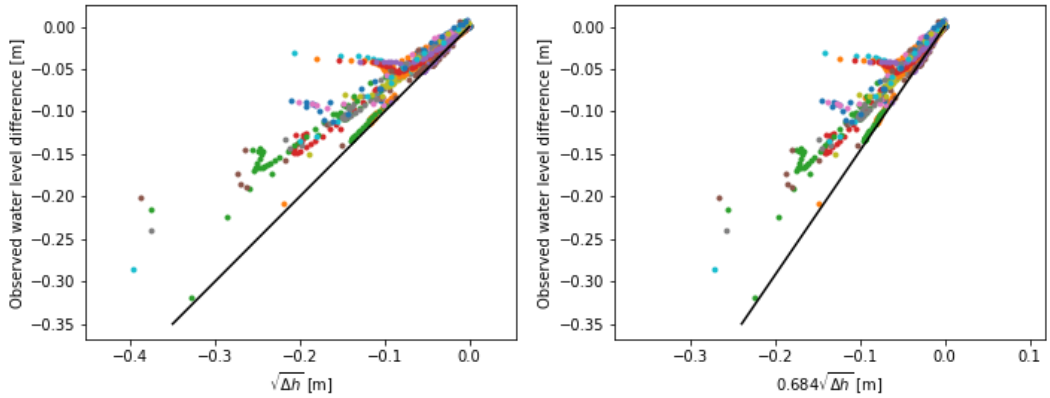


Figure 4.11: Left graph shows the square root of the steric height plotted to the observed water level difference and the right graph shows the fit for the same variables based on a linear regression. Each model is indicated by a different color in both graphs and the black line indicates a purely steric conditions in both graphs.

Although the fit for $\partial h = a\sqrt{\Delta h}$ is exceptionally good $r^2 = 0.92$, the physics and governing processes are neglected. Studying the slope of each individual model could potentially lead to better understanding of the physics and even better predictions for the effect on water levels. Furthermore, this approach seems unpractical due square root relation of steric height to observed water level differences. Therefore, an approach based on a hydrostatic balance has been developed.

Hydrostatic balance

When first analyzing this problem a partly steric effect was found. However, the relation based on the square root of the steric height to observed water level differences is not practical. Therefore, a new approach might be better. As stated before, steric height is based on equal pressure at the sea bed, the approach below described below is based on a fluid force balance.

For the sake of simplicity, the process is preferably hydrostatic. In order to test this hypothesis results from the numerical models for the parameterization (water levels and density information) are used. First a brief explanation on the hydrostatic pressure balance.

In this explanation (depth-)averaged densities and water levels are assumed for the sake of simplicity, an approximation which has no interference with the derivation. Figure 4.12 shows the starting point for the derivation. Water column 1 has a higher depth-averaged density compared to water column 2. In order to have a hydrostatic pressure balance, the water level in water column 2 will be higher. The derivation below shows the equation of ratio of densities to the ratio of water levels for which hydrostatic pressure balance holds.

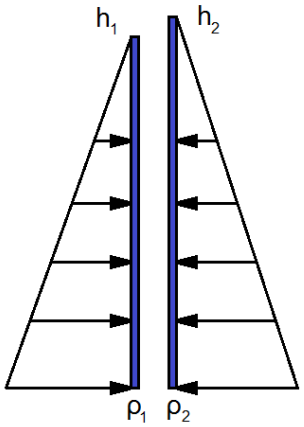


Figure 4.12: Hydrostatic pressure balance based on two water columns with both different densities. $\rho_1 > \rho_2$ and $h_2 > h_1$

$$F_1 = F_2 \qquad F_i = \int_0^{h_i} \overline{\rho}_i g z dz$$

$$\frac{1}{2}\overline{\rho}_1 g \overline{h}_1^2 = \frac{1}{2}\overline{\rho}_2 g \overline{h}_2^2$$

$$\frac{\overline{\rho}_1}{\overline{\rho}_2} = \frac{\overline{h}_2^2}{\overline{h}_1^2}$$

Based on this brief derivation, two scaling factors r_ρ and r_h can be distinguished as shown in eq. 4.5 (see next page). If $r_\rho = r_h$ then a hydrostatic pressure balance is at hand.

$$r_\rho = \frac{\overline{\rho}_1}{\overline{\rho}_2} \qquad r_h = \frac{\overline{h}_2^2}{\overline{h}_1^2} \tag{4.5}$$

The numerical model gives output at each observation location, which are, among other locations, regularly spaced cross-shore of the river mouth. At each observation location water levels and salinity (which can be converted to densities) is given as a time dependent output. By time averaging the water levels and densities over multiple tidal periods at each observation point r_h and r_ρ (of each observation point) can be computed. If this is done for all experiments, a linear trend is visible in each of the model results. Through a regression, based on ordinary least squares (OLS), the slope (α) of each linear trend is determined. This data set can later be used for a parameterization and possibly improving predictions methods compared to the steric height method.

$$\begin{aligned} \alpha = & regression\ based\ on\ \{r_\rho - 1, r_h - 1\} \\ & for\ \beta\ in \quad r_h - 1 = \beta(r_\rho - 1) \end{aligned} \tag{4.6}$$

Figure 4.14 shows a random selection of the regression fits of the hydrostatic analysis (please note the different x and y scales). The results show good fits and a linear fit is completely justifiable as shown in the figure. Again, the approach described above is applied in further analysis.

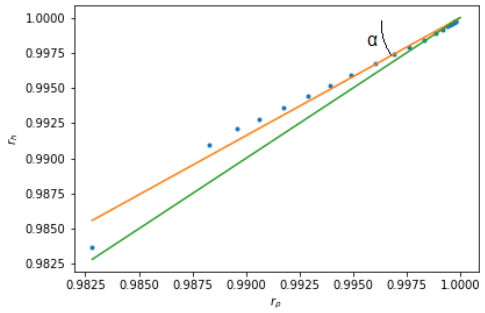


Figure 4.13: The blue dots represent the results from an individual observation location, orange line is the fitted line and the green line indicates purely hydrostatic conditions.

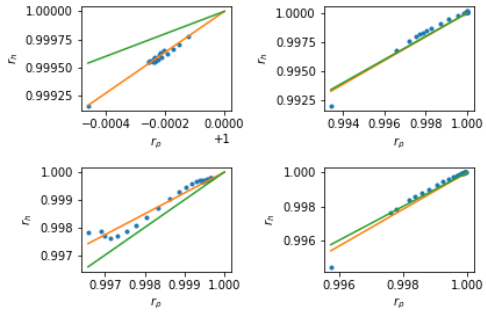


Figure 4.14: Random fits of the slope between r_ρ and r_h, the blue dots represent the observations from the hydraulic model, the orange line is the fitted slope and green line is the line if purely steric conditions are observed.

Repsonse of a to parameters

The slope a in the hydrostatic balance equation can be analyzed based on the results of the numerical experiment. These results of the analysis are summarized in figure 4.15. The depth and tidal influence show the highest response, in positive and negative linear trends respectively. Furthermore, the slope a in the hydrostatic balance equation seems to be insensitive to salinity and the Coriolis parameter and for discharge and width a mild negative trend line is visible.

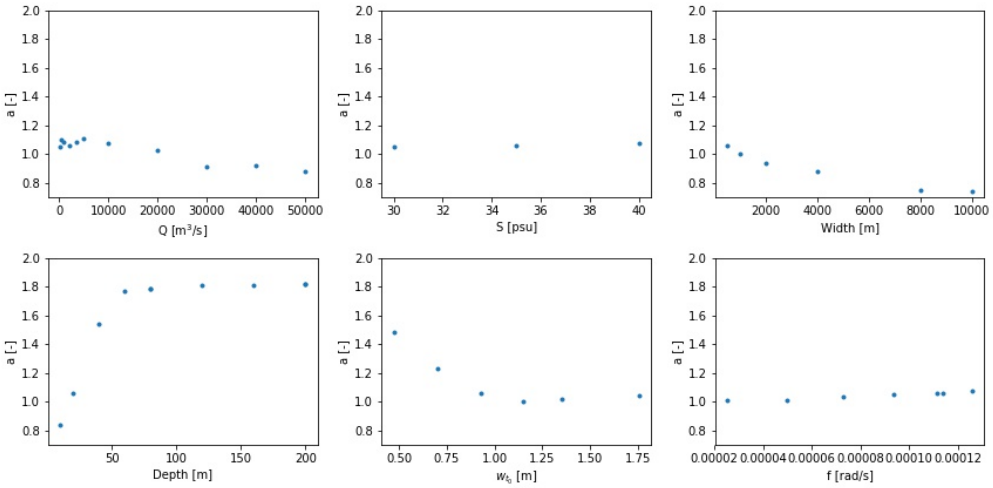


Figure 4.15: Overview of the analyzed numerical experiment results for the slope a in the hydrostatic balance equation.

4.4. Dynamic response

The size of a river plume responds for dynamically to changes in discharge and weather. So far, these changes have not been taken into account. This section will describe through a limited number of numerical simulations the implications of these dynamic changes in river discharge and weather.

Changes in river discharge

In order to assess the impact of changes in river discharge a restart file of the standard model[2] is created after simulating 270 days. This restart file is used to start the numerical model with a sudden increase in discharge per numerical experiment. The model simulates another 20 days and the discharge is varied for each experiment over a range of 4000 m^3/s to 20000 m^3/s. Based on the output files of the model, the plume width is calculated daily. The results are shown in figure 4.16. Unfortunately the model was set up to measure a maximum plume width of 50 km and the complete results for all models can not be shown in the graph.

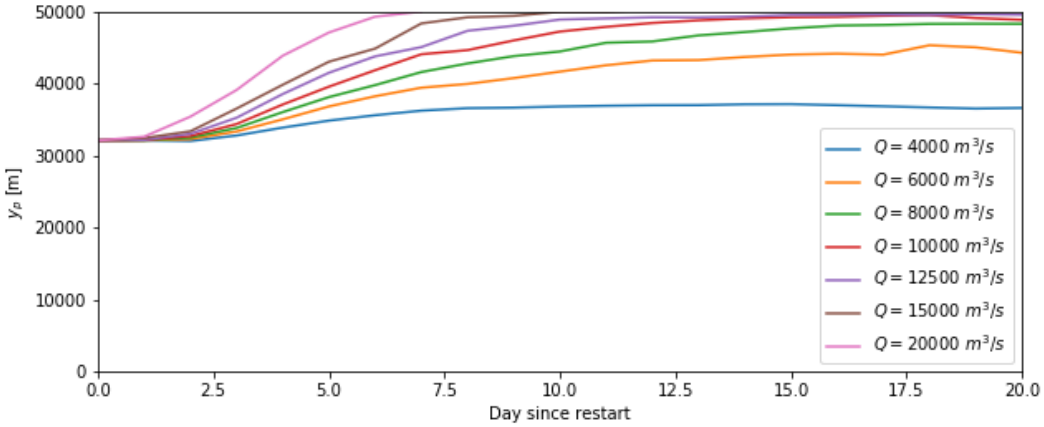


Figure 4.16: Time dependent behavior of plume width to changes in river discharge.

The results in figure 4.16 show that the plume width grows due to an increase in discharge. For discharges lower than 10000 m^3/s the plume grows within 10 to 25 days to the size of a fully developped plume, see figure 4.8. For higher discharge the plume grows at a similar rate, but not straight to the expected level. Furthermore, the figure shows that the plume width converges the the plume widths corresponding to the results presented in figure 4.8. Furthermore, figure 4.16 shows that the higher the discharge is increased the longer it takes to converge to the new plume size.

Changes in weather

A fresh water river plume will respond to weather. Mixing will be induced by wind speed through the stress at sea surface by wind. Furthermore, changes in wind direction and air pressure can induce mixing as well. In order to assess the influence two cases have been simulated.

The intention of this analysis is to model an extreme event and analysis the response of the river plume to the water level is measured. Both cases are quite similar, first the model is restarted after 270 days of simulation with no wind. Then the wind is gently increased, in a period of 2 days, to 5 m/s in NW direction (side cross-onshore). After 6 and 9 days (for case 1 and 2 respectively) a depression with maximum windspeeds of 35 m/s will cross, in NNW direction, the river plume.

The constant wind and air pressure are modeled uniformly over the complete grid. The airpressure of the depression is modeled as a uncorrelated bi-normal pdf subtracted from the uniform air pressure. The gradients in x and y direction ($\frac{\partial P}{dx}$ and $\frac{\partial P}{dy}$) of this uncorrelated bi-normal form the wind speed in x and y direction. By rotating the gradients ($u_x = \frac{\partial P}{dx}$ and $u_y = -\frac{\partial P}{dy}$) the wind speed is perpendicular and rotating anti-clockwise the air pressure gradients. Figure 4.17 shows the air pressure and the wind before the passing of the storm.

[2]Discharge: 2000 m^3/s, Salinity: 35 psu, Width river: 500 m, Depth: 20 m, Tidal amplitude: 1 m and latitude: 51.8°

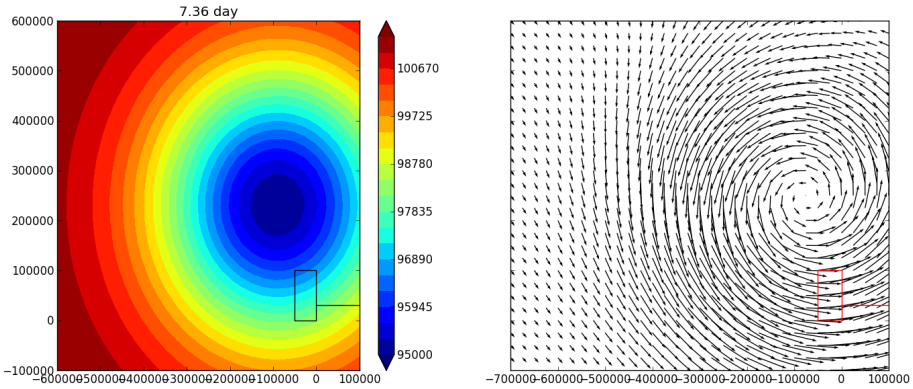


Figure 4.17: Left graph shows a contourplot of the airpressure, in black the river and region of interest are shown. Right graph shows a quiverplot to indicate the windspeed and -direction, in red the river and region of interest are shown. The quiverplot shows maximum windspeeds reaching 35 m/s and minimum 5 m/s. Both graphs are made at 7.36 days after restart.

The difference between both cases is the location of crossing the river plume. For case 1, the depression passes straight over the river mouth and for case 2, the depression passes North of the river mouth. For this case, the river plume is for a longer period of time to the high wind speeds. Figure 4.18 shows the windspeed and -direction for both cases.

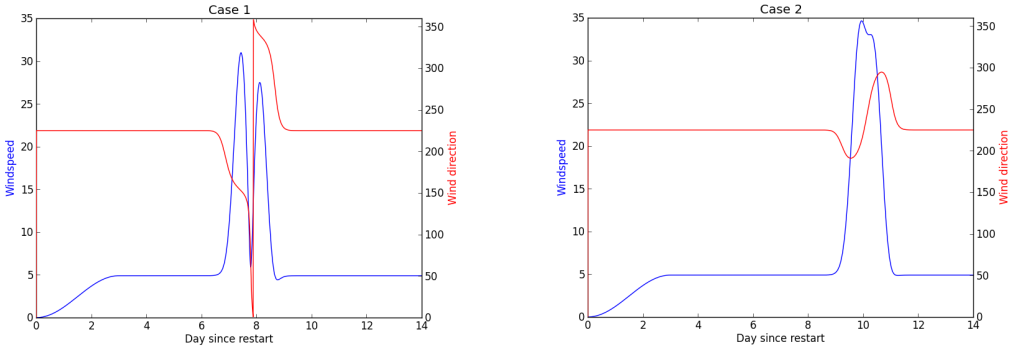


Figure 4.18: Windspeed at the river mouth for both cases. In case 1 the depression goes straight over the river mouth, as shown by the wind direction. In case 2, the storm arrives later and passes more North.

Plume width

First, the response of the plume width is analyzed to wind influence. The same method as for the static approach has been applied, only daily averages (instead of averages over 5 days) of sea surface salinity have been used to calculate the fresh water plume width. During this analysis it is of great importance to keep an eye out on the influence of boundaries on the numerical solution since the numerical model initially developed to model storms.

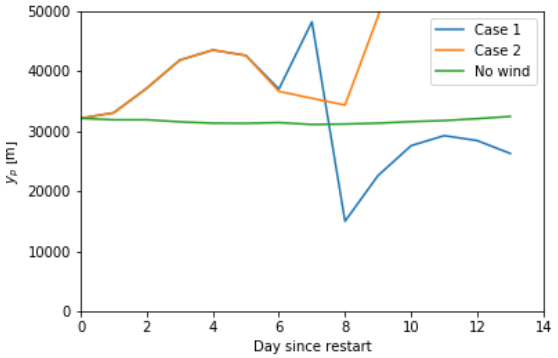


Figure 4.19: The plume width based on the two cases, case three is a case without wind.

The plume width is exactly the same for the first six days for both case 1 and 2, since the wind and air pressure are exactly the same for these cases (see figure 4.18). During this period the plume width increases approximately 15 km, which is expected by literature (Van Wiechen, 2011).

After 6 days the storm arrives, for case 1. The storm arrives from NNE direction, pushing the fresh water to the south. This results in a larger plume width. If the storm is right above the fresh water plume, the plume gets fully mixed. Due to this mixing the plume width becomes three times smaller. After the storm has passed the plume size starts to increase. However, results become very unreliable due to the fact that a lot of salinity has moved through the southern boundary. After 12 days, the plume width starts to decrease due to the fact that the southern boundary is located to close the river mouth to model storms with this intensity.

For case 2 the storm arrives from NNE direction after 9 days. The storm pushes the fully developed fresh water plume to the south. This results in a larger plume width, larger than the maximum observable plume width of 50 km. After the tenth day the plume width is too large and the southern boundary is located to close to the river mouth in order for the model to mix the fresh water with the saltier shelf waters.

Hydraulic effect

Second, the hydraulic effect of the fresh water plume is researched under the influence of a storm. The water levels for each case are given as an output for both a model with salinity and a model without salinity. By analyzing these water levels the hydraulic effect of the fresh water plume under the influence of an intense storm can be described. The results are shown in figure 4.20.

The maximum observed water level for case 1 is higher than for case 2, due to the fact the highest surge coincides with low tide in case 2. However, the surge is in case 2 slightly higher. As shown in the figure, the observed water level difference (∂h) is near constant and slightly increasing up to six days. For both cases applies that, if the storm arrives, the hydraulic effect gets smaller. High surges coincide with low observed water level differences, while under constant wind stress the effect tends to (slightly) increase.

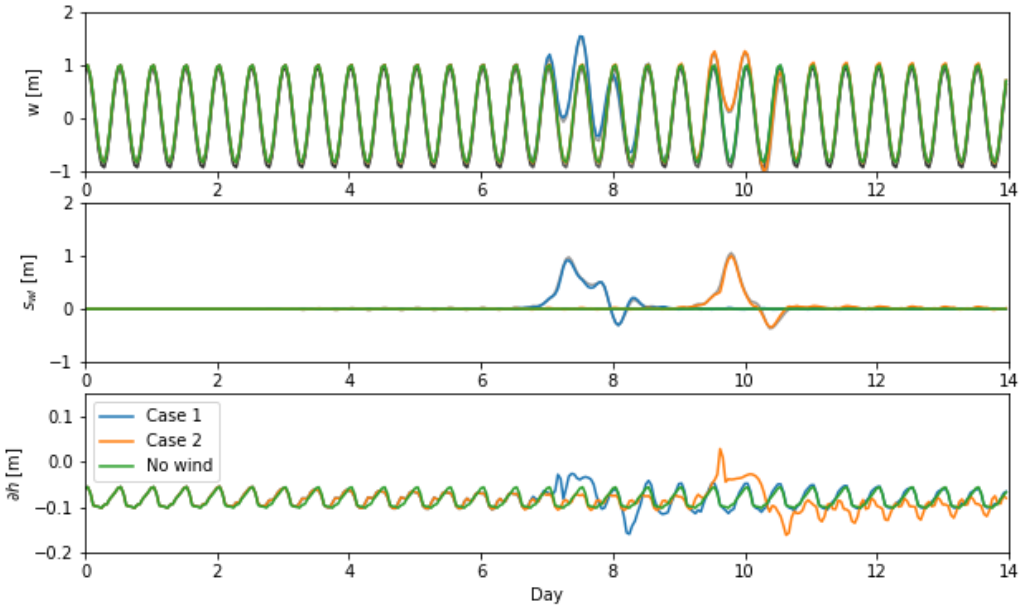


Figure 4.20: Top graph shows the observed water levels, gray lines (behind colored lines) indicate the corresponding water levels of numerical models without salinity. Middle graph shows the surgesgray lines (behind colored lines) indicate the corresponding water levels of numerical models without salinity. Bottom graph shows the observed water level differences as defined in eq. 4.3.

4.5. Discussion

The setup numerical model(s) has been presented, the method for the numerical experiments have been explained and the numerical results have been analyzed. Literature offers a possibility to compare earlier findings to these numerical results. Garvine (1999) made an attempt to find the cross shore width of a river plume numerical experiments, a comparison is shown below.

Figure 4.21 shows the results from this numerical experiment to the resulting fit of Garvine (1999). The scatter is quite similar compared to the scatter in Garvine (1999), maybe a bit higher. Furthermore Garvine (1999) states that the fit is "less succesfull" compared to other (for example alongshore plume length) fits he made. Another remarkable thing is that the parameterization of Garvine gives slightly lower results as observed in these results. Stricking is the fact that Garvine's fits have been made without tides and the results presented here are with tides, especially given the fact that Garvine found a lowering alongshore plume size influence of tides. On the other hand, the use of an idealized Kelvin wave (an approach not used by Garvine) will of course influence the results and possibly increasing plume width.

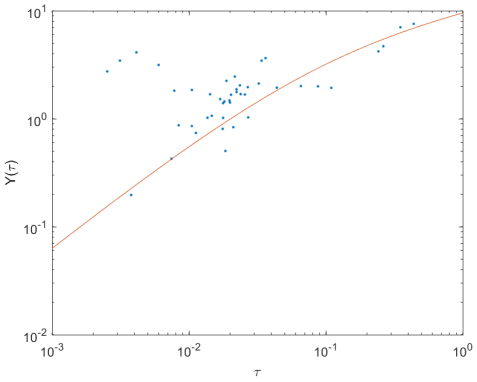


Figure 4.21: Results cross shore distance compared to Garvine (1999) fit.

5

Prediction process fresh water plumes and its uncertainty

The ultimate goal is to predict the fresh water plume width and hydraulic effect of a fresh water plume. This chapter describes the process and the uncertainties in this process. The first section elaborates the parameterizations and the physical interpretation of these parameterizations. The uncertainties are discussed and quantified in the second section.

5.1. Predition fresh water plume width and -hydraulic effect

The numerical results are used in a parameterization for interpolation purposes and a better understanding on the processes and physics. First, the plume width will be parameterized. Second, the hydraulic effect of salinity on water levels is parameterized. Finally, a physical analysis interpretation of the results is given.

5.1.1. Plume width

Garvine (1999) states in that paper "*In search of a scaling rule, I attempted a scaling analysis for* Δx_p[1] *based jointly on a simplified balance between alongshelf advection and vertical diffusion of density and on "thermal wind balance." This scaling analysis was unsuccessful because there was no simple, adequate approximation for diffusivity* $D = N$ *that was useful over a wide range of* τ*. Nevertheless, this analysis did suggest normalizing* Δx_p *by depth. ... A similar attempt at collapsing data for the coastal current width* y_c[2] *was less successful, but still useful. As buoyancy-driven coastal currents are all products of earth rotation, it is useful to scale their widths by* r_i*.*".

Garvine has in his numerical experiment a sloping bed level (and no tides), something that is not possible with the approximation of tides by a Kelvin wave as applied in this thesis. The choice for scaling the cross-shore plume length by r_i is unclear and one could argue that it is likely that Garvine's scaling of the cross-shore plume width by depth might be unsuccesfull due to the sloping bed level. The paper also shows dependency between bottom slope and cross-shore plume width which Garvine (1999) found as well in that paper.

Through symbolic regression (by the software package Eureqa) an attempt to find a parameterization with scaling to depth and internal Rosby radius (r_i) is made. The formulas for the parameterization are given in eq. 5.1 and 5.2 respectively. Furthermore, figure 5.1 shows the predicted plume width versus the observed plume width and a scaling by depth and internal Rosby radius.

[1] The alongshore plume length
[2] The cross-shore plume length

$$y_p = D\tau^{0.786}(1-\epsilon)^{\frac{6.862}{0.836K_i^{\tau-4.726}}e^{\left(\tau 0.836^{\tau^2}K_i^{20.778}\right)^{\sqrt{K_i}}}} \tag{5.1}$$

$$y_p = r_i(2.103\left(1+\frac{Re}{Re_s}\right))^{3.375^{\sqrt{\tau}}}(1-\epsilon)^{0.299e^{\frac{42.409\tau^2}{K_i}}K_i^{\frac{0.012}{3.45_{0.935}\frac{2.185}{\tau}}K_i}} \tag{5.2}$$

The scaling by Depth has a larger systematic over prediction than scaling by internal Rosby radius (mean error 7.9 km and 5.9 km respectively). Based on r^2 scores, both parameterization perform quite similar (0.96 and 0.97 respectively). However, based on maximum error (25.9 and 22.8 km respectively) and formula complexity of both scalings is equal. Therefore, the parameterization of scaling by internal Rosby radius is favored due to the fact that it is less complex and better performing based on mean- and maximum error. The results of this parameterization are shown in figure 5.1.

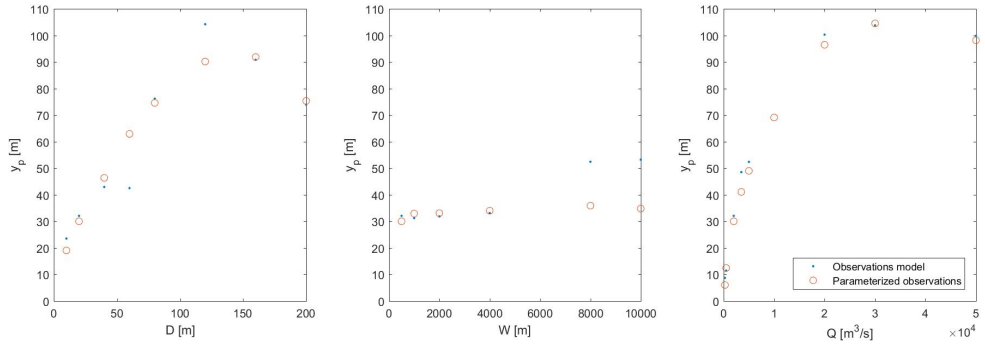


Figure 5.1: Parameterisation for scaling to Depth and internal Rosby radius

Discharge

The parameterization is able to fully capture the influence of discharge. Observations show a maximum plume width at $Q = 30.000\ m^3/s$ and the parameterization is able to reproduce this signal, as shown in figure 5.1. Furthermore, the deviations are remarkably small, with errors smaller than 8 km.

Depth

The influence of depth is complex. If the water level becomes smaller than 100 m, the plume size starts to decrease. This is probably because the tidal amplitude is roughly constant at $w_{t_0} \approx 1\ m$. Therefore, tidal influence becomes dominant which increases mixing and decreases the fresh water plume width. On a large scale the parameterization is able to capture this behavior. However, the parameterization smooths some effects and locally the parameterization produces errors.

Width

Quite some errors are found in the effect of width. Observations show, in figure 5.1, significantly larger plume width for $W > 4000\ m$ than for $W \leq 4000\ m$. However, the errors are quite, but not the observed maximum errors. Furthermore, the assumption that the parameterization produces larger errors for larger river mouths is valid.

5.1.2. Effect of salinity on water levels

Through symbolic regression (by the software package Eureqa) an attempt to find a parameterization for the slope a in the hydrostatic balance equation. The formula of the parameterization is given in eq. 5.3. Furthermore, figure 5.1 shows the predicted plume width versus the observed plume width and a scaling by depth and internal Rosby radius.

$$a = 0.881 \frac{e^{0.751(1-eps)^{Ir}}}{e^{\frac{2.19\tau \frac{g'}{9.81} 1.73^{K_i}}{K_i}}} \tag{5.3}$$

The parameterization for the slope a in the hydrostatic balance equation has no systematic over- or under-prediction and a $r^2 = 0.95$. Furthermore, the maximum error is equal to 0.14. Figure 5.2 shows a comparison of the parameterization to the most responsive parameters found in subsection 4.3.3.. The graph shows good agreement between the parameterization and depth, discharge and tidal amplitude. Errors are for these parameters very small. Width is significantly less good parameterized by the parameterization and errors increase for larger river mouths. Nevertheless, the parameterization is very useful for predictions.

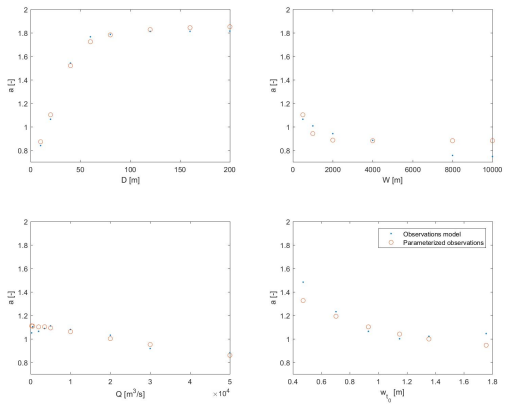


Figure 5.2: Parameterisation for the slope a in the hydrostatic balance equation.

5.1.3. Prediction process

In order to predict the plume width or the hydraulic effect of a fresh water plume the density difference between the river mouth and the shelf water is an essential input for the parameterization. Multiple parameters of the respective parameterizations (τ, K_i $Reand$ g') require the depth averaged density at the river mouth. This section presents two methods which can be used in order to predict the density difference at or near the river mouth, based on numerical model results and based on in situ measurements of the Rhine mouth. The first method is able to predict, with good accuracy, depth averaged density profiles and the last method is able to predict salinity at a constant depth. The last method has been added to show that a relatively simple classification ensemble is able to predict salinity based on in-situ measurements and potentially very well suitable for this application.

Reduced gravity parameterization

The results from the numerical model are used to estimate the reduced gravity (g'), based on discharge, depth- and width of the river mouth. However, this parameterization is not able to handle weather influences, since idealized conditions are present in the numerical model.

The parameterization of the reduced gravity is very accurate ($r^2 = 0.99$, parameter estimates found by non-linear regression). Under very high discharge the water column at the river mouth could become completely fresh water, implying a maximum for the reduced gravity. Furthermore, a relation based on Q, W and D, which units is agree with the units of g' $[m/s^2]$, is preferred. Unfortunately, after many tries a relationship with matching units is not found. The parameterized equation is shown in eq. 5.4 and the scaling of g' and performance of the parameterized equation are shown in figure 5.3. This parameterized equation is used for interpolation, which is necessary in understanding the influence of physical properties.

$$g' = min\left(0.065\left(\frac{Q}{D\sqrt{WD}}\right)^{0.684}, \frac{\rho_s - \rho_0}{\rho_s} g\right) \tag{5.4}$$

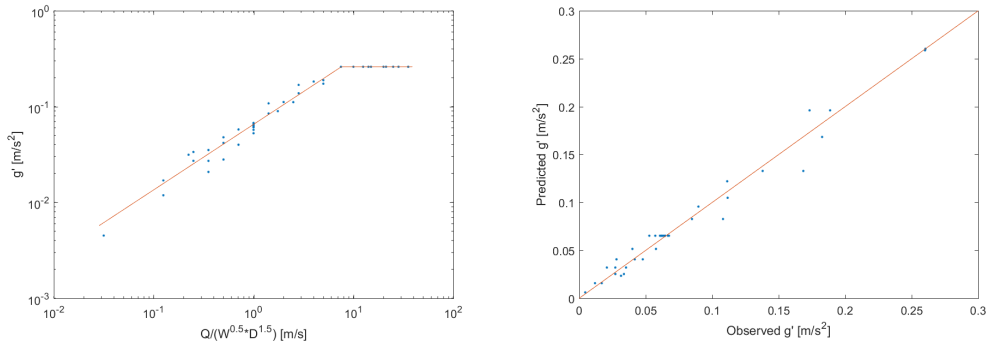


Figure 5.3: The left figure shows the scaling of g' to $Q/D\sqrt{WD}$. The right figure shows the performance of the parameterization

In-situ measurements

For this application the same dataset for water levels, surges and discharge has been used and a new time-continuous data set for salinity has been used. Salinity is calculated by a conductivity and temperature data set from Rijkswaterstaat (sampling location: Hoek van Holland, -4.5 m N.A.P.), which is open-source through https://waterinfo.rws.nl. Conductivity and temperature are later converted to salinity using the Unesco formula (Tables and Standards, 1981). This method is able to handle weather, produce real-time predictions (given prediction datasets) and able to handle discontinuities. The ultimate goal is to predict salinity values based on basic properties of the river Rhine, such as discharge, water level (and surge) and weather variables. Based on historical data from Rijkswaterstaat and KNMI (dutch weather institute) an effort has been made to create a model that is capable of predicting this. Currently, the model is only able to predict salinity at a single location in the water column. This method has been added to show it is likely that such a method is able to predict depth-averaged salinity, given a depth-averaged dataset.

However, the results of these attempts have been unsatisfactory ($r^2 < 0.5$), due to the fact that this problem is state-dependent. In other words, salinity depends not only on water levels, surge, discharge, but also strongly on salinity in previous time steps. This is found to be a major influence on salinity in future time steps. Even for predictions based on daily average historical data, which should show less state dependence, still showed state-dependency and therefore, even these attempts were unsatisfactory ($r^2 < 0.75$). For this reason, a prediction system, based on big data techniques which uses historical data and the state of the system in previous time steps, for salinity has been made.

A bootstrap-aggregated ensemble of regression trees has been trained to predict salinity the next timestep based on the current and the expected water levels, surges and discharge and the salinity in the previous time step. The model is able to cope with missing data, The model has been trained based on data (150.000 data points) in the period between 1984 and 2006 ($r^2 = 0.89$). In a later stage the model has been validated with data from the period between 2006 and 2016 ($r^2 = 0.91$). Unfortunately the model has not been used in collaboration with a prediction system for water levels and river discharges, however with perfect predictions the model performs excellent. Results from the model are shown in figure 5.4. Figure 5.5 summarizes the performance of the model.

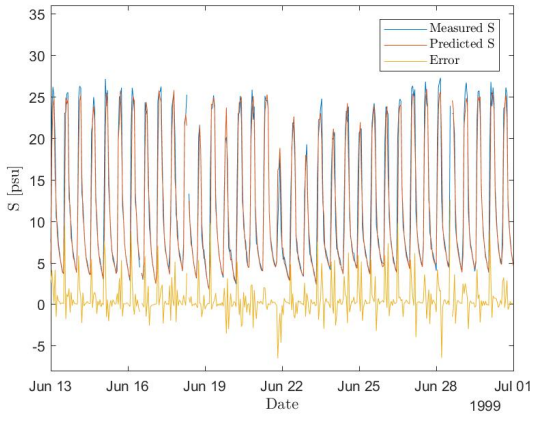


Figure 5.4: Example of the bootstrap-aggregated ensemble of regression trees trained to predict salinity

In order to come to this model several regression-models have been tested and even additional weather variables, such as wind speed, -direction and pressure, have been used. However, these additional weather related variables turned out to be irrelevant for the accuracy of the bootstrap-aggregated ensemble and are therefore not used in the final model. These weather variables could be irrelevant due to (1) the fact that conductivity and temperature are sampled at -4.5 m N.A.P. (roughly 4.5 m below mean sea level) and (2) the fact that the measurement location is located just (approximately 5.5 km) up stream of the river mouth. Due to these reasons, different mixing processes, in which weather variables have a smaller influence, govern salinity at that specific location.

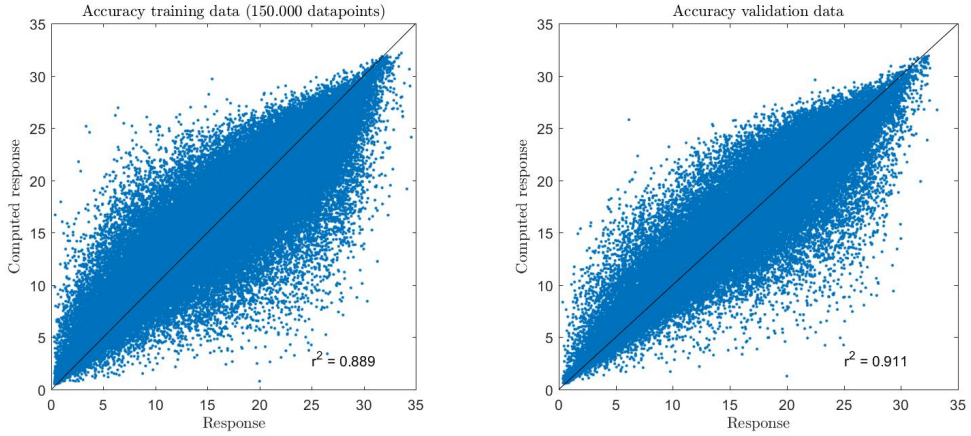


Figure 5.5: Bootstrap-aggregated ensemble of regression trees trained to predict salinity

Furthermore, one could question if this type of model is able to predict salinity for larger time steps. This can be achieved by predicting salinity for the next timestep (1 hour), then use that predicted value for a next prediction and so on. The main largest driver for mis predictions could be divergence of the prediction due to errors in the previous prediction (which are input for future predictions). After briefly testing this method, no divergence is shown and even a slightly higher coefficient of determination is found ($r^2 = 0.92$). The model is able to predict daily averages as well, figure 5.6 shows these results. The model is even more accurate with $r^2 = 0.95$ and $RMSE = 0.75$. Therefore, this prediction model can be used as an input for the parameterizations for plume width and the hydraulic effect of fresh water plumes.

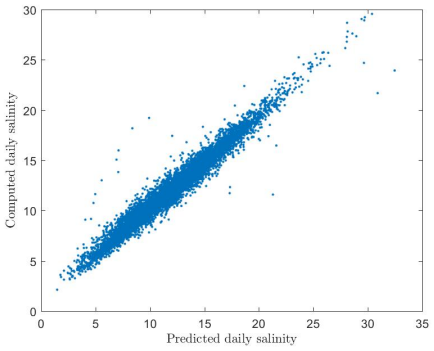


Figure 5.6: Bootstrap-aggregated ensemble of regression trees trained to predict salinity, daily performance

5.1.4. Physical analysis

An effort has been made in this subsection to capture the response of the fresh water plume, based on bulk parameters, through the numerical results and the corresponding parameterizations. These parameterizations ensure that a single expression captures all regimes determining the plume size and its effect on local water levels. This chapter analyzes the governing physical regimes in the parameterizations and its implications for fresh water plume width and the hydraulic effect on water levels of these plumes. Both parameterizations, for plume width and the hydraulic effect, are governed by the reduced density. Therefore the parameterization of this parameter, which is in dimensional terms incorrect, is used for interpretation purposes and it is interesting to see how these parameterizations combined perform.

Plume width

The plume width is, as shown in figure 4.6, mainly influenced by discharge, river mouth width, depth and tidal amplitude. The dependence of these parameters is shown in figure 5.7 and 5.8.

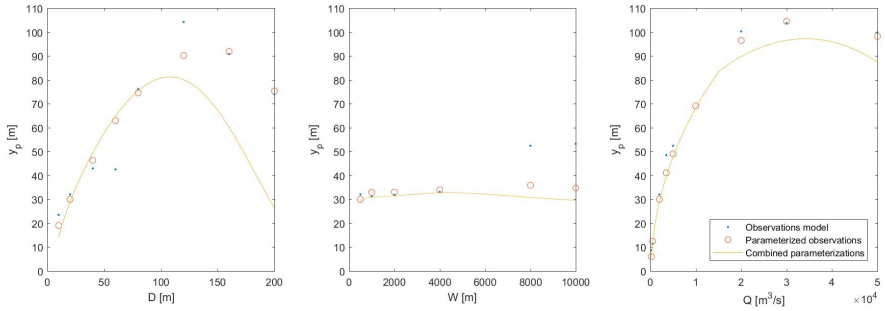


Figure 5.7: Parameterisation for plume width by scaling with internal Rosby radius, dependent on depth (left graph), width river mouth (middle graph) and discharge (right graph).

Figure 5.7 and 5.8 show that the parameterization is capable of producing quite good estimates. Furthermore, it is able to capture the complex processes governing plume width. Of course some errors are visible. These errors (mostly visible for depth) are mainly induced by a difference in observations in tidal amplitude and the assumed tidal amplitude in the parameterization. In the numerical model the tidal amplitude is measured at the river mouth. Furthermore, tides are generated at the southern boundary through a Kelvin wave with reducing along- and cross shore tidal amplitude, which location is dependent on discharge. Therefore, there is no exact influence on the tidal amplitude in the numerical model. The parameterization is very sensitive to tidal amplitude, since it is the only plume dissipating parameter. Figure 5.9 shows epsilon for each depth. Although the deviation is very small, the sensitivity to this parameter is large and it is a reasonable explanation for deviations between the combined parameterizations results and the numerical observations.

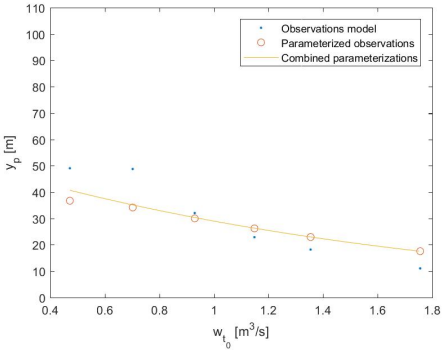


Figure 5.8: Parameterisation for scaling by internal Rosby radius, dependent on the tidal amplitude.

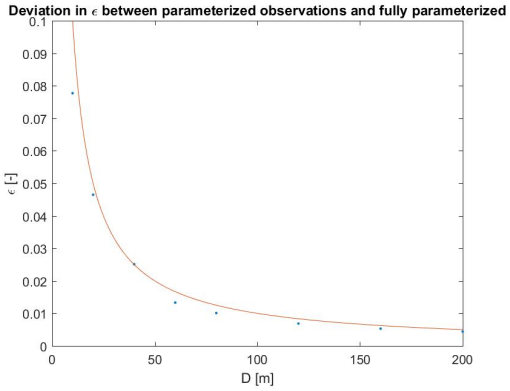


Figure 5.9: Relation between depth and epsilon. Blue dots show the observed epsilon in the numerical model and the red line is the assumed epsilon in the parameterization comparison.

Hydraulic effect of salinity on water levels

The hydraulic effect of salinity on water levels is governed by the slope a in the hydrostatic balance equation. The slope a in the hydrostatic balance equation is mainly influenced by discharge, river mouth width, depth and tidal amplitude. The dependence of these parameters is shown in figure 5.10.

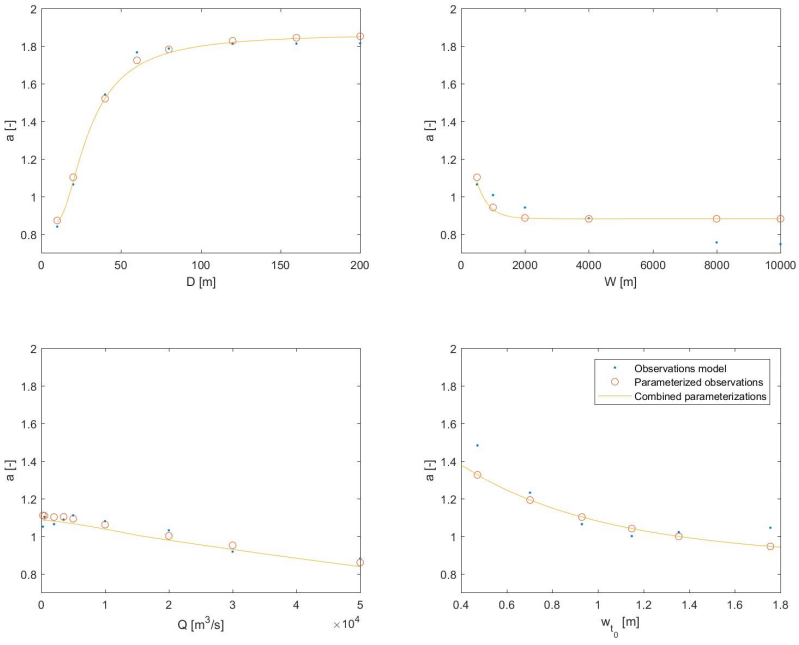


Figure 5.10: Parameterisation for a dependent on depth (top left graph), width river mouth (top right graph), discharge (bottom left graph) and tidal amplitude (bottom right graph).

The combined parameterizations perform similar to observations only. Larger errors in the parameterizaed slope a in the hydrostatic balance equation are only observed for larger widths. The conclusion can be drawn that the physical processes are fully captured by the parameterization.

However, by parameterizing the slope a in the hydrostatic balance equation the water level differences have not been computed yet. Therefore, using these parameterizations water level differences induced by salinity can be computed. This method or framework is most definitely use full for prediction purposes. The next page offers an overview of the perfomance of the complete procedure for water level difference predictions at the river mouth, the coastal location with the highest observed water level differences.

Figure 5.11 gives an overview of the computed water level differences through a combination of parameterizations. Predictions are quite well for width, discharge and tidal amplitude and only for depth the combination parameterizations seems to model the incorrect behavior, although the error are still reasonably small.

The maximum water levels difference observed at 20.000 m^3/s is parameterized correctly. Given this discharge the complete water column is fresh water at the river mouth and for higher discharges the observed water level difference becomes lower.

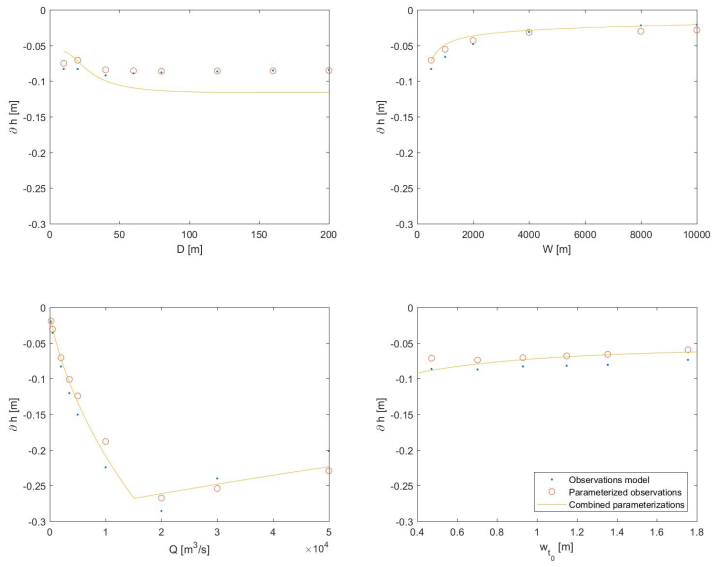


Figure 5.11: Results of approximating water level differences through a combination of parameterizations.

Rivers worldwide

The parameterizations can be used for the estimation of plume size and the effect of salinity on water levels for real rivers. Using the combination of parameterization for g' and the plume width or the hydraulic effect on water levels, the respective scale/effect can be quantified. This analysis is performed to illustrate the effect of salinity on water levels and scale for real applications. In order to perform this the coastline has to be schematized and information on discharge, salinity and latitude has to be estimated. The analysis is intended to give an indication of the effect and spatial scale. Therefore, the four biggest rivers have been selected for this application and precise schematizations of these rivers are not made.

Table 5.1 shows the input parameters for each river. Discharge and salinity are based on the mean discharge and mean salinity which can be found at Wikipedia. The width of the river mouth and the latitude have been estimated based on Google Maps. The Depth has been estimated on the bathymetry of the Global Tide and Surge Model (and Emodnet). The tidal amplitude is based on a M_2 tidal amplitude chart.

Table 5.1: My caption

River	Q $[m^3/s]$	W $[m]$	D $[m]$	S $[psu]$	w_{t_0} $[m]$	Latitude [°]
Amazon	209,000	70,000	20	36	1	0°
Congo	41,200	5,000	35	34	0.4	6°
Ganges	38,120	10,000	20	31	0.75	22.5°
Yangtze	30,166	12,000	12.5	31	0.8	30°

Figure 5.12 gives an overview of the results. The plume width is largest for the River Amazon. However, the hydraulic effect of salinity on water levels is largest for the river Congo, due to the larger water depths near the river mouth. The two largest rivers have an effect of approximately 0.24 m and the other two have an effect of 0.15 and 0.12 m. The effect of salinity on water levels is mainly governed by discharge and depth (the effect gets larger with increasing discharge and depth). This means that these river are not necessarily the rivers with the largest effect on water levels.

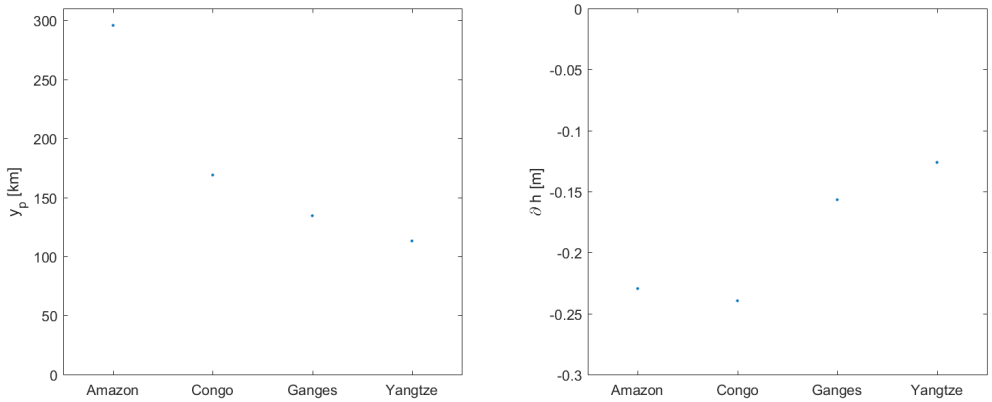


Figure 5.12: Right graph gives an overview of the plume width of each river. The left graph gives an overview of the hydraulic effect of salinity on water levels. Both graphs have been made using the parameterization for g' and the respective parameterization for plume width or the hydraulic effect of salinity on water levels.

5.2. Parameterization uncertainty

The reliability and range of applicability of the parameterization is governed by two factors, the model/parameterization and conditions present at the moment of assessing the hydraulic effect and plume size. This section captures both aspects, first the model uncertainty is assessed through a vine copula. The next section contains the uncertainty through a data set at the river mouth. This chapter will end with a description on how these methods can be combined, to give real time worldwide predictions on both the hydraulic effect of river plumes and the plume width.

5.2.1. Model uncertainty

This section gives an example for the application of the plume width - discharge relation which is parameterized in section 5.1. By fitting a copula to the observed plume width and the parameterized plume width the model uncertainty can be quantified and visualized. Figure 5.13 shows the performance of the parameterization, the graph shows a high correlation between the observed and predicted fresh water plume width ($r^2 = 0.96$).

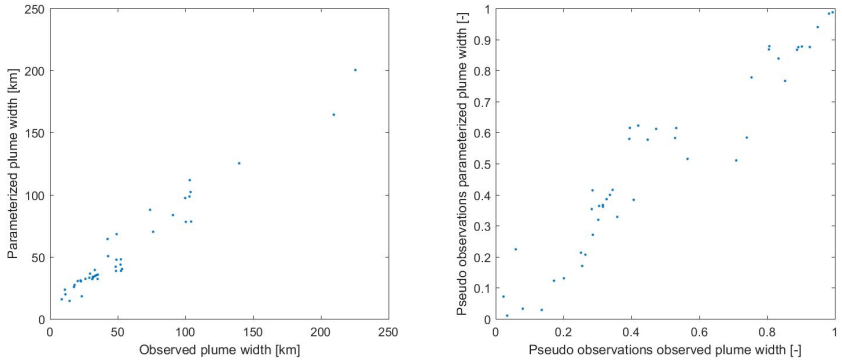


Figure 5.13: Left graph shows the observed plume width versus the parameterized plume width. The right graph shows the pseudo observations of the same variables, which are obtained by fitting parametric distributions to the dataset.

The parameterization for the plume width with scaling by internal rosby radius (r_i, see eq. 5.2) is used to predict the plume width. Predictions are made based on bulk parameters (for example: discharge, depth, width, etc) using the parameterization for the reduced gravity (g', see eq. 5.4). This is the same method as the method in section 5.1. Figure 5.13 shows the performance of the parameterization. The pseudo observations are obtained by fitting parametric distributions and comparing each fit by AIC. The survival plots of the five best fitted distributions are shown in appendix G; figure G.1. In this case, consistency is preferred and for both variables the Inverse Gaussian distribution has been used.

A copula is fitted to pseudo observation of the observed plume width and the parameterized plume width. Based on automatic fitting procedure of the R-package "VineCopula", with selection criterion AIC, a Frank copula (par = 22.32) is selected and used in a conditional sampling procedure. The conditional sampling procedure is shown in eq. 5.5, where $u_{1;fixed}$ are the conditional value (n the same values) and u_2 are n independent random values between 0 and 1. By estimating the 5% and the 95% quantiles and repeating the experiment for each conditional the confidence bounds are estimated as pseudo observations.

$$\begin{aligned} v_1 &= u_{1;fixed} \\ v_2 &= F^{-1}_{r12;x_1}(u_2) \end{aligned} \tag{5.5}$$

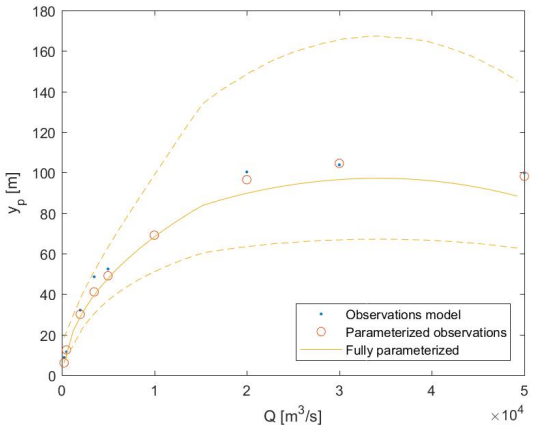


Figure 5.14: The parameterized plume width as a relation of the discharge (Q), yellow dashed lines indicate the 5% and 95% confidence lines.

The pseudo observations of the confidence bounds of the parameterization are estimated, through a conditional sampling procedure. Using the inverse cdf of the inverse Gaussian distribution the actual confidence bounds are estimated ($X_i = \hat{F}(v_i)$). The result is shown in figure 5.14.

This section shows, briefly, the use of statistical models for the application of acquiring confidence bounds for parameterized equations. Flexibility is ensured with the use of these multivariate statistical models, since several copula families can be used in the fitting process.

5.2.2. Data uncertainty

In order to fully understand the (un)certainty of the parameterization a statistical model is developed which captures the dependency structure of the data. First the dataset is discussed, then the correlation structure is discussed. Then the transformation of each variable to it's marginal is presented. Finally, the development of the statistical model, a vine copula, is described.

Data

The same dataset as used in section 3.2 is used for the statistical model. More information on the collection of data can be found in chapter 3. The dataset consists of daily average water levels and surges, which have both been sampled hourly, daily mean discharge and a irregularly sampled (approximately once per two weeks) data set for sea surface salinity. All these measurements are made near Hoek van Holland. The time series of this dataset covers a large period, from 1971 until 2016. An overview of the dataset is given in figure 5.15. As shown in the graphs, the data set for salinity is significantly smaller as for the other variables.

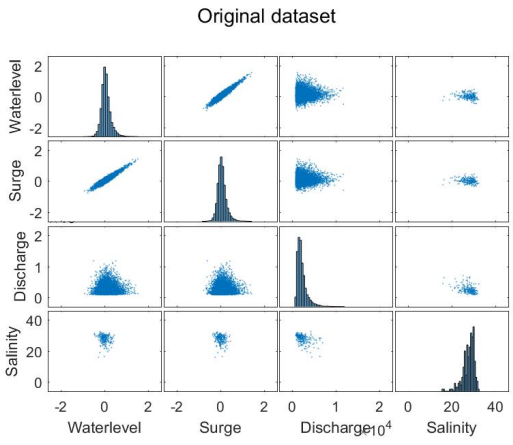


Figure 5.15: Correlation structure for data uncertainty

Correlation structure

The complete correlation structure is given in table 5.2. As shown in the graphs of figure 5.15, the data set for salinity is significantly smaller as for the other variables. It should be noted that this implies that the confidence bounds for rank correlations (as shown in eq. 2.15) are significantly larger due to this smaller sample size.

As expected, water levels and surges are highly correlated. Discharge is nearly uncorrelated to water levels and surges and it shows quite strong negative correlation to salinity. Furthermore, salinity shows some negative correlation to water levels and surges.

Table 5.2: Correlation original dataset

	Var 1, Water level	Var 2, Surge	Var 3, Discharge	Var 4, Salinity
Water level	1.00	0.89	0.05	-0.17
Surge		1.00	0.027	-0.23
Discharge			1.00	-0.45
Salinity				1.00

Pseudo observations

The psuedo observations are obtained by fitting a distribution to each variable of the dataset, all used fits are shown in table 5.3. The survival plots of the five best fits (selected by AIC) are shown in appendix H; figure H.1. By transferring the variables to pseudo observations (u_i) through the CDF's of the fitted function ($u_i = \hat{F}_x(x_i)$).

Table 5.3: Distribution families used for finding the marginal distribution of each variable. Each distribution is fitted to the dataset by maximum likelihood estimators (except for the gaussian mixtures, these are fitted by the Expectation Maximization algorithm), as stated in section 2.4.3.

	Distribution family	Parameters	Parameter values
Water levels	Gaussian mixture distribution - 3 components	π_1, μ_1, σ_1	4.54e-1, 1.32e-2, 1.2e-2
		π_2, μ_2, σ_2	4.64e-1, 8.02e-2, 4.81e-2
		π_3, μ_3, σ_3	8.15e-2, 3.14e-1, 1.37e-2
Surges	Gaussian mixture distribution - 3 components	π_1, μ_1, σ_1	5.56e-1, 1.65e-2, 1.46e-2
		π_2, μ_2, σ_2	3.81e-1, 1.04e-1, 4.61e-2
		π_3, μ_3, σ_3	6.27e-2, 3.16e-1, 1.27e-2
Discharge	Gaussian mixture distribution - 6 components	π_1, μ_1, σ_1	1.84e-1, 1.20e3, 3.61e4
		π_2, μ_2, σ_2	3.86e-1, 1.73e3, 9.51e4
		π_3, μ_3, σ_3	2.66e-2, 2.45e3, 1.88e4
		π_4, μ_4, σ_4	1.22e-1, 1.65e-2, 5.02e5
		π_5, μ_5, σ_5	3.62e-2, 1.04e-1, 9.94e5
		π_6, μ_6, σ_6	6.18e-3, 3.16e-1, 1.93e6
Salinity	Extreme value distribution	μ,σ	28.95, 1.94

Figure 5.16 shows the pseudo observations of each variable. The correlation structure between each set of pseudo observation is now visible. Furthermore, the same correlations are visible for the sets of pseudo observations, since rank correlations are used and the cdf's are monotomely increasing functions. For the next steps in building a multivariate statistical model it is essential that each marginal is uniformly distributed. The histograms of each variable are shown in figure 5.16 on the diagonal. The figure shows that the marginals of the variables water levels, surges and discharge are clearly uniformly distributed. However, the marginal of the variable salinity is quite uniform, but most certainly not perfect. Due to a lack of observations (in total 250) this is inevitable.

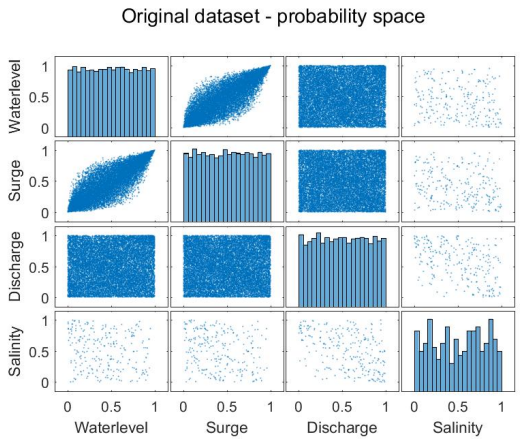


Figure 5.16: Correlation structure for pseudo observations data uncertainty

Vine copulas

Vine copulas are used to model the statistical dependence structure. In order to fit a vine copula model the R-package VineCopula (Schepsmeier et al., 2018) is used. The R-package VineCopula has the implication that it is unable the assess the goodness-of-fit for all copula families. Therefore, the vine structure is selected by CvM and in a next step the new vine structure is selected via the fitting algorithm of the VineCopula package.

Vine structure

A c- and d-vine have been fitted to the dataset with the structure selection function of the R-package VineCopula. In this structure selection function the selection method is AIC for selecting a pair-copula. The structure of these vines is shown on the right side. The c-vine performs better compared to the d-vine based on a Cramér-von Mises goodness of fit statistic (with CvM scores of 0.0696 and 0.0987 respectively). Therefore the c-vine is used in the next analysis.

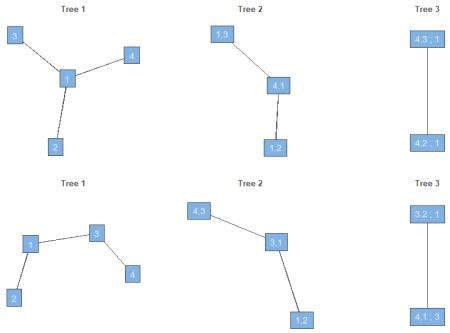


Figure 5.17: Top figure gives an overview of the c-vine and the bottom figure gives an overview of the d-vine

Table 5.4: Overview of c- and d-vine trees and families

C-vine copula with the following pair-copulas:		D-vine copula with thefollowing pair-copulas:	
Tree 1:	Family	Tree 1:	Family
1,2 1,3 4,1	Gumbel (par = 3.71) t (par = 0.06, par2 = 11.97) Rotated Clayton 90 degrees (par = -0.28)	1,2 3,1 4,3	Gumbel (par = 3.71) t (par = 0.06, par2 = 11.97) Gaussian (par = -0.42)
Tree 2:	Family	Tree 2:	Family
4,2;1 4,3;1	Independence Gaussian (par = -0.45, tau = -0.3)	3,2;1 4,1;3	Frank (par = -0.31) Rotated Clayton 90 degrees (par = -0.33)
Tree 3:	Family	Tree 3:	Family
3,2;4,1	Independence	4,2;3,1	Rotated Clayton 270 degrees (par = -0.24)

Vine copula model

Based on CvM (see Vine structure) a c-vine is the preferred vine structure. Via the R-package VineCopula a new c-vine is fitted to the dataset, using all copula families. This implies that it is not possible to asses the fit of the vine copula to data by CvM. Due to the fact that the number of copula families has increased, the fit will be at least as good and most likely better. The results of this fitting procedure is shown in table 5.5. Only the copula familiy between Variable 1 (Water levels) and Variable 2 (Surges) has changed compared to the previous fit and a BB7 copula has replaced the Gumble copula family.

Table 5.5: Overview of c- and d-vine trees and families

C-vine copula with the following pair-copulas:	
Tree 1:	Family
1,2 1,3 4,1	BB7 (par = 4.26, par2 = 1.92) t-copula (par = 0.06, par2 = 11.97) Rotated Clayton 90 degrees (par = -0.28)
Tree 2:	Family
4,2;1 4,3;1	Independence Gaussian (par = -0.45, tau = -0.3)
Tree 3:	Family
3,2;4,1	Independence

Resampled pseudo observations can be obtained for the fitted c-vine through the sampling procedure of the vine structure (see chapter 2.6.4). These psuedo observations, with the same number of samples in the dataset, are shown in figure 5.18 (left graph). After sampling the pseudo observations a re-sampled dataset can be obtained by transferring the pseudo observations back via their original distributions (see eq. 5.6). Figure 5.18 (right graph) shows the final results of the re-sampled dataset (for a same sized dataset as the original dataset). The difference (apart from the missing data in the original dataset) is very small. Furthermore, the model is able to fully capture the dependency structure of all variables.

$$x_i = \hat{F}_i^{-1}(u_i) \tag{5.6}$$

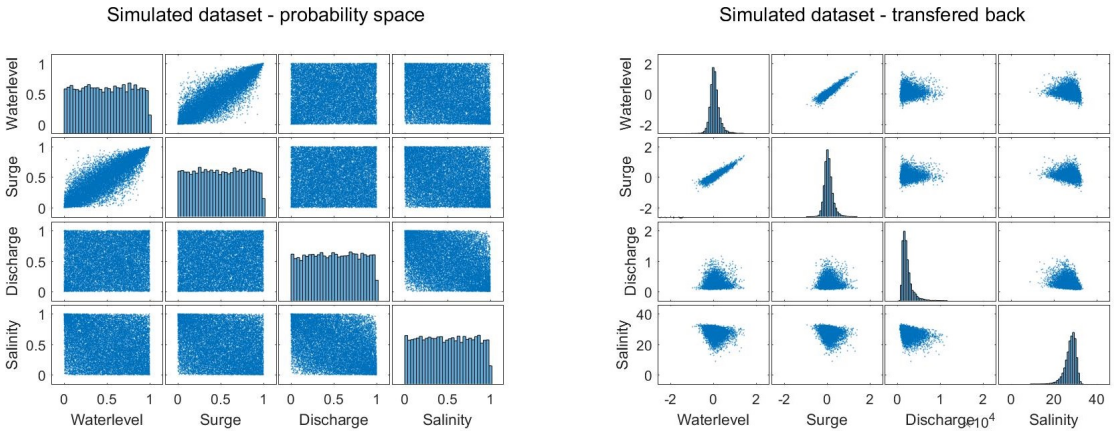


Figure 5.18: Correlation structure resampled vine copula data uncertainty, left graph shows the resampled pseudo observations and right graph shows the resampled data.

6

Conclusions and recommendations

First the conclusions are presented in the first section and the second section contains the recommendations.

6.1. Conclusions

The general conclusions are presented in the first subsection. The research questions have been answered in the second subsection.

6.1.1. General conclusions

This thesis aims to estimate the size and hydraulic effect of fresh water plumes induced by river outflow to coastal waters. Through a combination of data analysis, a numerical model and -experiments, parameterization and statistical analysis insights in the complex behavior of fresh water river plumes, compound floods and the effect of salinity on water levels have been gained. A numerical model has been developed to research the effect for the Rhine-Meuse river mouth and by changing bulk parameters (discharge, width river mouth, depth, tidal amplitude and latitude) the effect has been researched for worldwide applications. Based on literature, a priori knowledge of the system was needed in order to make these predictions. This chapter gives the most important conclusions.

The data analysis focuses on the location of the Rhine ROFI, the effect on water levels, the correlation structure of compound floods (the combination of coastal set-up and high river discharge) and return periods. Based on these analysis, salinity has a minor effect on water levels, but the scale of the Rhine ROFI is significant. The effect of salinity on water levels is not well visible in the analysis of return periods. The effect of tidal amplitude (which increases towards the south of the Netherlands) and coastal shape dominate the analysis. Compound flood events (the combination of coastal- and pluvial flooding) are quite low correlated in the Neterlands, if no lag is introduced between the time series of water levels and discharge. Furthermore, the correlation becomes largest for a lag of 6 days if the discharge time series is lagged with respect to the water level time series. The semi-correlation r_{NE}, which is an important measure for the tail dependence, is maximum at a lag of 4 days.

In order to research the effect of plume width and salinity on water levels an idealized hydraulic numerical model of the river Rhine has been developed. The model is able to simulate fresh water river outflow in combination with tides, in which tides (S_2) are approximated by a Kelvin wave with linearized bottom friction. The numerical model can be adapted in such a way that the depth, width of the river, discharge, salinity and tidal amplitude are varied, while performing up to standard. A tremendous amount of time has been spent on the development of this model and getting boundary conditions correct for the complete parameter space.

The numerical model is used in a numerical experiment. By varying the discharge, the uniform bathymetry, width of the river, salinity and tidal amplitude the influence of each parameter can be estimated. The results of the numerical experiment are used to analyze (1) the cross-shore plume width and (2) the hydraulic effect of salinity on water levels by dimensionless parameters used in previous research. The plume width is most responsive to discharge, depth and tidal amplitude. Furthermore, restarting the numerical model with a higher discharge has shown that the plume will grow over time to the expected plume width and under the presence of surges the hydraulic effect of salinity on water levels decreases (based on two numerical simula-

tions).

In literature comparable experiments, without tidal interactions, have been performed and scaling the results was challenging in these experiments. For this reason and the fact that complex behavior in the numerical results is observed, the results of the numerical experiment have been parameterized by symbolic regression. The symbolic regression are intended for interpolation, analysis and prediction purposes. The results of symbolic regression shows quite good performance and the parameterized plume width shows a high sensitivity to tidal amplitude. Furthermore, the slope a in the hydrostatic balance equation has been parameterized. Both parameterizations are dependent on the reduced gravity g' (via τ and K_i). an effort has been made to parameterize this parameter as well, unfortunately a regression with dimension mismatch has been found which is performing well ($r^2 = 0.99$). Using a combination of parameterizations the performance of parameterizations for the plume width and hydraulic effect are assessed. Both parameterizations are able to capture the complex behavior which is present in the results of the numerical experiments and able to produce predictions for world wide applications.

Multivariate statistical vine copula models were used with two purposes; (1) to estimate uncertainty of the parameterizations and (2) for the application to compound floods. The uncertainty of the parameterization is quantified through a conditional copula sampling procedure and the uncertainty in the dataset of the river Rhine has been researched through a vine copula model. The c-vine structure has been selected by CvM, based on a vine copula with a limited number of copula families. The final c-vine copula model is able to reproduce the complete correlation structure and sample a matching dataset.

The ultimate goal of this thesis is to understand the influence of salinity on water levels, the relation to compound flood events and estimate the importance of this effect for global water level predictions. The maximum observed water level different between a numerical model with salinity and without salinity (but with the same discharge, depth, salinity, tidal amplitude and width) is 0.31 m, a results that is closely approximated by the combination of parameterizations. Furthermore, numerical results for two analysis showed that the hydraulic effect of salinity decreases under the presence of surges, which were induced by extreme wind conditions. All together, the conclusion can be drawn, based on the results of the idealized numerical model, that the effect of salinity is relevant for world wide water level prediction and, to a lesser extend, for flood predictions and compound flood events.

6.1.2. Research questions

First the main question is answered and second the sub-questions are individually answered.

Main research question

Can the effect of salinity differences on (spatial) variability in water levels near a river mouth be computed by a combination of numerical modeling and probabilistic models and reduce the amount of data needed while holding good estimates of exceedance probabilities for water levels?

In this thesis a numerical Delft3D-FM model was used to research the fresh water plume width and the hydraulic effect of salinity on water levels. The results, based on dimensionless parameters, of the numerical experiment showed complex behavior. Therefore, probabilistic models, which are only able to model monotonously increasing- or decreasing relations, have not been used (although this was initially intended). Symbolic regression has been used to parameterize in order to interpolate the numerical results, for a physical analysis and for prediction purposes. The numerical models produce large amounts of data, however after processing these results efficient estimates can be made.

Subquestions
Below each subquestion is answered.

1. *What is the exceedance probability in water levels near the mouth of the Meuse based on observations only?*
   The exceedance probabilities in water levels and surges are presented in figure 3.12. Tidal influence is dominant for the river mouth Rhine-Meuse, which results in higher water levels for the same exceedance probability in southern direction. The coastal shape is also expected to be a major influence on the spatial distribution of flood risks. These results are also compared to the Dutch flood risk standards and these confirm the results. Furthermore, surges show the same behavior which is explained by the funnel shape of the North Sea.

2. *What is the relationship between salinity differences and water levels?*
   The relationship between salinity and water levels has been researched by (1) a data analysis and (2) a idealized numerical model. The data analysis shows that salinity influences water levels, however the influence is small compared to tidal- and weather influences. The idealized numerical model shows influence of salinity on water levels and depending on the bulk parameters (discharge, width river mouth, depth, tidal amplitude, etc.) and the maximum observed water level difference is 0.31 m. The hydraulic effect of salinity has been researched based on steric height anomaly and a hydrostati balance equation. The last is favored since (1) steric height anomaly shows a square root relationship and (2) time-averaged water level differences were researched. Based on two numerical model experiments, the hydraulic effect tends to decrease under the presence of surges which are induced by extreme wind conditions.

3. *What is the influence of salinity differences on the variability in water levels at the mouth of the Meuse compared to the variability in water levels at other locations at the North Sea?*
   The variability in water levels are shown in table 3.1 and figure 3.4 and the variability in water levels is mainly influenced by the tidal amplitude, which is decreasing from South to North. Influence of salinity on the standard deviation in water levels is negligible. The same holds for the exceedance probabilities in water levels. However, the influence of the Rhine ROFI is visible in the mean water level, higher mean water levels are visible near the river mouth.

4. *How does the numerical model perform compared to real observations and is this performance sufficient for this application?*
   The numerical model is an idealized model, which aims to model the most important dynamic interactions between salinity and water levels. The model performs comparable to (De Boer, 2009) numerical model, while being computationally more efficient. Furthermore, the model can be modified to model The performance for this applications is sufficient.

5. *What is an efficient way of parameterizing the results from the numerical model for the statistical analysis?*
   The numerical results have been parameterized by symbolic regression, which is an efficient method if complex relations and simplifications are hard to estimate. (Garvine, 1995) had the same problem, however at that time such methods were not yet available. However, in a computational sense the method of symbolic regression is far from efficient, due to the evolutionary process.

6. *Does a probabilistic model combined with a simple physical model give a better fit compared to a probabilistic model based on observations only?*
   The strength of the simple physical (numerical) model combined with a probabilistic model is that it is able to model events which are not present in datasets. The simplified numerical model will introduce additional uncertainties and errors. Therefore, a probabilistic model combined with a physical model will not give a better fit compared to a probabilistic model based on observations only.

6.2. Recommendations

Within this thesis the size and hydraulic effect of fresh water river plumes is researched. This chapter gives the recommendations which are found by working on this thesis. The first subsection gives the general recommendations and a conceptual model is discussed in the second subsection.

6.2.1. General recommendations

This subsection contains the general recommendations arising from this research.

The developed numerical model performs currently adequately for this application. However, the numerical model introduces the assumption of a simplified, uniform coastal, bathymetry. At this point the implications of this assumption are unclear and further research on these implications is interesting. For instance, a sloping coastal bathymetry can be researched by a new numerical model. Figure 6.1 gives an graphical overview of a proposed model setup. The model grid is characterized by its large circular shape and its irregular grid structure, allowing computational efficiency and local refinements. Furthermore, technically tides, Kelvin wave shaped, can be generated through artificial movements of celestial bodies. This technique has the potential to even model tidal waves with decreasing or increasing tidal amplitudes. Currently such a model is non existing to our knowledge and it could potentially increase physical understanding of ROFI's.

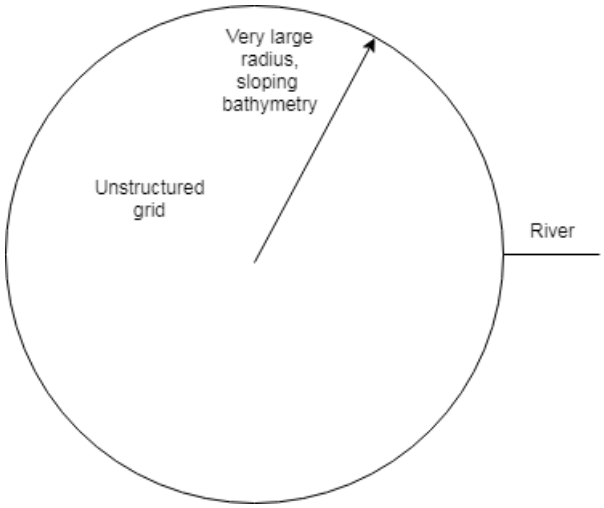


Figure 6.1: Proposal for new numerial model

Furthermore, it would be very interesting to research coastal current structures, which are induced by density differences. The developed numerical model and numerical experiment setup would be suitable for this application. One only has to keep an eye out for data storage, velocity profiles tend to take a lot of data. Researching this could lead to better physical understanding and give a better understanding on the tidal ellipses observed by de Boer et al. (2009).

At this moment computational limitations restrict the possibility of a numerical 3D model for worldwide applications with a fine unstructured grid. Therefore, currently 2D models are used for this application. However, 2D models imply that it is not possible to incorperate the effect of freshwater river outflows on coastal waters, since these models are based on the depth-averaged shallow water equations. Advances in computational efficiency will eventually solve this problem and of course uncertainty remains over the time span of these developments. Advances are made on quantum computing and this technique could potentially solve this limitation. Further developing the current approach requires computational resources and one could think that using resources for something that will be solved by computational technical advances is not useful. However, it is very usefull to develop this approach for two reasons (1) In order to verify future, computationally more efficient techniques and, models using simplified approaches can easily be used with lower computational efforts. (2) By developing these techniques a larger physical understanding of ROFI's will be obtained, which potentially could lead to significant advances in environmental sciences and economic benefits can be obtained, by for instance: shipping could become more reliable from these developments.

Statistical methods are extensively used throughout this thesis, for instance to estimate the uncertainty in the parameterizations. The use of multivariate statistical methods is inevitable in numerous application, since compound events could lead to failure (which is neglected for univariate cases). Therefore, the use of for instance univariate statistical distributions and extreme value theory are currently common practice in science and business. However, the use of multivariate statistical models, such as vine copulas, will in the end lead to more use of these techniques. Development of these multivariate statistical methods is complex and currently a field with many developments. While working with these statistical models it becomes clear

that numerous software packages (in a range of programming languages) are currently developed, but not a single package has all functions. Therefore, the development of such a package is necessary.

6.2.2. Conceptual model

The numerical results are parameterized by symbolic regression. However, a simplified physical approach could potentially be applied as well. This section contains an introduction and a description of saline mixing processes for riverine and coastal applications. Equations that need to be solved, this approach still needs some work and not all solutions are presented. Furthermore, it has not been possible to apply this approach, due to time limitations in this thesis project. Therefore, this has been added as an recommendation which can be picked up by others for further development.

The conceptual model is far from finished, but it has a lot of potential and it could lead to better physical understanding of the governing processes. Further development of the model will lead to more insight on the draw backs of assumptions. However, the conceptual model will not be able to replace a complete 3D numerical model, since mixing processes are related to water levels and vice versa through non-linear interactions.

Introduction

Fresh riverine water will mix with the saltier coastal waters. This process of mixing can be divided in two regimes: riverine- and coastal mixing, as shown in 6.2. At this stage of development, the conceptual model assumes idealized conditions: no mixing by tides and weather. Further developments of the conceptual model will allow these kind of features.

There is a large difference between riverine mixing and coastal mixing. Riverine mixing can be approximated by a 2D (down/upstream direction and depth) process, while coastal mixing is inevitable a 3D (along-, cross-shore and depth) process.

The conceptual model is based on the assumption that the influx of potential energy, which is induced by the difference in salinity between river- and coastal water, and the mixing rate. At this moment the assumption of stationary conditions and uniform bed levels are used, to simplify the process. After developments have progressed, time dependent processes could potentially be interpreted by this method.

The method could potentially be used to map a fresh water plume, induced by a river outflow. However, the method cannot fully incorporate all interactions between different process for real time processes. For example, the current structures induced by a river plume cannot be mapped over numerical results for prediction purposes, due to the non-linear interactions between these processes.

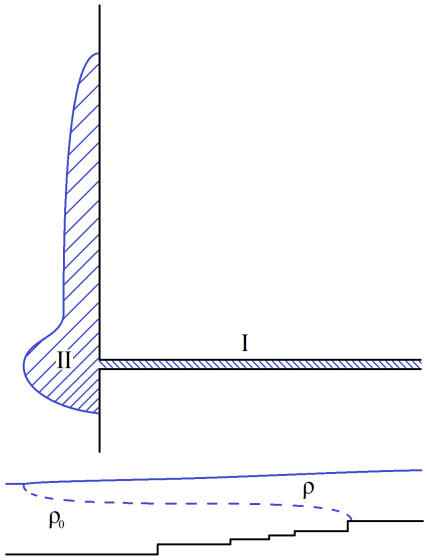


Figure 6.2: The mixing processes in the conceptual model can be split in two; riverine (I, in top graph) and coastal (II, in top graph). Bottom graph shows a cross section over the center of the river which is extended to coastal waters.

River

This section gives a descriptive overview of the mixing process on a river together with some equations. The approach is not finished, but can be used for further developments.

The water upstream of the river mouth has generally a lower salinity. Figure 6.3 gives an graphical overview of this process under idealized conditions (without mixing). The first observation under these idealized conditions is that the layer thickness of the fresh water will decrease in down stream. Consequential, based on volume balance, the speed of the layer will increase.

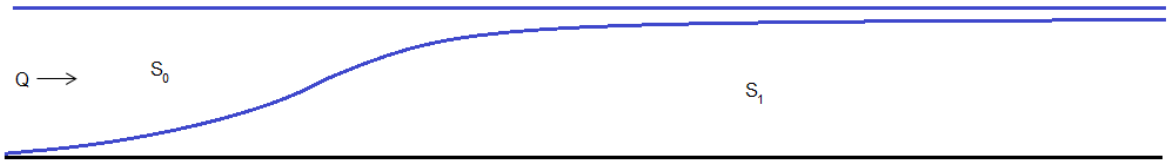


Figure 6.3: Simplified graphical representation of fresh water discharge close to a river mouth.

The depth-averaged salinity is shown in the eq. 6.1. In this equation α is equal to the ratio of the height of fresh water to the total water depth. The discharge is equal to the product of width (w_0), depth (d_0) and speed (v_0) ($Q = v_0 d_0 w_0$) and de density of the discharge is, among others, a function of the salinity S_0 ($\rho = f(S_0)$).

$$\bar{S} = \alpha S_0 + (1-\alpha) S_1 \tag{6.1}$$

Volume balance

Under the assumption that no water is added over the length of the river, the complete river holds a volume balance which is equal to:

$$Q = w v(x_0) d(x_0) = w v(x) d(x) \tag{6.2}$$

Energy balance

Under the assumption that no energy is added over the length of the river, the complete river holds a volume balance which is equal to:

$$\begin{aligned} E_0 &= w_0 v_0 \int_{-d_0}^{0} \rho_0 g z dz = -\frac{1}{2}\rho_0 g Q d_0^2 \\ E &= w v(x) \int_{-d(x)}^{0} \rho(x) g z(x) dz = -\frac{1}{2}\rho(x) g Q d(x)^2 \end{aligned} \tag{6.3}$$

Based on these equations and some additional assumptions this should be solvable. This could lead to better physical understanding and it would be interesting to see a comparison to the parameterizations. A next step would be to assume a fraction of the energy of bottom friction induced by tides goes to mixing.

Coastal

The coastal mixing process is a more complicated process, since it is a 3D process. Mixing, under the influence of tides and weather, will decrease the size of the fresh water plume and fresh water inflow will increase the size of the fresh water plume. The plume will all ways try to find a balance in size between these aspects.

The fresh water inflow from the river will be dissipated in the coastal fresh water plume through mixing, as indicated in figure 6.4. Therefore, under stationary conditions eventually a balance will be found between dissipation and inflow of fresh water. The balance can be described using the equation below.

$$A_{plume} u_{plume} \tau_{plume} = \frac{1}{2}\Delta\rho g h_0^2 \tag{6.4}$$

Based on this simple balance, if the fresh water discharge increases, the plume size (A) is expected to grow as well, in order to dissipate more salinity and vice versa. Results can be obtained under the assumption that a fraction of the tidal dissipation is used for mixing.

Figure 6.4: Characteristic shape of a fresh water plume.

Bibliography

Aas, K., C. Czado, A. Frigessi, and H. Bakken
2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

Aydin, D. and B. Şenoğlu
2015. Monte Carlo Comparison of the Parameter Estimation Methods for the Two-Parameter Gumbel Distribution. *Journal of Modern Applied Statistical Methods*, 14(2):12.

Bedford, T. and R. M. Cooke
2001. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268.

Boon, J.
2007. Ernesto: Anatomy of a Storm Tide. *Virginia Institute of Marine Science, College of William and Mary*.

Breiman, L.
1996. Bagging Predictors. *Machine Learning*, 24(2):123–140.

Centre, T. G. R. D.
2017. Dataset discharge river Rhine. Koblenz, Germany.

De Boer, G.
2009. On the interaction between tides and stratification in the Rhine Region of Freshwater Influence.

de Boer, G. J., J. D. Pietrzak, and J. C. Winterwerp
2009. SST observations of upwelling induced by tidal straining in the Rhine ROFI. *Physics of Estuaries and Coastal Seas: Papers from the PECS 2006 Conference*, 29(1):263–277.

De Quay, J.
1967. Samenhang tussen stormvloeden en hoge rivierafvoeren. *Letter from the minister of transport, public works and water management's, Gravenhage*.

de Ruijter, W., A. Visser, and W. Bos
1997. The Rhine outflow: A prototypical pulsed discharge plume in a high energy shallow sea. *Journal of Marine Systems*, 12(1):263–276. 263.

Durack, P. and S. Wijffels
2010. Fifty-year trends in global ocean salinities and their relationship to broad-scale warming. *Journal of Climate*, 23(16):4342–4362.

Durack, P., S. Wijffels, and R. Matear
2012. Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *science*, 336(6080):455–458.

Fischer, E. M. and R. Knutti
2013. Robust projections of combined humidity and temperature extremes. *Nature Climate Change*, 3(2):126–130.

Flannigan, M., M. Krawchuk, W. de Groot, B. Wotton, and L. Gowman
2009. Implications of changing climate for global wildland fire. *International journal of wildland fire*, 18(5):483–507.

Garvine, R.
1995. A dynamical system for classifying buoyant coastal discharges. *Continental Shelf Research*, 15(13):1585–1596.

Garvine, R.
1999. Penetration of buoyant coastal discharge onto the continental shelf: A numerical model experiment. *Journal of Physical Oceanography*, 29(8):1892–1909.

Geerse, C.
2013. Correlatie tussen stormvloeden en afvoeren voor de benedenrivieren. *Mate van correlatie en geschatte invloed op de Toetspeilen, Report PR2442*, 10.

Giessen, A. v. d., W. De Ruijter, and J. Borst
1990. Three-dimensional current structure in the Dutch coastal zone. *Netherlands Journal of Sea Research*, 25(1):45–55. 45.

Gutierrez, B., N. Plant, and E. Thieler
2011. A Bayesian network to predict coastal vulnerability to sea level rise. *Journal of Geophysical Research: Earth Surface*, 116(F2):n/a–n/a.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae
2013. Global flood risk under climate change. *Nature Climate Change*, 3(9):816–821. 816.

Hollander, M., D. Wolfe, and E. Chicken
2013. *Nonparametric statistical methods.*, Wiley series in probability and statistics., third edition / edition. Hoboken, New Jersey :: John Wiley & Sons, Inc.,.

Horsburgh, K. and C. Wilson
2007. Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research: Oceans*, 112(C8).

Hurk, B. v. d., E. v. Meijgaard, P. d. Valk, K. v. Heeringen, and J. Gooijer
2015. Analysis of a compounding surge and precipitation event in the Netherlands. *Environmental Research Letters*, 10(3).

Jacobs, W.
2004. Modelling the Rhine river plume.

Kernkamp, H., G. Stelling, and E. Goede
2011. Efficient scheme for the shallow water equations on unstructured grids with application to the Continental Shelf. *Ocean Dynamics*, 61(8):1175–1188. 1175.

Kew, S., F. Selten, G. Lenderink, and W. Hazeleger
2013. The simultaneous occurrence of surge and discharge extremes for the Rhine delta. *Natural Hazards and Earth System Sciences*, 13(8):2017–2029.

Klerk, W., H. Winsemius, W. v. Verseveld, A. Bakker, and F. Diermanse
2015. The co-incidence of storm surges and extreme discharges within the Rhine-Meuse Delta. *Environmental Research Letters*, 10(3).

Koza, J. R.
1994. Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4(2):87–112.

Kumar, A., M. Chen, M. Hoerling, and J. Eischeid
2013. Do extreme climate events require extreme forcings? *Geophysical Research Letters*, 40(13):3440–3445.

Kurowicka, D. and R. Cooke
2006. *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons.

Leonard, M., S. Westra, A. Phatak, M. Lambert, B. van den Hurk, K. McInnes, J. Risbey, S. Schuster, D. Jakob, and M. Stafford-Smith
2014. A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1):113–128.

Leroux, B. G.
1992. Consistent estimation of a mixing distribution. *The Annals of Statistics*, Pp. 1350–1360.

Lian, J.
2013. Joint impact of rainfall and tidal level on flood risk in a coastal city with a complex river network: a case study of Fuzhou City, China. *HYDROLOGY AND EARTH SYSTEM SCIENCES*, 17(2):679–690.

Marin, A.
2010. Riders under storms: contributions of nomadic herders' observations to analysing climate change in Mongolia. *Global Environmental Change*, 20(1):162–176.

McLachlan, G. and D. Peel
2004. *Finite mixture models*. John Wiley & Sons.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne, and T. L. Delworth
2002. Increasing risk of great floods in a changing climate. *Nature*, 415(6871):514–517. 514.

Muis, S., M. Verlaan, H. Winsemius, J. C. Aerts, and P. Ward
2016. A global reanalysis of storm surges and extreme sea levels. *Nature communications*, 7.

Mutua, F.
1994. The use of the Akaike Information Criterion in the identification of an optimum flood frequency model. *Hydrological Sciences Journal*, 39(3):235–244. 235.

Pawlowicz, R., B. Beardsley, and S. Lentz
2002. Classical tidal harmonic analysis including error estimates in MATLAB using T_tide. *Computers & Geosciences*, 28(8):929–937.

Rappaport, E.
2000. Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bulletin of the American Meteorological Society*, 81(9):2065–2073.

Schepsmeier, U., J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler, T. Erhardt, C. Almeida, A. Min, C. Czado, and M. Hofmann
2018. Package 'VineCopula'.

Schmidt, M. and H. Lipson
2009. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85.

Schoelzel, C. and P. Friederichs
2008. Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. *Nonlin. Processes Geophys.*, 15(5):761–772.

Seneviratne, S., N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, contributors, , L. Alexander, S. Allen, G. Benito, and T. Cavazos
2012. *Chapter 3 Changes in climate extremes and their impacts on the natural physical environment*.

Sibson, R.
1981. A brief description of natural neighbor interpolation. *Interpreting multivariate data*, Pp. 21–36.

Simpson, J., W. Bos, F. Schirmer, A. Souza, T. Rippeth, S. Jones, and D. Hydes
1993. Periodic stratification in the rhine ROFI in the north-sea. *Oceanologica Acta*, 16(1):23–32.

Suijlen, J. and R. Duin
2002. Atlas of near-surface Total Suspended Matter concentrations in the Dutch coastal zone of the North Sea:.

Svensson, C. and D. Jones
2002. Dependence between extreme sea surge, river flow and precipitation in eastern Britain. *International Journal of Climatology*, 22(10):1149–1168.

Tables, J. P. o. O. and Standards
1981. *Tenth Report of the Joint Panel on Oceanographic Tables and Standards, Sidney, B. C., Canada 1-5 September 1980*. Unesco.

Taylor, G.
1922. Tidal oscillations in gulfs and rectangular basins. *Proceedings of the London Mathematical society*, 2(1):148–181.

Van den Brink, H., G. Können, J. Opsteegh, G. Van Oldenborgh, and G. Burgers
2005. Estimating return periods of extreme events from ECMWF seasonal forecast ensembles. *International Journal of Climatology*, 25(10):1345–1354.

van Verkeer en Waterstaat, M.
2006. *Hydraulische Randvoorwaarden 2006*. Ministerie van Verkeer en Waterstaat.

Van Wiechen, J.
2011. *Modelling the wind-driven motions in the Rhine ROFI*. PhD thesis, TU Delft, Delft.

Verlaan, M., S. De Kleermaeker, and L. Buckman
2015. GLOSSIS: Global storm surge forecasting and information system. P. 229. Engineers Australia and IPENZ.

Vinogradov, S. and R. Ponte
2010. Annual cycle in coastal sea level from tide gauges and altimetry. *Journal of Geophysical Research: Oceans*, 115(C4).

Wahl, T., S. Jain, J. Bender, S. Meyers, and M. Luther
2015. Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, 5(12):1093–1097.

Xu, S.
2007. *Analysis and predictions of extreme coastal water levels*. The Florida State University.

Zhang, Z., L. Ding, F. Zhang, and Z. Zhang
2015. Optimal Currency Composition for China's Foreign Reserves: A Copula Approach. *The World Economy*, 38(12):1947. 1947.

Zheng, F., M. Leonard, and S. Westra
2015. Efficient joint probability analysis of flood risk. *Journal of Hydroinformatics*, 17(4):584–597.

Zhong, H.
2013. A joint probability approach using a 1-D hydrodynamic model for estimating high water level frequencies in the Lower Rhine Delta. *Natural Hazards and Earth System Sciences*, 13(7):1841–1852.

A

Data information

Table A.1: Salinity data period 1975 untill 1985

Location	Start	End	Number of observations	Number of years	Number of observations per month
Appelzak 1 km uit de kust	16-06-75	21-02-83	174	7,7	1,9
Appelzak 10 km uit de kust	16-06-75	21-02-83	173	7,7	1,9
Appelzak 2 km uit de kust	16-06-75	17-12-85	226	10,5	1,8
Appelzak 20 km uit de kust	16-06-75	17-12-85	227	10,5	1,8
Appelzak 30 km uit de kust	16-06-75	21-02-83	173	7,7	1,9
Appelzak 4 km uit de kust	16-06-75	21-02-83	173	7,7	1,9
Appelzak 50 km uit de kust	16-06-75	21-02-83	173	7,7	1,9
Appelzak 70 km uit de kust	16-06-75	18-12-85	217	10,5	1,7
Callantsoog 1 km uit de kust	19-06-75	23-02-83	148	7,7	1,6
Callantsoog 10 km uit de kust	19-06-75	23-02-83	169	7,7	1,8
Callantsoog 2 km uit de kust	19-06-75	18-12-85	203	10,5	1,6
Callantsoog 20 km uit de kust	19-06-75	18-12-85	222	10,5	1,8
Callantsoog 30 km uit de kust	19-06-75	23-02-83	167	7,7	1,8
Callantsoog 4 km uit de kust	19-06-75	23-02-83	169	7,7	1,8
Callantsoog 50 km uit de kust	19-06-75	23-02-83	169	7,7	1,8
Callantsoog 70 km uit de kust	19-06-75	23-02-83	169	7,7	1,8
Egmond aan Zee 1 km uit de kust	18-06-75	23-02-83	158	7,7	1,7
Egmond aan Zee 10 km uit de kust	18-06-75	23-02-83	170	7,7	1,8
Egmond aan Zee 2 km uit de kust	18-06-75	23-02-83	169	7,7	1,8
Egmond aan Zee 20 km uit de kust	18-06-75	23-02-83	169	7,7	1,8
Egmond aan Zee 30 km uit de kust	18-06-75	23-02-83	169	7,7	1,8
Egmond aan Zee 4 km uit de kust	18-06-75	23-02-83	170	7,7	1,8
Egmond aan Zee 50 km uit de kust	18-06-75	23-02-83	168	7,7	1,8
Egmond aan Zee 70 km uit de kust	19-06-75	23-02-83	169	7,7	1,8
Goeree 10 km uit de kust	17-06-75	22-02-83	170	7,7	1,8
Goeree 20 km uit de kust	17-06-75	17-12-85	224	10,5	1,8
Goeree 30 km uit de kust	17-06-75	22-02-83	170	7,7	1,8
Goeree 50 km uit de kust	18-01-77	22-02-83	138	6,1	1,9
Goeree 6 km uit de kust	17-06-75	17-12-85	224	10,5	1,8
Goeree 70 km uit de kust	18-01-77	22-02-83	138	6,1	1,9
IJmuiden 20 km uit de kust	15-03-83	17-12-85	50	2,8	1,5
IJmuiden 3 km uit de kust	15-03-83	18-12-85	51	2,8	1,5
Noordwijk 1 km uit de kust	18-06-75	23-02-83	158	7,7	1,7
Noordwijk 10 km uit de kust	18-06-75	23-02-83	170	7,7	1,8
Noordwijk 2 km uit de kust	18-06-75	18-12-85	221	10,5	1,8
Noordwijk 20 km uit de kust	18-06-75	17-12-85	224	10,5	1,8

Table A.1: Salinity data period 1975 untill 1985

Location	Start	End	Number of observations	Number of years	Number of observations per month
Noordwijk 30 km uit de kust	18-06-75	23-02-83	169	7,7	1,8
Noordwijk 4 km uit de kust	18-06-75	23-02-83	169	7,7	1,8
Noordwijk 50 km uit de kust	18-06-75	23-02-83	168	7,7	1,8
Noordwijk 70 km uit de kust	18-06-75	23-02-83	168	7,7	1,8
Rottumerplaat 10 km uit de kust	24-06-75	24-02-83	151	7,7	1,6
Rottumerplaat 100 km uit de kust	23-06-75	24-02-83	153	7,7	1,7
Rottumerplaat 15 km uit de kust	24-06-75	24-02-83	152	7,7	1,7
Rottumerplaat 20 km uit de kust	24-06-75	09-12-85	204	10,5	1,6
Rottumerplaat 3 km uit de kust	24-06-75	16-12-85	268	10,5	2,1
Rottumerplaat 30 km uit de kust	24-06-75	24-02-83	153	7,7	1,7
Rottumerplaat 5 km uit de kust	24-06-75	25-02-83	147	7,7	1,6
Rottumerplaat 50 km uit de kust	24-06-75	24-02-83	154	7,7	1,7
Rottumerplaat 70 km uit de kust	23-06-75	27-11-85	163	10,4	1,3
Schouwen 1 km uit de kust	17-06-75	22-02-83	149	7,7	1,6
Schouwen 10 km uit de kust	17-06-75	17-12-85	223	10,5	1,8
Schouwen 20 km uit de kust	17-06-75	17-12-85	224	10,5	1,8
Schouwen 30 km uit de kust	17-06-75	28-03-83	173	7,8	1,9
Schouwen 4 km uit de kust	17-06-75	22-02-83	149	7,7	1,6
Schouwen 50 km uit de kust	17-06-75	22-02-83	172	7,7	1,9
Schouwen 70 km uit de kust	17-06-75	22-02-83	172	7,7	1,9
Ter Heide 1 km uit de kust	18-06-75	17-12-85	220	10,5	1,7
Ter Heide 10 km uit de kust	18-06-75	18-12-85	219	10,5	1,7
Ter Heide 2 km uit de kust	18-06-75	17-12-85	284	10,5	2,3
Ter Heide 20 km uit de kust	18-06-75	17-12-85	223	10,5	1,8
Ter Heide 30 km uit de kust	18-06-75	18-12-85	219	10,5	1,7
Ter Heide 4 km uit de kust	18-06-75	18-12-85	220	10,5	1,7
Ter Heide 50 km uit de kust	18-06-75	18-12-85	218	10,5	1,7
Ter Heide 70 km uit de kust	18-06-75	18-12-85	216	10,5	1,7
Terschelling 10 km uit de kust	23-06-75	24-02-83	156	7,7	1,7
Terschelling 20 km uit de kust	23-06-75	18-12-85	206	10,5	1,6
Terschelling 30 km uit de kust	23-06-75	24-02-83	156	7,7	1,7
Terschelling 4 km uit de kust	23-06-75	18-12-85	206	10,5	1,6
Terschelling 50 km uit de kust	23-06-75	24-02-83	155	7,7	1,7
Walcheren 1 km uit de kust	17-06-75	22-02-83	170	7,7	1,8
Walcheren 10 km uit de kust	17-06-75	22-02-83	173	7,7	1,9
Walcheren 2 km uit de kust	17-06-75	17-12-85	226	10,5	1,8
Walcheren 20 km uit de kust	17-06-75	17-12-85	226	10,5	1,8
Walcheren 30 km uit de kust	17-06-75	22-02-83	172	7,7	1,9
Walcheren 4 km uit de kust	17-06-75	22-02-83	175	7,7	1,9
Walcheren 50 km uit de kust	18-01-77	22-02-83	139	6,1	1,9
Walcheren 70 km uit de kust	18-01-77	22-02-83	138	6,1	1,9

Table A.2: Water level data period 1975 untill 1985

Location	Start	End	Number of observations	Number of years	Number of observations per month
Brouwershavensche Gat 08	1-1-1980	30-12-1985	52584	6.0	730.8
Cadzand	2-1-1975	30-12-1985	96383	11.0	730.7
Den Helder	2-1-1975	30-12-1985	96383	11.0	730.7
Hoek van Holland	2-1-1975	30-12-1985	96383	11.0	730.7
IJmuiden buitenhaven	1-1-1981	30-12-1985	43799	5.0	730.9
Noordwijk meetpost	2-9-1982	30-11-1985	131890	3.2	3387.6
Oostkapelle	1-1-1979	30-12-1985	61343	7.0	730.8
Scheveningen	2-1-1975	30-12-1985	96383	11.0	730.7
Westkapelle	2-1-1975	30-12-1985	96382	11.0	730.7

Table A.3: Water level data period 2001 untill 2016

Location	Start	End	Number of observations	Number of years	Number of observations per month
Brouwershavensche Gat 08	02-01-01	30-12-16	836846	16,0	4360,7
Cadzand	02-01-01	30-12-16	841242	16,0	4383,6
Den Helder	02-01-01	30-12-16	846694	16,0	4412,0
Euro platform	01-07-01	30-12-16	815327	15,5	4383,7
Hoek van Holland	02-01-01	30-12-16	862996	16,0	4497,0
IJmuiden buitenhaven	02-01-01	30-12-16	841248	16,0	4383,7
Lichteiland Goeree	01-07-01	30-12-16	821058	15,5	4414,5
Noordwijk meetpost	02-01-01	05-07-06	289493	5,5	4383,7
Scheveningen	02-01-01	30-12-16	841248	16,0	4383,7
Westkapelle	02-01-01	30-12-16	841248	16,0	4383,7

Table A.4: Correlation salinity waterlevel overlapping period

Salinity location	Waterlevel location	Date start	Date end
Appelzak 2 km uit de kust	Cadzand	16-06-75	17-02-88
Ter Heide 1 km uit de kust	Hoek van Holland	18-06-75	19-02-88
IJmuiden 3 km uit de kust	IJmuiden buitenhaven	15-03-83	16-02-88
Noordwijk 10 km uit de kust	Noordwijk meetpost	02-09-82	05-07-06
Schouwen 1 km uit de kust	Oostkapelle	01-01-79	22-02-83
Ter Heide 1 km uit de kust	Scheveningen	18-06-75	19-02-88
Walcheren 2 km uit de kust	Westkapelle	17-06-75	14-12-16

B

Salinity - Mean water level transformed to standard normal

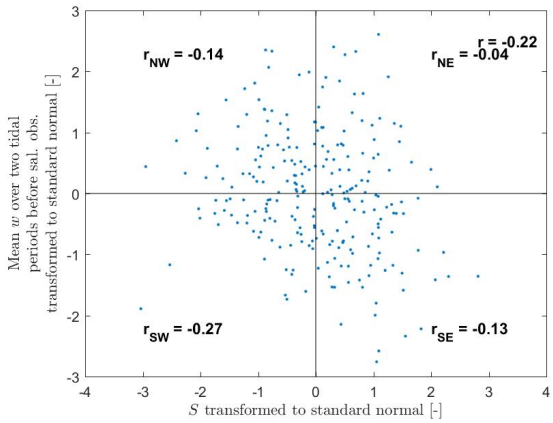


Figure B.1: Correlation mean water level (averaged over two tidal periods) and salinity observations, observations are transformed to standard normal

C

Storm surge definition

This appendix gives the definition, which is applied in this thesis, of storm surge.

Stormsurge

For this research, the definitions of tide (w_t) and storm surge (s_{wl}) are of importance. Tide is defined as the harmonic variation in water level under average meteorological conditions, due to the gravitational pull of celestial bodies. Storm surge is defined as the meteorological contributions to the water level (Horsburgh and Wilson, 2007). Therefore we can consider storm surge is defined as the difference between the observed water level (w) and an predicted tidal water level, as described by Boon (2007) and formula C.1. It should be noted that, with this definition, storm surge contains the genuinely meteorological storm surge, tide–surge interaction, and harmonic prediction errors.

$$s_{wl}(t) = w(t) - w_t(t) \tag{C.1}$$

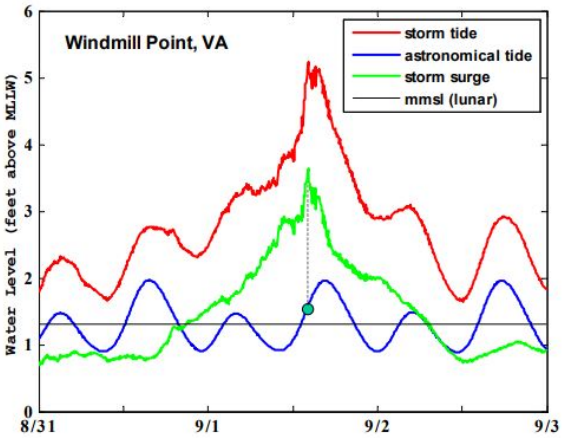


Figure C.1: The definition of storm surge by Boon (Boon, 2007), during the tropical depression ERNESTO at Windmill point, Virginia

The storm surge is in this case a time dependent signal, since the water level observation signal and the astronomical tide signal are time dependent. Figure C.1 shows the definition of a storm surge, based on a case in Virginia.

D

Discharge - Mean water level transformed to standard normal

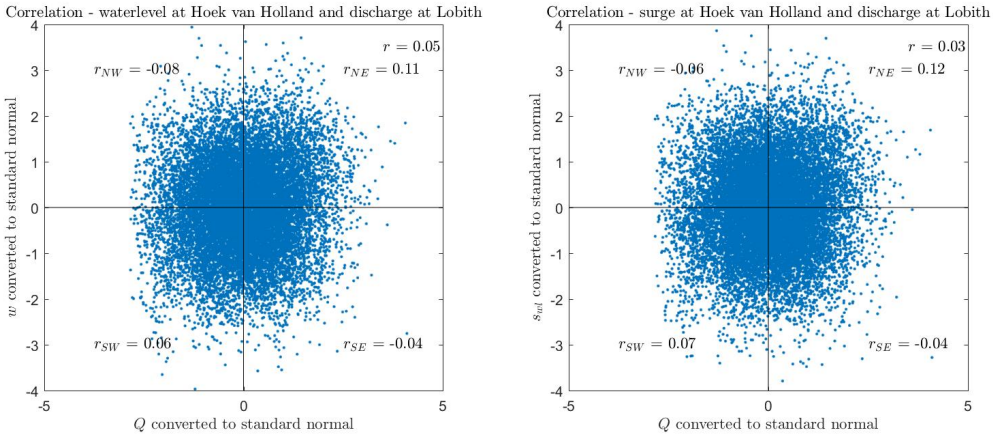


Figure D.1: Correlation structure for the discharge at Lobith and the daily mean water level (left) and -surge (right) at Hoek van Holland. The observations have been transformed to standard normal in these graphs.

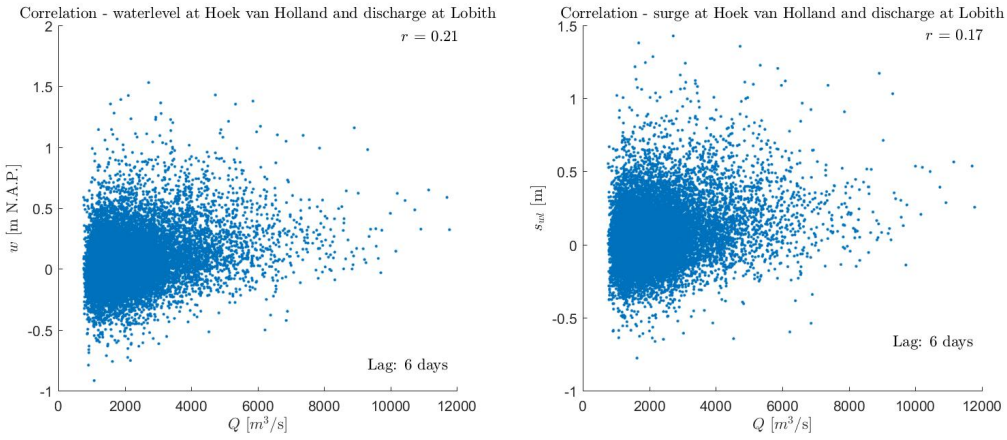


Figure D.2: Correlation structure for the discharge at Lobith and the daily mean water level (left) and -surge (right) at Hoek van Holland, with a lag of 6 days between the time series.

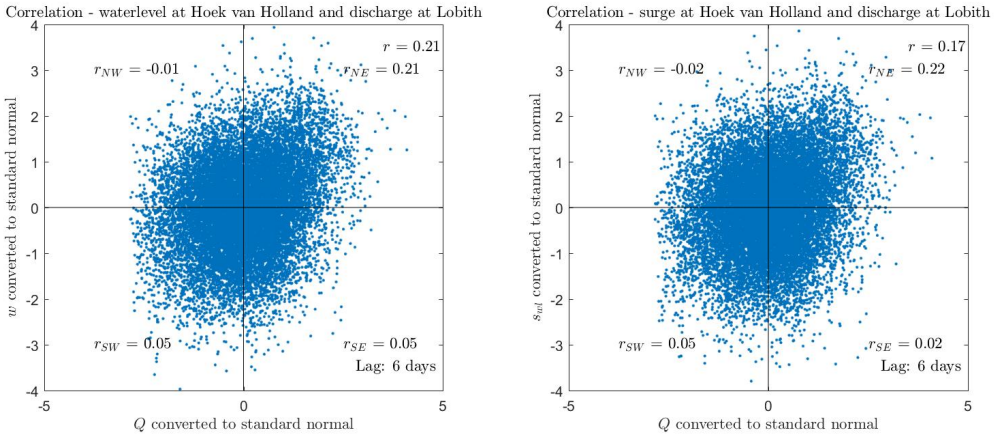


Figure D.3: Correlation structure for the discharge at Lobith and the daily mean water level (left) and -surge (right) at Hoek van Holland, with a lag of 6 days between the time series. The observations have been transformed to standard normal in these graphs.

E

Salinity - Discharge transformed to standard normal

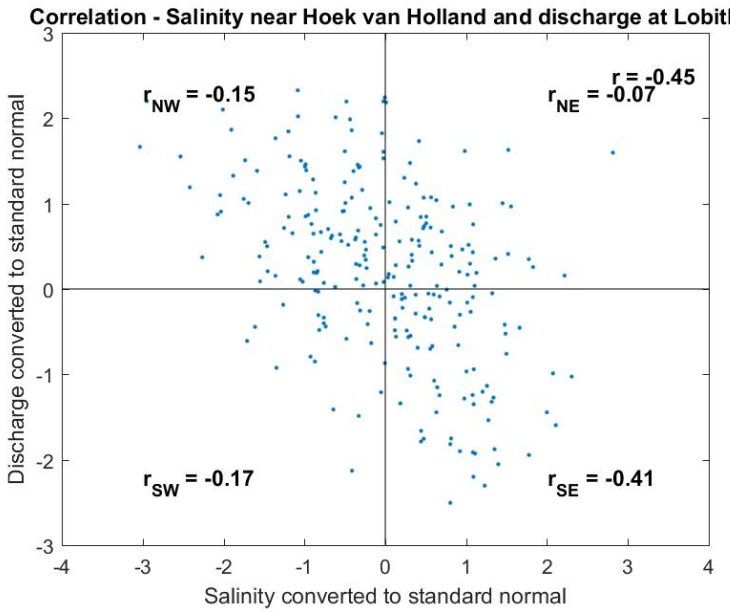


Figure E.1: Correlation structure for salinity near Hoek van Holland and discharge at Lobith. The observations have been transformed to standard normal in these graphs.

F

Results numerical experiments

Table F.1: Results numerical experiments, first column indicates the number of the experiment. The standard case is experiment number 15.

#	Q $[m^3/s]$	S $[psu]$	W $[m]$	D $[m]$	w_{t_0} $[m]$	f $[rad/s]$	g' $[m/s^2]$	ci $[m/s]$	ri $[m]$
1	1000	35	500	10	0.764	1.14E-4	0.108	1.040	9.11E+3
2	8000	35	500	80	1.011	1.14E-4	0.028	1.497	1.31E+4
3	16000	35	500	160	1.022	1.14E-4	0.021	1.825	1.60E+4
4	4000	35	1000	20	0.906	1.14E-4	0.085	1.303	1.14E+4
5	2000	35	4000	20	0.918	1.14E-4	0.035	0.839	7.35E+3
6	2000	35	500	200	1.070	1.14E-4	0.005	0.950	8.32E+3
7	2000	35	500	40	1.005	1.14E-4	0.027	1.042	9.13E+3
8	250	35	500	20	0.920	1.14E-4	0.017	0.582	5.10E+3
9	2000	35	2000	20	0.934	1.14E-4	0.048	0.978	8.57E+3
10	500	35	500	20	0.924	1.14E-4	0.027	0.737	6.46E+3
11	2000	35	500	20	1.354	1.14E-4	0.065	1.137	9.97E+3
12	2000	35	500	20	0.472	1.14E-4	0.053	1.027	9.00E+3
13	10000	35	500	20	0.93	1.14E-4	0.173	1.861	1.63E+4
14	16000	35	4000	20	0.94	1.14E-4	0.138	1.661	1.46E+4
15	2000	35	500	20	0.93	1.14E-4	0.063	1.121	9.82E+3
16	2000	30	500	20	0.933	1.14E-4	0.057	1.07	9.37E+3
17	2000	35	8000	20	0.944	1.14E-4	0.034	0.82	7.18E+3
18	8000	35	2000	20	0.901	1.14E-4	0.112	1.494	1.31E+4
19	2000	35	10000	20	0.942	1.14E-4	0.031	0.793	6.95E+3
20	2000	35	500	20	1.148	1.14E-4	0.064	1.133	9.93E+3
21	2000	35	500	20	0.702	1.14E-4	0.061	1.102	9.65E+3
22	2000	35	500	10	0.777	1.14E-4	0.169	1.298	1.14E+4
23	32000	35	8000	20	0.963	1.14E-4	0.183	1.911	1.67E+4
24	3500	35	500	20	0.909	1.14E-4	0.090	1.339	1.17E+4
25	2000	35	500	80	0.995	1.14E-4	0.012	0.975	8.54E+3
26	4000	35	500	40	0.993	1.14E-4	0.040	1.265	1.11E+4
27	2000	40	500	20	0.926	1.14E-4	0.068	1.162	1.02E+4
28	1000	35	500	20	0.928	1.14E-4	0.042	0.914	8.01E+3
29	5000	35	500	20	0.918	1.14E-4	0.111	1.492	1.31E+4
30	2000	35	1000	20	0.934	1.14E-4	0.058	1.075	9.42E+3
31	2000	35	500	20	1.756	1.14E-4	0.062	1.111	9.74E+3
32	2000	35	500	20	0.934	1.11E-4	0.063	1.12	1.01E+4
33	2000	35	500	20	0.962	9.35E-5	0.064	1.128	1.21E+4
34	2000	35	500	20	0.91	1.26E-4	0.062	1.114	8.85E+3
35	2000	35	500	20	0.892	1.37E-4	0.061	1.108	8.11E+3
36	2000	35	500	20	0.995	7.27E-5	0.065	1.137	1.56E+4
37	2000	35	500	20	1.03	4.97E-5	0.067	1.155	2.32E+4
38	2000	35	500	20	1.068	2.53E-5	0.067	1.156	4.58E+4
39	2000	35	500	20	0.881	1.43E-4	0.061	1.102	7.69E+3
40	50000	35	16000	5	0.11	1.14E-4	0.260	1.141	1.00E+4
41	50000	35	16000	10	0.37	1.14E-4	0.260	1.614	1.41E+4
42	40000	35	16000	5	0.12	1.14E-4	0.260	1.141	1.00E+4
43	40000	35	500	20	0.916	1.14E-4	0.260	2.282	2.00E+4
44	30000	35	16000	5	0.128	1.14E-4	0.260	1.141	1.00E+4
45	20000	35	16000	5	0.135	1.14E-4	0.260	1.141	1.00E+4
46	50000	35	500	20	0.913	1.14E-4	0.260	2.282	2.00E+4
47	40000	35	16000	10	0.375	1.14E-4	0.260	1.614	1.41E+4
48	30000	35	500	20	0.919	1.14E-4	0.260	2.282	2.00E+4
49	20000	35	16000	10	0.358	1.14E-4	0.189	1.373	1.20E+4
50	20000	35	500	20	0.921	1.14E-4	0.260	2.282	2.00E+4
51	30000	35	16000	10	0.382	1.14E-4	0.260	1.613	1.41E+4
52	2000	35	500	120	0.823	1.14E-4	0.008	0.959	8.40E+3
53	2000	35	500	200	0.876	1.14E-4	0.005	0.95	8.32E+3
54	2000	35	500	80	0.808	1.14E-4	0.012	0.975	8.54E+3
55	2000	35	500	160	0.846	1.14E-4	0.006	0.954	8.36E+3
56	2000	35	500	60	0.798	1.14E-4	0.016	0.986	8.64E+3

Table F.2: Results numerical experiments, first column indicates the number of the experiment. The standard case is experiment number 15.

#	τ [−]	K_i [−]	ϵ [−]	I_r [−]	$1+\frac{Re}{Re_s}$ [−]	a [−]	$y_p[m]$
1	2.11E-2	5.49E-2	0.924	50	1.464	0.91	1.45E+4
2	1.02E-2	3.81E-2	0.987	6.25	6.342	1.76	
3	6.85E-3	3.13E-2	0.994	3.13	14.02	1.81	
4	2.69E-2	8.76E-2	0.955	50	2.162	1.06	5.21E+4
5	3.24E-2	5.44E-1	0.954	200	1.749	0.88	3.31E+4
6	2.53E-3	6.01E-2	0.995	2.5	9.47	1.82	
7	1.05E-2	5.48E-2	0.975	12.5	2.859	1.54	4.29E+4
8	8.41E-3	9.8E-2	0.954	25.00	1.519	1.05	8.68E+3
9	2.39E-2	2.33E-1	0.953	100	1.872	0.94	3.18E+4
10	1.05E-2	7.74E-2	0.954	25	1.657	1.10	1.14E+4
11	1.76E-2	5.02E-2	0.932	25	2.014	1.02	1.82E+4
12	2.16E-2	5.56E-2	0.976	25	1.916	1.48	4.91E+4
13	3.29E-2	3.07E-2	0.953	25	2.66	1.08	
14	6.62E-2	2.75E-1	0.953	200	2.481	1.04	
15	1.82E-2	5.09E-2	0.954	25	2.000	1.06	3.21E+4
16	2.00E-2	5.34E-2	0.953	25	1.954	1.05	2.96E+4
17	3.39E-2	1.11E+0	0.953	400	1.731	0.75	5.24E+4
18	4.09E-2	1.53E-1	0.955	100	2.332	1.07	
19	3.63E-2	1.44E+0	0.953	500	1.707	0.75	5.32E+4
20	1.78E-2	5.04E-2	0.943	25	2.01	1.00	2.29E+4
21	1.88E-2	5.18E-2	0.965	25	1.983	1.23	4.88E+4
22	2.71E-2	4.4E-2	0.922	50	1.579	0.84	2.35E+4
23	1.00E-1	4.78E-1	0.952	400	2.705	0.92	
24	2.23E-2	4.26E-2	0.955	25	2.194	1.09	4.85E+4
25	6.01E-3	5.86E-2	0.988	6.25	4.477	1.79	
26	1.43E-2	4.51E-2	0.975	12.5	3.256	1.58	4.92E+4
27	1.69E-2	4.91E-2	0.954	25	2.036	1.08	3.52E+4
28	1.37E-2	6.24E-2	0.954	25	1.815	1.09	1.77E+4
29	2.56E-2	3.83E-2	0.954	25	2.331	1.11	5.24E+4
30	1.98E-2	1.06E-1	0.953	50	1.958	1.01	3.12E+4
31	1.85E-2	5.13E-2	0.912	25	1.991	1.04	1.10E+4
32	1.78E-2	4.97E-2	0.953	25	1.999	1.06	3.17E+4
33	1.47E-2	4.15E-2	0.952	25	2.006	1.05	2.91E+4
34	2.03E-2	5.65E-2	0.955	25	1.994	1.07	3.33E+4
35	2.23E-2	6.17E-2	0.955	25	1.988	1.09	3.42E+4
36	1.12E-2	3.20E-2	0.95	25	2.014	1.04	2.62E+4
37	7.45E-3	2.15E-2	0.949	25	2.03	1.01	2.25E+4
38	3.78E-3	1.09E-2	0.947	25	2.031	1.01	2.05E+4
39	2.36E-2	6.5E-2	0.956	25	1.983	1.09	3.53E+4
40	1.75E+0	1.6E+0	0.995	3200	1.254	0.77	
41	4.38E-1	1.13E+0	0.982	1600	1.719	0.71	2.25E+5
42	1.40E+0	1.6E+0	0.994	3200	1.254	0.84	
43	8.77E-2	2.5E-2	0.954	25	3.035	0.92	1.03E+5
44	1.05E+0	1.6E+0	0.994	3200	1.254	0.82	
45	7.01E-1	1.6E+0	0.993	3200	1.254	0.76	
46	1.10E-1	2.5E-2	0.954	25	3.035	0.88	9.98E+4
47	3.51E-1	1.13E+0	0.981	1600	1.719	0.72	2.10E+5
48	6.58E-2	2.50E-2	0.954	25	3.035	0.92	1.04E+5
49	2.42E-1	1.33E+0	0.982	1600	1.612	0.95	1.03E+5
50	4.38E-2	2.5E-2	0.954	25	3.035	1.03	1.00E+5
51	2.63E-1	1.13E+0	0.981	1600	1.719	0.90	1.40E+5
52	4.14E-3	5.95E-2	0.993	4.17	6.132	1.81	1.04E+5
53	2.53E-3	6.01E-2	0.996	2.5	9.47	1.81	7.39E+4
54	6.01E-3	5.86E-2	0.990	6.25	4.477	1.79	7.62E+4
55	3.13E-3	5.98E-2	0.995	3.13	7.808	1.81	9.08E+4
56	7.82E-3	5.79E-2	0.987	8.33	3.639	1.77	4.25E+4

G

Fits parameterization uncertainty - survival plots

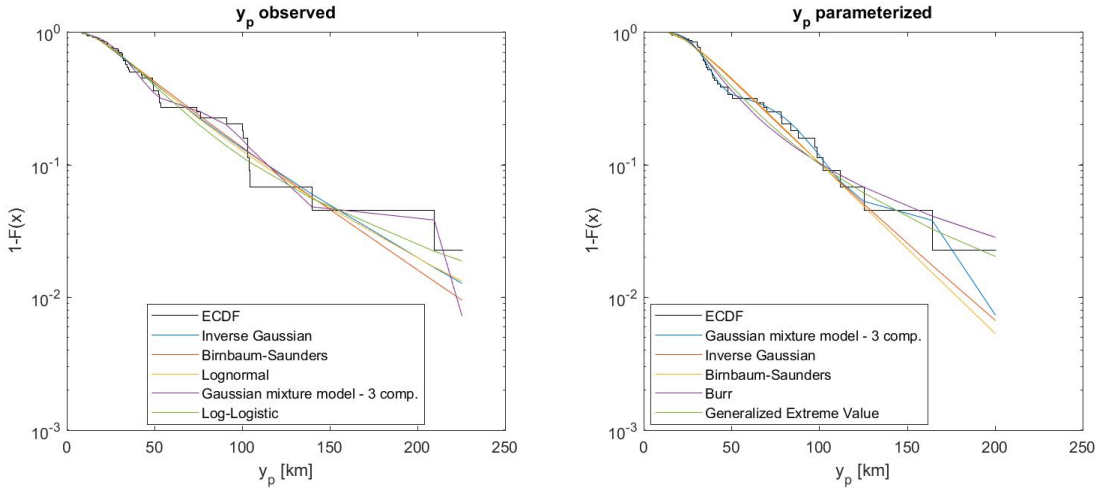


Figure G.1: All parameteric distribution have been fitted by MLE. Except for the Gaussian mixture models, these have been fitted by the Expectation Maximization algorithm. The fits are sorted by AIC and for both data sets the inverse gaussian distribution has been used.

H

Fits data uncertainty - survival plots

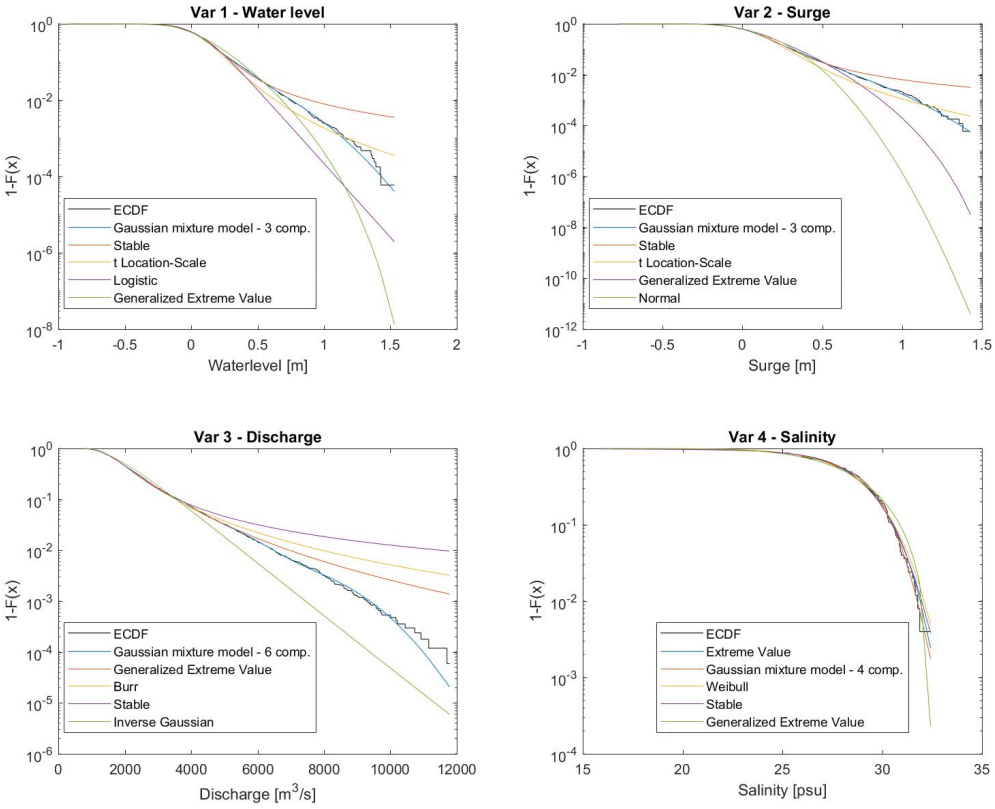


Figure H.1: All parameteric distribution have been fitted by MLE. Except for the Gaussian mixture models, these have been fitted by the Expectation Maximization algorithm. The fits are sorted by AIC and the top distribution in each legend has been used in further analysis.